GOVERNANCE OF/ THROUGH BIG DATA

Volume I

A cura di
Giorgio Resta
Vincenzo Zeno-Zencovich

Consumatori e Mercato **13**

Università degli Studi Roma Tre
Dipartimento di Giurisprudenza

Nella stessa Collana

1. V. Zeno-Zencovich (a cura di), *Cosmetici. Diritto, regolazione, bio-etica*, 2014
2. M. Colangelo, V. Zeno-Zencovich, *Introduction to European Union transport law*, I ed. 2015; II ed. 2016; III ed. 2019
3. G. Resta, V. Zeno-Zencovich (a cura di), *Il diritto all'oblio su Internet dopo la sentenza Google Spain*, 2015
4. V. Zeno-Zencovich, *Sex and the contract* (II ed.), 2015
5. G. Resta, V. Zeno-Zencovich (a cura di), *La protezione transnazionale dei dati personali. Dai "safe harbour principles" al "privacy shield"*, 2016
6. A. Zoppini (a cura di), *Tra regolazione e giurisdizione*, 2017
7. C. Giustolisi (a cura di), *La direttiva consumer rights. Impianto sistematico della direttiva di armonizzazione massima*, 2017
8. R. Torino (a cura di), *Introduction to European Union internal market law*, 2017
9. M.C. Paglietti, M.I. Vangelisti (a cura di), *Innovazione e regole nei pagamenti digitali. Il bilanciamento degli interessi nella PSD2*, 2020
10. L. Scaffardi, V. Zeno-Zencovich (a cura di), *Cibo e diritto. Una prospettiva comparata*, 2020
11. A.M. Mancaleoni, E. Poillot (a cura di), *National Judges and the Case Law of the Court of Justice of the European Union*, 2020
12. E. Poillot, G. Lenzini, G. Resta, V. Zeno-Zencovich, *Data Protection in the Context of Covid-19. A Short (Hi)Story of Tracing Applications*, 2021

Università degli Studi Roma Tre
Dipartimento di Giurisprudenza

GOVERNANCE OF/ THROUGH BIG DATA

Volume I

A cura di
Giorgio Resta
Vincenzo Zeno-Zencovich

Consumatori e Mercato 13

Roma TrE-Press
2023

Coordinamento redazionale e editoriale:
Gruppo di Lavoro *Roma TrE-Press*

Collana pubblicata nel rispetto del Codice etico adottato dal Dipartimento di Giurisprudenza dell'Università degli Studi Roma Tre, in data 22 aprile 2020.

Elaborazione grafica della copertina: MOSQUITO. mosquitoroma.it

Caratteri tipografici utilizzati:
Brandon Grotesque (copertina e frontespizio)
Adobe Garamond Pro (testo)

Impaginazione e cura editoriale: Colitti-Roma colitti.it

Edizioni: *Roma TrE-Press* ©
Roma, maggio 2023
ISBN: 979-12-5977-173-5

http://romatrepress.uniroma3.it

Questo volume è pubblicato nel quadro del PRIN 2017-2017BAPSXF "Governance of/through Big Data: Challenges for European Law", finanziato dal Ministero dell'Università e della Ricerca.

L'attività della *Roma TrE-Press* è svolta nell'ambito della
Fondazione Roma Tre-Education, piazza della Repubblica 10, 00185 Roma

Presentazione della collana "Consumatori e Mercato"

Direttore: Vincenzo Zeno-Zencovich
Comitato Scientifico:
Guido Alpa, Marcello Clarich, Alberto Musso

La Collana "Consumatori e mercato", pubblicata in open access dalla Roma TrE-Press, intende essere una piattaforma editoriale multilingue, avente ad oggetto studi attinenti alla tutela dei consumatori e alla regolazione del mercato. L'intento è di stimolare un proficuo scambio scientifico attraverso una diretta partecipazione di studiosi appartenenti a diverse discipline, tradizioni e generazioni.

Il dialogo multidisciplinare e multiculturale diviene infatti una componente indefettibile nell'àmbito di una materia caratterizzata da un assetto disciplinare ormai maturo tanto nelle prassi applicative del mercato quanto nel diritto vivente. L'attenzione viene in particolare rivolta al contesto del diritto europeo, matrice delle scelte legislative e regolamentari degli ordinamenti interni, e allo svolgimento dell'analisi su piani differenti (per estrazione scientifica e punti di osservazione) che diano conto della complessità ordinamentale attuale.

****

The "Consumer and market" series published, in open access, by Roma TrE-Press, aims at being a multilingual editorial project, which shall focus on consumer protection and market regulation studies. The series' core mission is the promotion of a fruitful scientific exchange amongst scholars from diverse legal systems, traditions and generations. This multidisciplinary and multicultural exchange has in fact become fundamental for a mature legal framework, from both the market practice and the law in action standpoints. A particular focus will be given on European law, where one can find the roots of the legislation and regulation in the domestic legal systems, and on the analysis of different levels, in line with the current complexity of this legal sector.

Contents

VOLUME I

SECTION I
Algorithms and Artificial Intelligence

Fabiana Di Porto
Algorithmic Disclosure Rules

Annalisa Signorelli
La prevedibilità della e nella decisione giudiziaria

SECTION II
ANTITRUST, A.I. AND BIG DATA

Contents

VOLUME II

SECTION III
Big Data

SECTION IV
Data Governance

Preface

This volume collects some of the articles and papers published within the "Governance of/through Big Data: Challenges for European Law" research project.

When, in 2016, it was presented to the Italian Ministry of Universities the topics of Big Data and of Artificial Intelligence were relatively unexplored. Since then, there has been a flourish of other projects, publications and congresses.

We trust that the wide participation of academic institutions in this project – Roma Tre University, as project lead; Bocconi University in Milan; LUMSA University in Rome; Salento University in Lecce and Turin Polytechnic – has contributed to the awareness and involvement of Italian scholars in novel territories of legal research and to their outreach abroad.

We are extremely grateful to Dr. Vanessa Villanueva Collao for her careful editing of the many contributions and organization of the volume.

Rome, May 2023

Giorgio Resta
Vincenzo Zeno-Zencovich

SECTION I

Algorithms and Artificial Intelligence

Alessandro Mantelero

Electronic Democracy and Digital Justice

Abstract: A growing debate in several European fora is paving the way for future rules for Artificial Intelligence (AI). A principles-based approach prevails, with various lists of principles drawn up in recent years. These lists, which are often built on human rights, are only a starting point for a future regulation. It is now necessary to move forward, turning abstract principles into a context-based response to the challenges of AI. This article therefore places the principles and operational rules of the current European and international human rights framework in the context of AI applications in two core, and little explored, areas of digital transformation: electronic democracy and digital justice. Several binding and non-binding legal instruments are available for each of these areas, but they were adopted in a pre-AI era, which affects their effectiveness in providing an adequate and specific response to the challenges of AI. Although the existing guiding principles remain valid, their application should therefore be reconsidered in the light of the social and technical changes induced by AI. To contribute to the ongoing debate on future AI regulation, this article outlines a contextualised application of the principles governing e- democracy and digital justice in view of current and future AI applications.

1. *AI Challenges and Human Rights*

Artificial Intelligence (AI) is part of our daily life. It is used to moderate public debate, fashion the social environment and support human decision-makers in various fields, including justice. AI is therefore a component of many decision-making processes affecting individuals and groups, actively shaping our communities and personal lives[1]. This means

[*] This article was published in *Direito Público*, 2022, 18(100). https://doi.org/10.11117/rdp.v18i100.6199.

[1] For an analysis of the different impacts of AI on individuals and society, see Council of Europe, *Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications*, 2018. Available at <https://rm.coe.int/algorithms-and-humanrights-en-rev/16807956b5>, accessed on Jan. 15, 2019); A. Mantelero & M.S. Esposito, *An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems*, 41 *Computer L. & Sec. Rev.* 1 (2021), DOI: 10.1016/j.clsr.2021.105561; F. Zuiderveen Borgesius, *Strengthening Legal Protection against Discrimination by Algorithms and*

that AI is no longer a mere technical or marketing trend but a regulatory issue[2], given the social consequences and, in some cases, legal effects.

To correctly frame this debate, it is important to keep in mind the difference between natural and artificial intelligence, where the latter is nothing more than a data-driven and mathematical form of information processing[3]. AI is not able to think, elaborate concepts or develop theories of causality: AI merely takes a path recognition approach to order huge amounts of data and infer new information and correlations.

Data dependence is both the strength and the weakness of these systems. Poor data undermines the quality of their results[4], datafication can only partially represent reality[5] and incredibly large datasets and complex AI solutions often do not allow human decision makers to inspect and check the 'reasoning' of the machine[6]. The upshot of these technical and structural constraints can be summed up under three main headings: bias, obscurity, and ownership.

Regarding bias, the design and development of AI tools can be affected

Artificial Intelligence, 24 *Int'l. J. of Human Rights* 1572 (2020).

[2] See European Commission, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, COM(2021)206 final, 2021; Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI), *Feasibility Study*, CAHAI(2020)23, 2020, available at <https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da>. Accessed on Jul. 29, 2021); Council of Europe, *Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems*, 2020; Council of Europe, *Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data (Convention 108)*, 2019; OECD, *Recommendation of the Council on Artificial Intelligence*, 2019; UNESCO, *Draft Text of the Recommendation on the Ethics of Artificial Intelligence*, 2021, available at < https://unesdoc.unesco.org/ark:/48223/ pf0000377897>, accessed on Sept. 3, 2021. See also A. Veronese, A. Nunes, Lopez Espiñeira Lemos, *Trayectoria normativa de la inteligencia artificial en los países de Latinoamérica con un marco jurídico para la protección de datos: límites y posibilidades de las políticas integradoras*, in *Revista Latinoamericana de Economía y Sociedad Digital*, 2, 2021, available at <https://revistalatam.digital/article/210207/>, accessed on Aug. 27, 2021.

[3] See Hildebrandt, *The Issue of Bias. The Framing Powers of Machine Learning*, in M. Pelillo, T. Scantamburlo (eds.) *Machines We Trust. Perspectives on Dependable AI*, Cambridge, 2021.

[4] See European Union Agency for Fundamental Rights, *Data Quality and Artificial Intelligence—Mitigating Bias and Error to Protect Fundamental Rights*, 2019.

[5] P.E. Agre, *Surveillance and Capture: Two Models of Privacy*, 10 *The Information Soc'y* 101 (1994); Hildebrandt, *Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning*, 20 *Theoretical Inquiries in L.* 83 (2019).

[6] D. Kolkman, *The (in)Credibility of Algorithmic Models to Non-Experts*, 25 *Information, Communication & Soc'y* 93 (2020), doi <10.1080/1369118X.2020.1761860>.

by different biases that, in many cases, differ from human bias[7]. Bias does not only concern the much debated data quality (for example selection bias)[8], but also the methodologies adopted (e.g., pre- processing and data cleaning biases, measurement bias, bias in survey methodologies)[9], the target of investigation (e.g., historical bias in pre- existing data-sets and under- or over-representation of certain groups in new data-sets), and the psychological attitude of the data scientists (e.g., confirmation bias).

This brief listing of potential biases also reveals the human component of AI solutions, often underestimated in a misleading comparison between humans and machines. This dichotomy understates the role of human intervention in AI data processing[10] and the intentional or unintentional transposition of developers' views into the AI reference values used for classification[11].

As for obscurity, this concerns both the AI tools used and the way they

[7] M.L. Cummings et al., Chatham House Report, *Artificial Intelligence and International Affairs. Disruption Anticipated*, London, 2018, available at <https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial--intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf>, accessed on Mar. 21, 2020; R. Caruana et Al., *Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission*, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015; K. Eykholt et al., *Robust Physical-World Attacks on Deep Learning Visual Classification*, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, available at https://openaccess.thecvf.com/content_cvpr_2018/ papers/Eykholt_Robust_Physical-World_Attacks_CVPR_2018_paper.pdf, accessed on Apr. 23, 2021.

[8] Ai Now Institute, *AI Now 2017 Report*, New York, 2017, pp. 4, 16-17, available at <https://assets.contentful.com/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf>, accessed on Oct. 26, 2017.

[9] M. Veale, R. Binns, *Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data*, 4 *Big Data & Soc'y* 1 (2017), doi:10.1177/2053951717743530.

[10] P. Tubaro, A. Casilli, M. Coville, *The Trainer, the Verifier, the Imitator: Three Ways in Which Human Platform Workers Support Artificial Intelligence*, 7 *Big Data & Soc'y* 1 (2020), doi: 10.1177/2053951720919776; K. Crawford, V. Joler, *Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources*, 2018, available at <http://www.anatomyof.ai>, accessed on Dec. 27, 2019; R. Adams, N. Loideain, *From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of Data Protection Impact Assessments*, 36 *Computer L. & Sec. Rev*. (2020), doi:10.1016/j.clsr.2019.105366.

[11] S. M. West, M. Whittaer, K. Crawford, *Discriminating Systems. Gender, Race, and Power in AI*, 2019, available at <https://ainowinstitute.org/ discriminatingsystems.pdf>, accessed on May 15, 2019.

impact on individuals, whose circumstances are analysed and represented through them. Not only is the way some AI applications actually function and process information unknown[12], even to data scientists, but individuals are often unaware of their being dynamically grouped on the basis of unseen correlations and inferences, without being able to know the identity of the other members of the group. Obscurity therefore entails two different consequences: first, data scientists are unable to clearly justify the specific decisions suggested by AI; and second, people are passively scrutinised by AI without having a meaningful or effective role in AI design or the opportunity to voice their collective interests[13].

This level of obscurity and the limitations to democratic participation in AI development is heightened by a third feature of many AI products: ownership. The proprietary nature of the algorithms used and, in certain cases, of the data silos used to train and implement them mean that intellectual property rights are a further barrier to access to the architecture of these applications and to public oversight[14].

These three inherent constraints – bias, obscurity, and ownership – have a direct impact on the challenges of AI and its social acceptance in monitoring and governing human activities (e.g., smart cities),[15] offering personalised services (e.g., predictive medicine)[16] and, more in general,

[12] A. D. Selbst, *Disparate Impact in Big Data Policing*, 52 *Georgia L. Rev*. 109, 163 (2017); J. Burrell, *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, 3 *Big Data & Soc'y* (2016), doi: 10.1177/2053951715622512; R. Brauneis, E. P. Goodman, *Algorithmic Transparency for the Smart City*, 20 *Yale J.L. & Technol.* 103, 131 (2018).

[13] C. B. Graber, *Artificial Intelligence, Affordances and Fundamental Rights*, in M. Hildebrandt, K. O'Hara (eds.), *Life and the Law in the Era of Data-Driven Agency*, Glos-Massachusets, 2020; A. Mantelero, *Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection*, 32 *Computer L. & Security Rev*. 238 (2016).

[14] F. Pasquale, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Cambridge-London, 2015, p. 193.

[15] Privacy International, *Smart Cities: Utopian Vision, Dystopian Reality*, 2017, available at <https://privacyinternational.org/sites/default/files/2017-12/Smart%20Cities-utopian%20Vision%2C%20Dystopian%20Reality.pdf>, accessed on May 12, 2020; E. Goodman, J. Powles, *Urbanism Under Google: Lessons from Sidewalk Toronto*, 88 *Fordham L. Rev*. 457 (2019); J. E. Cohen, *Between Truth and Power. The Legal Construction of Informational Capitalism*, New York, 2019, pp. 62-3.

[16] K. Ferryman, M. Pitcan, *Fairness in Precision Medicine*, 2018, available at <https://datasociety.net/wp-content/uploads/2018/02/Data.Society.Fairness.In_. Precision. Medicine.Feb2018.FINAL-2.26.18.pdf>, accessed on Apr. 8, 2018.

supporting humans in the decision-making process.

Issues surrounding data-intensive solutions and their use in decision-making processes concern a variety of interests related to human rights and freedoms[17]. To address the growing concern about the potential impact of AI on human rights and freedoms, several initiatives have been proposed at local, national and international levels, and a variety of guidelines have been drawn up by NGOs, research centres and corporate entities. Several proposals have focused on ethics[18], often blurring the line between law and ethics, describing human rights and freedoms as ethical values with their 'ethicisation' and relativization.

This emphasis on the ethical dimension can entail the risk of extending to the field of data science an ethical imperialism whose effects are already known in biomedicine and the social sciences[19]. In this regard, previous experience in ethical assessment of scientific research suggests that careful consideration should be given to the distinction between ethical and legal values and the differences between ethical approaches[20]. Several documents providing guidelines on AI refer to ethics in a fairly broad and indefinite manner, with no clarification (or justification) of the ethical framework used[21].

Ethical responses to uncertainty in a rapidly changing technological and social environment may paradoxically become a new source of ambiguity. Discretionary and, in some cases, interest-based values risk weakening the legal framework or indirectly redefining it without following an appropriate procedure as required by the regulatory process[22].

Without underestimating the role of ethics in technology development,

---

[17] Mantelero & Esposito, *An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems,* cit.; Council of Europe, *Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications*, cit.

[18] A. Jobin, M. Ienca, E. Vayena, *The Global Landscape of AI Ethics Guidelines*, in *Nature Machine Intelligence*, 1, 2019, p. 389; T. Hagendorff, *The Ethics of AI Ethics: An Evaluation of Guidelines,* in *Minds and Machines* 30, 2020, p. 99.

[19] Z. M. Schrag, *Ethical Imperialism. Institutional Review Boards and the Social Sciences 1965-2009*, Baltimore, 2017.

[20] Hildebrandt, *The Issue of Bias. The Framing Powers of Machine Learning,* cit.

[21] Raab, *Information Privacy, Impact Assessment, and the Place of Ethics*, cit.; Independent High-Level Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 2019, available at <https://ec.europa.eu/digital-single-market/en/ news/ethics-guidelines-trustworthyai>, accessed on Mar. 2, 2020.

[22] See P. Nemitz, *Constitutional Democracy and Technology in the Age of Artificial Intelligence*, 378 *Phil. Transactions Royal Soc'y A* (2018), doi:10.1098/rsta.2018.0089.

these considerations suggest a more balanced integration of law and ethics in AI regulation, based on the emphasis on the role of human rights as the universal cornerstone of the future architecture of AI regulation. From a regulatory perspective, the main challenge is to contextualise the legal principles and provisions enshrined in international human rights instruments, drafted in a pre-AI era, within the current scenario where predictive policing tools, automated digital propaganda and other new AI-based applications are reshaping many aspects of our society and human relations.

Regulatory initiatives have been proposed in several countries[23], many of them referring explicitly to all or some human rights. However, these are often generic statements without a proper contextualisation of the rights and freedoms considered. Although it is relatively easy to agree on a general list of rights and freedoms that should underpin AI development, these lists do little to advance the regulatory process, since general principles, such as transparency or participation, can be interpreted in many different ways.

An effective contribution to the human rights debate in this field can therefore only come from a proper contextualisation of these guiding principles within the AI scenario. This means placing such rules, including the operational ones, in the context of the changes to society produced by AI and providing a more refined and specific formulation of the guiding principles with a view to possible future AI regulation.

This contextualisation of the guiding principles and rules can provide a more refined and elaborate formulation, taking into account the specific nature of AI products and services, and helping to better address the challenges arising from AI.

From a methodological perspective, an analysis of international legally binding instruments is the obligatory starting point in defining the existing legal framework, identifying its guiding values and verifying whether this framework and its principles properly address the issues raised by AI, with a view to preserving the harmonisation of the existing legal framework in the fields of democracy and justice.

The methodology adopted is therefore necessarily deductive, extracting

[23] J. Gesley, *Regulation of Artificial Intelligence in Selected Jurisdictions*, 2019, available at <https://www.loc.gov/law/help/artificial-intelligence/index.php>, accessed on Dec. 30, 2019; Council of Europe, *Graphical Visualisation of the Distribution of Strategic and Ethical Frameworks Relating to Artificial Intelligence*, 2020; See also K. Manheim, L. Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, 21 *Yale J. of L. & Technol.* 106 (2019).

the guiding principles from the variety of regulations concerning the fields in question. The theoretical basis of this approach relies on the assumption that the general principles provided by international human rights instruments should underpin all human activities, including AI-based innovation[24].

These guiding principles should be considered within the scenario of AI-driven transformation, which in many cases requires adaptation. They remain valid, but their implementation must be reconsidered in the light of the social and technical changes brought about by AI. This will deliver a more contextualised and granular application of these principles so that they can make a concrete contribution to the shape of future AI regulation.

Against this background, the following sections examine two critical areas of AI application: electronic democracy and digital justice. While in other areas, such as data protection and biomedicine, the specific nature of the sectors and recent soft-law regulatory initiatives[25] make it possible to draft some provisions for future AI regulation[26], in these two realms this is much more difficult. In addition, key principles that can be seen as guiding elements of future AI regulation, such as transparency and explainability[27], are open to varying interpretations and implementations, given the higher

[24] Council of Europe, *Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems*, cit.

[25] Council of Europe, Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data (Convention 108), *Guidelines on artificial intelligence and data protection*, 2019, available at <https://rm.coe.int/ guidelines-on-artificial-intelligence-and-data-protection/168091f9d8>, accessed on Feb. 20, 2020; CEPEJ, European Commission for the Efficiency of Justice, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*, 2018.

[26] See A. Mantelero, *Artificial Intelligence and Data Protection: Challenges and Possible Remedies. Report on Artificial Intelligence*, Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of personal data, 2019, available at <https://rm.coe.int/2018-lignes-directrices-sur-l-intelligence-artificielle-et-la-protecti/168098e1b7>, accessed on Feb. 20, 2020.

[27] A. D. Selbst, S. Barocas, *The Intuitive Appeal of Explainable Machines*, 87 *Fordham L. Rev*. 1085 (2018); M. Veale, R. Binns, *Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data*, cit.; N. Diakopoulps, *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*, 2013, available at <https://academiccommons.columbia.edu/ doi/10.7916/D8ZK5TW2>, accessed on Mar. 18, 2018; M. E. Kaminski, G. Malgieri, *Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR*, in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery*, 2020, doi: 10.1145/3351095.3372875.

political significance of both democracy and justice. The analysis therefore focuses on high-level principles and their contextualisation, resulting in a more limited elaboration of key guiding provisions.

2. *AI and Electronic Democracy*

Democracy covers an extremely wide array of societal and legal issues[28], most of them likely to be implemented with the support of ICT[29]. In this scenario, AI can play an important role in the present and future development of digital democracy in an information society.

The broad dimension of this topic makes it difficult to identify a single binding sector-specific legal instrument for reference. Several international instruments deal with democracy and its different aspects, starting with the UN Declaration of Human Rights and the International Covenant on Civil and Political Rights. Similarly, in the European context, key principles for democracy are present in several international sources.

Based on Article 25 ICCPR, we can identify two main areas of intervention related to electronic democracy: (i) participation[30] and good governance, and (ii) elections. Undoubtedly, it is difficult or impossible to draw a red line between these fields as they are interconnected in various ways. AI can have an impact on all of them: participation (e.g., citizens engagement, participation platforms), good governance (e.g., e-government, decision-making processes, smart cities), pre-electoral phase (e.g., financing, targeting and profiling, propaganda), elections (e.g., prediction of election results, e-voting), and the post-election period

[28] Council of Europe, Directorate General of Democracy, European Committee on Democracy and Governance. *The Compendium of the most relevant Council of Europe texts in the area of democracy*, 2016.

[29] Council of Europe, Directorate General of Democracy and Political Affairs and Directorate of Democratic Institutions, *Project 'Good Governance in the Information Society'*, CM(2009)9 Addendum 3, 2009; Council of Europe, *Additional Protocol to the European Charter of Local Self-Government on the right to participate in the affairs of a local authority*, 2009, art. 2.2.iii.

[30] A. Faye Jacobsen, *The Right to Public Participation. A Human Rights Law Update*, Issue Paper, 2013, available at <https://www.humanrights.dk/publications/ right-public-participation-human-rights-law-update>, accessed on Jan. 14, 2021; N. Maisley, *The International Right of Rights? Article 25(a) of the ICCPR as a Human Right to Take Part in International Law-Making*, in *Eur. J. Int'l L.*, vol. 28, 2017, p. 89.

(e.g., electoral dispute resolution).

As in any classification, this distinction is characterised by a margin of directionality. It is worth pointing here out that this is a functional classification based on different AI impacts, with no intention to provide a legal or political representation of democracy and its different key elements. The relationship between participation, good governance, and elections can therefore be considered from different angles and shaped in different ways, unifying certain areas or further subdividing them.

Participation is expressed both through taking part in the democratic debate and through the electoral process, but the way that AI tools interact with participation in these two cases differs and there are distinct international legal instruments specific to the electoral process.

2.1. *Participation and good governance*

The right to participate in public affairs (Article 25 Covenant) is based on a broad concept of public affairs[31], which includes public debate and dialogue between citizens and their representatives, with a close link to freedom of expression, assembly, and association[32]. In this respect, AI is relevant from two different perspectives: as a means to participation and as the subject of participatory decisions.

Considering AI as a means, technical and educational barriers can undermine the exercise of the right to participate. Participation tools based on AI should therefore consider the risks of under-representation and lack of transparency in participative processes (for example platforms for the drafting of bills). At the same time, AI is also the subject of participatory decisions, as they include decisions on the development of AI in general and its use in public affairs.

AI-based participative platforms (e.g., Consul, Citizenlab, Decidim[33]) can make a significant contribution to the democratic process, facilitating citizen interaction, prioritising of objectives, and collaborative approaches

[31] UN Human Rights Committee (HRC). *CCPR General Comment No. 25: The right to participate in public affairs, voting rights and the right of equal access to public service (Art. 25)*, CCPR/C/21/Rev.1/Add.7, 12 July 1996.

[32] UN Committee on Economic, Social and Cultural Rights (CESCR), *General Comment No. 1: Reporting by States Parties*, 27 July 1981, par. 5.

[33] Information on these platforms is available at <https://decidim.org/>; <https://consulproject.org/en/>; <https://www. citizenlab.co/>, accessed on Dec. 29, 2019.

in decision-making[34] on topics of general interests at different levels (neighbourhood, municipality, metropolitan area, region, country)[35].

Specific issues arise in relation to AI tools for democratic participation (including those for preventing and fighting corruption[36]), which are associated with the following four main areas: transparency, accountability, inclusiveness, and openness. In this regard, the general principles set out in international binding instruments have an important implementation in the Recommendation CM/Rec(2009)1 of the Committee of Ministers of the Council of Europe to member states on electronic democracy (e-democracy), which provides a basis for further elaboration of the guiding principles in the field of AI with regard to democracy.

Transparency is a requirement for the use of technological applications for democratic purposes[37]. This principle is common to other fields, such as healthcare[38], but is a context-based notion. While in healthcare transparency is closely related to self-determination, here it is not only a requirement for citizens' self-determination with respect to a technical tool but is also a component of the democratic participatory process[39]. Transparency no longer has an individual dimension but assumes a collective dimension as a guarantee of the democratic process.

In this context, the use of AI-based solutions for e-democracy must be transparent in respect of their logic and functioning (e.g., content selection in participatory platforms) providing clear, easily accessible, intelligible, and updated information about the AI tools used and their justification[40].

Moreover, the implementation of this notion of transparency should

[34] See also Council of Europe, *Guidelines for civil participation in political decision making*, CM(2017)83-final, 2017a.

[35] See also Council of Europe, *Recommendation CM/Rec(2009)2 on the evaluation, auditing and monitoring of participation and participation policies at local and regional level*, 2009.

[36] See United Nations, *Convention against Corruption*, 2003, art. 13.

[37] See Council of Europe, *Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy)*, 2009, par. 6.

[38] See Council of Europe, *Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine*, 1997

[39] See also Council of Europe, *Guidelines for civil participation in political decision making*, cit.

[40] See Council of Europe, *Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy)*, par. 6 and Appendix, par. P.57. See also Council of Europe, *Recommendation CM/Rec(2016)5 on Internet freedom*, 2016, Appendix, par. 2.1.3 and 3.2. On the importance of justification see M. Hildebrandt, *Primitives of Legal Protection in the Era of Data-Driven Platforms*, 2 *Georgetown L. Technol. Rev.* 252, 2018, pp. 271-3.

also consider the range of different users of these tools, adopting an accessible approach[41] from the early stages of the design of AI tools. This is to ensure effective transparency with regard to vulnerable and impaired groups, giving added value to accessibility in this context.

Transparency and accessibility are closely related to the nature of the architecture used to build AI systems. Open source and open standards can therefore contribute to democratic oversight of the most critical AI applications[42]. There are cases where openness is affected by limitations, due to the nature of the specific AI application (for example crime prevention). In these cases, auditability, as well as certification schemes, play a more important role than they already do in relation to AI systems in general[43].

In the context of AI applications to foster democratic participation, an important role can be also played by interoperability[44] as it facilitates integration between different services/platforms for e-democracy and at different geographical levels. This aspect is already relevant for e-democracy in general[45], and should therefore be extended to the design of AI-based systems.

Another key principle in e-democracy is accountability. In this regard, to be accountable, AI service providers and entities using AI-based solutions for e-democracy shall adopt forms of algorithm vigilance that promote the accountability of all relevant stakeholders by assessing and documenting the expected impacts on individuals and society in each phase of the AI system lifecycle on a continuous basis, to ensure compliance with human rights, the rule of law and democracy[46].

Finally, given the role of media in the context of democratic

[41] See also Council of Europe, *Recommendation CM/Rec(2018)4 on the participation of citizens in local public life*, 2018, Appendix, par. B.IV.

[42] See Council of Europe, *Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy)*, cit. par. 6 and Appendix, par. G.58 and P.54.

[43] It is worth to underline that auditing and certification schemes play an important role also in cases of open- source AI architecture, as this nature does not imply per se absence of bias or any other shortcomings. See also Council of Europe, *Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy)*, cit., Appendix, par. P. 55 and G.57.

[44] See also Council of Europe, *Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy)*, cit., Appendix par. P.56, 59 and 60.

[45] See also Council of Europe, *Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy)*, cit., par. 6.

[46] See also Council of Europe, *Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems*, cit.

participation[47], AI applications must not compromise the confidentiality and security of communications and protection of journalistic sources and whistle-blowers[48].

In addressing the different aspects of developing AI solutions for democratic participation, a first consideration is that a democratic approach is incompatible with a techno-determinist approach. AI solutions to address societal problems should therefore be the result of an inclusive process. Hence, legal values such as the protection of minorities, pluralism and diversity should be a necessary consideration in the development of these solutions.

From a democratic perspective, the first question we should ask is: do we really need an AI-based solution to a given problem as opposed to other options[49], considering the potential impact of AI on rights and freedoms? If the answer to this question is yes, the next step is to examine value- embedding in AI development[50].

The proposed AI solutions must be designed from a human rights-oriented perspective, ensuring full respect for human rights and fundamental freedoms, including the adoption of assessment tools and procedures for this purpose[51]. In the case of AI applications with a high impact on human rights and freedoms, such as electoral processes, legal compliance should be prior assessed. In addition, AI systems for public tasks should be auditable and, where not excluded by competing prevailing interests, audits should be publicly available.

Another important aspect to be considered is the public-private partnership that frequently characterises AI services for citizens[52], weighing

[47] See Council of Europe, *Recommendation CM/Rec(2016)4 on the protection of journalism and safety of journalists and other media actors*, 2016, Appendix par. 2; Council of Europe Parliamentary Assembly, *Resolution 2254 (2019)1. Media freedom as a condition for democratic elections*, 2019.

[48] See also Council of Europe Parliamentary Assembly, *Resolution 2300 (2019)1. Improving the protection of whistle-blowers all over Europe*, 2019; Council of Europe, *Recommendation CM/Rec(2014)7 on the protection of whistleblowers*, 2014.

[49] See also Council of Europe, *Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems*, cit.

[50] See also Council of Europe, *Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes*, 2019, par. 7.

[51] See Council of Europe, *Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy)*, cit., par. 5 and 6, and Appendix, par. G.67. See also A. Mantelero, *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*, 34 *Computer L. & Security Rev*. 754, 2018.

[52] J. Mikhaylov, M. Esteve, A. Campion, *Artificial Intelligence for the Public Sector:*

which is the best choice between in-house and third-party solutions, including the many different combinations of these two extremes. In this regard, when AI solutions are fully or partially developed by private companies, transparency of contracts and clear rules on access and use of citizens' data have a critical value in terms of democratic oversight.

Restrictions on access and use of citizens' data are not only relevant from a data protection perspective (principles of data minimisation and purpose limitation) but more generally with regard to the bulk of data generated by a community, which also includes non-personal data and aggregated data. This issue should be considered as a component of democracy in the digital environment, where the collective dimension of the digital resources generated by a community should entail forms of citizen control and oversight, as happens for the other resources of a territory/ community.

The considerations already expressed above on openness as a key element of democratic participation tools should be recalled here, given their impact on the design of AI systems. Furthermore, the design, development and deployment of these systems should also consider the adoption of an environmentally friendly and sustainable strategy[53].

Finally, it is worth noting that while AI-design is a key component of these systems, design is not neutral. Values can be embedded in technological artefacts[54], including AI systems. These values can be chosen intentionally and, in the context of e-democracy, this must be based on a democratic process. But they may also be unintentionally embedded into AI solutions, due to the cultural, social and gender composition of AI developer teams. For this reason, inclusiveness has an added value here, in terms of inclusion and diversity[55] in AI development.

The principles discussed for e-democracy can be repeated with regard to good governance[56]. This is the case with smart cities and

Opportunities and Challenges of Cross-Sector Collaboration, 376 *Phil. Trans. R. Soc. A*, 2018, doi: 10.1098/rsta.2017.0357.

[53] See also Council of Europe, *Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy)*, cit., appendix, par. P.58.

[54] See also P. P. Verbeek, *Understanding and Designing the Morality of Things*, Chicago-London, 2011, pp. 41-65.

[55] See Council of Europe, *Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems*, cit. Appendix, par. 3.5.

[56] See Council of Europe, *Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy)*, cit. Appendix, par. P.4; Council of Europe, *Recommendation CM/Rec(2004)15 on electronic governance ("e-governance")*, 2004; Committee of Ministers of the Council of Europe, *The 12 Principles of Good Governance enshrined in the*

sensor- based environmental management, where open, transparent and inclusive decision-making processes play a central role[57]. Similarly, the use of AI to supervise the activities of local authorities[58], for auditing and anticorruption purposes[59], should be based on openness (open source software), transparency and auditability.

More generally, AI can be used in government/citizen interaction to automate citizen' inquiries and information requests[60]. However, in these cases, it is important to guarantee the right to know we are interacting with a machine[61] and to have a human contact point. Moreover, access to public services must not depend on the provision of data that is unnecessary and not proportionate to the purpose.

Special attention should also be paid to the potential use of AI in human-machine interaction to implement nudging strategies[62]. Here, due to the complexity and obscurity of the technical solutions adopted, AI can increase the passive role of citizens and negatively affect the democratic decision-making process. Otherwise, an active approach based on conscious and active participation in community goals should be preferred and better managed by AI participation tools. Where adopted, nudging strategies should still follow an evidence-based approach.

Finally, the use of AI systems in governance tasks raises challenging questions about the relationship between human decision-makers and the

Strategy on Innovation and Good Governance at local level, 2008.

[57] Privacy International, *Smart Cities: Utopian Vision, Dystopian Reality*, cit.

[58] Council of Europe, *Recommendation CM/Rec(2019)3 on supervision of local authorities'activities*, 2019, Appendix, par. 4 and 9.

[59] See also P. Savaget, T. Chiarini, S. Evans, *Empowering Political Participation through Artificial Intelligence*, 46 *Science and Public Pol'y* 369, 2019.

[60] See H. Mehr, *Artificial Intelligence for Citizen Services and Government*, 2017, available at <https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_ services.pdf>, accessed on Mar. 15, 2021.

[61] See also Council of Europe, Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data (Convention 108), *Guidelines on artificial intelligence and data protection*, cit. par. 2.11.

[62] On the use of nudging in the smart city context, see S. Ranchordás, *Nudging Citizens through Technology in Smart Cities*, 34 *Int'l Rev. L. Computers & Technol.* 254, 2020; O. H. Gandy Jr., S. Nemorin, *Toward a Political Economy of Nudge: Smart City Variations*, 22 *Information, Communication & Soc'y* 2112, 2019. See generally C. R. Sunstein, *The Ethics of Nudging*, 32 *Yale J. Regulation* 412, 2015; C. R. Sunstein, *Why Nudge? The Politics of Libertarian Paternalism*, New Haven, 2015; R. H. Thaler, C. R. Sunstein, *Nudge. Improving Decisions about Health, Wealth, and Happiness*, New Haven, 2008; C. R. Sunstein, R. H. Thaler, *Libertarian Paternalism in Not an Oxymoron*, 70 *U. Chicago L. Rev*. 1159, 2003.

role of AI in the decision-making process[63]. These issues are more relevant with regard to the functions that have a high impact on individual rights and freedoms, as in the case of jurisdictional decisions[64].

2.2. *Elections*

The impact of AI on electoral processes is broad and concerns the pre-election, election, and post-election phases in different ways. However, an analysis focused on the stages of the electoral process does not adequately highlight the different ways in which AI solutions interact with it.

The influence of AI is therefore better represented by the following distinction: AI for the electoral process (e-voting, predictions of results, and electoral dispute resolution) and AI for electoral campaigns (micro-targeting and profiling, propaganda and fake news). While in the first area AI is mainly a technological improvement of an existing process, in the field of electoral campaigning AI-based profiling and propaganda raise new concerns that are only partially addressed by the existing legal framework. In addition, several documents have emphasised the active role of states in creating an enabling environment for freedom of expression[65].

As regards the technological implementation of e-democracy (e-voting, prediction of results, and electoral dispute resolution), some of the key principles mentioned with regard to democratic participation are also

[63] D. K. Citron, R. Calo, *The Automated Administrative State: A Crisis of Legitimacy*, 70 *Emory L. J.* 797, 2021.

[64] See Section 3.

[65] See Council of Europe, *Recommendation CM/Rec(2018)1 on media pluralism and transparency of media ownership*, 2018; United Nations (UN), Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE), Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on "Fake News," Disinformation and Propaganda*, 2017. See also Council of Europe, *Recommendation CM/Rec(2016)5 on Internet freedom*, cit. Appendix, par. 1.5, 2.1 and 3; European Commission for Democracy trough Law (Venice Commission), *Joint Report of the Venice Commission and of the Directorate of Information society and Actions Against Crime of the Directorate General of Human Rights and Rule of Law (DGI) on Digital Technologies and Elections*, 2019, par. 151.E; B. Bukovska, *Spotlight on Artificial Intelligence and Freedom of Expression 'SAIFE'*. Organization for Security and Co-operation in Europe, 2020. Available at <https://www.osce.org/files/f/documents/9/f/456319_0.pdf>, accessed on Aug. 11, 2020.

relevant here. Accessibility, transparency, openness, risk management and accountability (including the adoption of certification and auditing procedures) are fundamental elements of the technological solutions adopted in these stages of the electoral process[66].

As regards AI for campaigning (micro-targeting and profiling, propaganda and fake news), some of the issues raised concern the processing of personal data in general. The principles set out in Convention 108+ can therefore be applied and properly contextualised[67].

More specific and new responses are needed in the case of propaganda and disinformation[68]. Here the existing binding and non-binding instruments do not set specific provisions, given the novelty of the disinformation based on new forms of communication, such as social networks, which differ from traditional media[69] and often bypass the professional mediation of the journalists.

However, general principles, such as the principle of non-interference by public authorities on media activities to influence elections[70], can be extended to these new forms of propaganda and disinformation. Considering the use of AI to automate propaganda, future AI regulation should extend the scope of the general principles of non-interference to AI-based systems used to provide false, misleading and harmful information. In addition, to prevent such interference, states[71] and social media providers should adopt a by-design approach to increase their resilience to disinformation and propaganda.

---

[66] See Council of Europe, *Recommendation CM/Rec(2017)5 on standards for e-voting*, 2017, Appendix I, par. 1, 2, 32, and 35-40. See also Council of Europe, Directorate General of democracy and Political Affairs – Directorate of Democratic Institutions, 2011.

[67] See Council of Europe, *Recommendation CM/Rec(2010)13 on the protection of individuals with regard to automatic processing of personal data in the context of profiling*, 2010; Council of Europe, Consultative Committee of the Convention of the Protection of Individuals with Regard to Automatic Processing of Personal Data, 2019.

[68] See Manheim, Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, cit.; European Commission, Directorate-General for Communications Networks, Content and Technology, 2018.

[69] See also Council of Europe, *Recommendation CM/Rec(2011)7 on a new notion of media*, 2011.

[70] See Council of Europe, *Recommendation CM/Rec(2007)15 on measures concerning media coverage of election campaigns*, 2007, par. I.1.

[71] See also United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., *Joint Declaration on 'Fake News,' Disinformation and Propaganda*, cit. par. 2.c.

Similarly, the obligation to cover election campaigns in a fair, balanced, and impartial manner[72] should entail obligations for media and social media operators regarding the transparency of the logic of the algorithms used for content selection,[73] ensuring pluralism and diversity of voices[74], including critical ones[75].

Moreover, states and intermediaries should promote and facilitate access to tools to detect disinformation and non-human agents, as well as support independent research on the impact of disinformation and projects offering fact-checking services to users[76].

Given the important role played by advertising in disinformation and propaganda, the criteria used by AI-based solutions for political advertising should be transparent[77], auditable and provide equal conditions to all the political parties and candidates[78]. In addition, intermediaries should review their advertising models to ensure that they do not adversely affect the diversity of opinions and ideas[79].

---

[72] Council of Europe, *Recommendation CM/Rec(2007)15 on measures concerning media coverage of election campaigns*, cit., par. II.1

[73] See also United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., *Joint Declaration on 'Fake News,' Disinformation and Propaganda*, cit. Appendix, par. 2.1.3 and 2.3.5.

[74] See also EU Code of Practice on Disinformation, 2018, available at <https://digital-strategy. ec.europa.eu/en/policies/code-practice-disinformation>, accessed on Mar. 24, 2021.

[75] See also Council of Europe, *Recommendation CM/Rec(2016)4 on the protection of journalism and safety of journalists and other media actors*, 2016, Appendix, par. 15.

[76] See also United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., *Joint Declaration on 'Fake News,' Disinformation and Propaganda*, cit. par. 4.e; European Commission for Democracy trough Law (Venice Commission), *Joint Report of the Venice Commission and of the Directorate of Information society and Actions Against Crime of the Directorate General of Human Rights and Rule of Law (DGI) on Digital Technologies and Elections*, cit. par. 151. D.

[77] See also Council of Europe Parliamentary Assembly, *Resolution 2254 (2019)1. Media freedom as a condition for democratic elections*, 2019; European Commission for Democracy trough Law (Venice Commission), *Joint Report of the Venice Commission and of the Directorate of Information society and Actions Against Crime of the Directorate General of Human Rights and Rule of Law (DGI) on Digital Technologies and Elections*, par. 151. A and 151.B.

[78] See also Council of Europe, *Recommendation CM/Rec(2007)15 on measures concerning media coverage of election campaigns*, cit. par. II.5.

[79] See also United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., *Joint Declaration on 'Fake News,' Disinformation and Propaganda*, cit. par. 4.e.

3. *AI and Digital Justice*

As in the case of democracy, the field of justice is a broad domain and analysing the whole spectrum of the consequences of AI on justice would be too ambitious. In line with the scope of this contribution, this section sets out to describe the main challenges associated with the use of AI in digital justice and the principles which, based on international legally binding instruments, can contribute to its future regulation.

This analysis is facilitated by the European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment, adopted by the CEPEJ in 2019, which directly addresses the relationship between justice and AI. Although this non-binding instrument is classed as an ethical charter, to a large extent it concerns legal principles enshrined in international instruments.

Guiding principles for the development of AI in the field of digital justice can be derived from the following binding instruments: the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, the Convention for the Protection of Human Rights and Fundamental Freedoms, the International Convention on the Elimination of All Forms of Racial Discrimination, the Convention on the Elimination of All Forms of Discrimination against Women, and the Convention for the Protection of Human Rights and Fundamental Freedoms[80].

Given the range of types and purposes of operations in this field and the various professional figures and procedures involved, this section makes a functional distinction between two areas: (i) judicial decisions and alternative dispute resolutions (ADRs) and (ii) crime prevention/prediction. Before analysing and contextualising the key principles relating to these two areas, we should offer some general observation, which may also apply to the action of the public administration as a whole[81].

First of all, it is worth noting that – compared to human decisions, and more specifically judicial decisions – the logic behind AI systems does not resemble legal reasoning. Instead, they simply execute codes based on a data-centric and mathematical/statistical approach.

In addition, error rates for AI are close to, or lower than, the human

[80] See also, with regard to the EU area, the Charter of Fundamental Rights of the European Union.
[81] See Section 2.

brain in fields such as image labelling, but more complicated decision-making tasks have higher error rates. This is the case with legal reasoning in problem solving[82]. At the same time, while a misclassification of an image of a cat may have limited adverse effects, an error rate in legal decisions[83] has a high impact on rights and freedom of individuals.

It is worth pointing out that the difference between errors in human and machine decision-making has an important consequence in terms of scale: while human error affects only individual cases, poor design and bias in AI inevitably affect all people in the same or similar circumstances, with AI tools being applied to a whole series of cases. This may cause group discrimination, adversely affecting individuals belonging to different traditional and non-traditional categories[84].

Given the textual nature of legal documents, natural language processing (NLP) can play an important role in AI applications for the justice sphere[85]. This raises several critical issues surrounding commercial solutions developed with a focus on the English-speaking market, making them less effective in a legal environment that uses languages other than English[86]. Moreover, legal decisions are often characterised by implicit unexpressed reasoning, which may be amenable to expert systems, but not by language-based machine learning tools. Finally, the presence of general clauses requires a prior knowledge of the relevant legal interpretation and

[82] O. A. Osoba, W. Welser, *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*, 2017, p. 18, available at <https://www.rand.org/pubs/research_reports/RR1744.html>, accessed on May 20, 2020; See also M.L. Cummings et al., Chatham House Report, *Artificial Intelligence and International Affairs. Disruption Anticipated*, cit. p. 13.

[83] N. Aletras et al., *Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective*, in *PeerJ Computer Science* 2, e93, 2016, doi:10.7717/peerj-cs.93; F. Pasquale, G. Cashwell, *Prediction, Persuasion, and the Jurisprudence of Behaviourism*, 68 *U. Toronto L. J.* 63, 2018; M. Hildebrandt, *Algorithmic Regulation and the Rule of Law*, 376 *Phil. Trans. Royal Soc'y* 1, 2018.

[84] S. Wachter, *Affinity Profiling and Discrimination by Association in Online Behavioral Advertising*, 35 *Berkeley Technol. L.J.* 367 (2021); B. Mittelstadt, *From Individual to Group Privacy in Big Data Analytics*, 30 *Phil. & Technol.* 475, 2017; L. Taylor, L. Floridi, B. Van der Sloot, B. (eds), *Group Privacy: New Challenges of Data Technologies*, Dordrecht, 2017.

[85] But see M. Oswald, *Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power*, 376 *Phil. Trans. Royal Soc'y A*, 2018, doi: 10.1098/rsta.2017.0359; Pasquale, Cashwell, *Prediction, Persuasion, and the Jurisprudence of Behaviourism*, cit.

[86] See Council of Bars & Law Societies of Europe, *CCBE Considerations on the Legal Aspects of Artificial Intelligence*, 2020, p. 29.

continual updates which cannot be derived from text mining.

All these constraints suggest a careful and more critical adoption of AI in the field of justice than in other domains and, with regard to court decisions and ARDs, suggest following a distinction between cases characterised by routinely and fact-based evaluations and cases characterised by a significant margin for legal reasoning and discretion[87].

3.1. *Adrs and court decisions*

Several so-called Legal Tech AI products do not have a direct impact on the decision-making processes in courts or alternative dispute resolutions (ADRs), but rather facilitate content and knowledge management, organisational management, and performance measurement[88]. These applications include, for example, tools for contracts categorisation, detection of divergent or incompatible contractual clauses, e-discovery, drafting assistance, law provision retrieval, assisted compliance review. In addition, some applications can provide basic problem-solving functions based on standard questions and standardised situations (e.g., legal chatbots).

Although AI has an impact in such cases on legal practice and legal knowledge that raises various ethical issues[89], the potential adverse consequences for human rights, democracy and the rule of law are limited. To a large extent, they are related to inefficiencies or flaws of these systems.

In the case of content and knowledge management, including research and document analysis, these flaws can generate incomplete or inaccurate representations of facts or situations, but this affects the meta-products, the results of a research tool that need to be interpreted and adequately motivated when used in court. Liability rules, in the context of product liability, for instance, can address these issues.

In addition, bias (poor case selection, misclassification etc.) affecting standard text-based computer-assisted search tools for the analysis

[87] See the following Section on the distinction between codified justice and equitable justice.

[88] See CEPEJ, European Commission for the Efficiency of Justice, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*, cit. Appendix II.

[89] See also C. Nunez, *Artificial Intelligence and Legal Ethics: Whether AI Lawyers Can Make Ethical Decisions*, 20 *Tulane J. Technol. Intellectual Property* 189, 2017.

of legislation, case-law, and literature[90], can be countered by suitable education and training of legal professionals and the transparency of AI systems (that is the description of their logic, potential bias and limitations) can reduce the negative consequences.

Transparency should also characterise the use by courts of AI for legal research and document analysis. Judges must be transparent as to which decisions depend on AI and how the results provided by AI are used to contribute to the arguments, in line with the principles of fair trial and equality of arms[91].

Finally, transparency can play an important role with regard to legal chatbots based on AI, making users aware of their logic and the resources used (for example list of cases analysed). Full transparency should also include the sources used to train these algorithms and access to the database used to provide answers. Where these databases are private, third-party audits should be available to assess the quality of datasets and how potential biases have been addressed, including the risk of under- or over-representation of certain categories (non-discrimination).

Further critical issues affect AI applications designed to automate alternative dispute resolution or to support judicial decision. Here, the distinction between codified justice and equitable justice[92] suggests that AI should be circumscribed for decision-making purposes to cases characterised by routine and fact-based evaluations. This entails the importance to carry out further research on the classification of the different kind of decisional processes to identify those routinised applications of legal reasoning that can be demanded to AI, preserving in any case human overview that also guarantees legal creativity of decision-makers[93].

Regarding equitable justice, as the literature points out[94], its logic is

[90] See the notion of e-justice in Council of Europe, *Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy)*, cit. Appendix, par. 38.

[91] See also CEPEJ, European Commission for the Efficiency of Justice, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*, cit.

[92] R. M. Re, A. Solow-Nieederman, *Developing Artificially Intelligent Justice*, 22 *Stanford Technol. L. Rev*. 242, 2019, pp. 252-4.

[93] See also T. Clay (ed.), *L'arbitrage en ligne. Rapport du Club des Juristes*, 2019, available at <https://www.leclubdesjuristes.com/les-commissions/larbitrage-en-ligne/>. Accessed on May 30, 2020. In this regard, for example, a legal system that provides compensation for physical injuries on the basis of the effective patrimonial damages could be automatised, but it will not be able to reconsider the foundation of the legal reasoning and extend compensation to non-personal and existential damages.

[94] See R. M. Re, A. Solow-Nieederman, *Developing Artificially Intelligent Justice*, cit.

more complicated than the simple outcome of individual cases. Expressed and unexpressed values and considerations, both legal and non-legal, characterise the reasoning of the courts and are not replicable by the logic of AI. ML-based systems are not able to perform a legal reasoning. They extract inferences by identifying patterns in legal datasets, which is not the same as the elaboration of legal reasoning.

Considering the wider context of the social role of courts, jurisprudence is an evolving system, open to new societal and political issues. AI path-dependent tools could therefore stymie this evolutive process: the deductive and path-dependent nature of certain AI solutions can undermine the important role of human decision-makers in the evolution of law in practice and legal reasoning.

Moreover, at the individual level, path-dependency may also entail the risk of 'deterministic analyses'[95], prompting the resurgence of deterministic doctrines to the detriment of doctrines of individualisation of the sanction and with prejudice to the principle of rehabilitation and individualisation in sentencing.

In addition, in several cases, including ADR, both the mediation between the parties' demands and the analysis of the psychological component of human actions (fault, intentionality) require emotional intelligence that AI systems do not have.

These concerns are reflected in the existing legal framework provided by the international legal instruments. The Universal Declaration of Human Rights (Articles 7 and 10), the International Covenant on Civil and Political Rights (Article 14), the Convention for the Protection of Human Rights and Fundamental Freedoms (Article 6) and also the Charter of Fundamental Rights of the European Union (Article 47) stress the following key requirements with regard to the exercise of judicial power: equal treatment before the law, impartiality, independence and competency. AI tools do not possess these qualities, and this limits their contribution to the decision-making process as carried out by courts.

As stated by the European Commission for the Efficiency of Justice,

'the neutrality of algorithms is a myth, as their creators consciously or unintentionally transfer their own value systems into them'. Many cases of biases regarding AI applications confirm that these systems too often – albeit in many cases unintentionally – provide a partial representation of society and individual cases, which is not compatible with the principles

[95] See CEPEJ, European Commission for the Efficiency of Justice, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*, cit. p. 9.

of equal treatment before the law and non-discrimination[96]. Data quality and other forms of quality assessment (impact assessment, audits, etc.) can reduce this risk but, given the degree of potentially affected interests in the event of biased decisions, the risks remain high in the case of equitable justice and seem disproportionate to the benefits largely in terms of efficiency for the justice system[97].

Further concerns affect the principles of fair trial and of equality of arms[98], when court decisions are based on the results of proprietary algorithms whose training data and structure are not publicly available[99]. A broad notion of transparency might address these issues in relation to the use of AI in judicial decisions, but the transparency of AI – a challenging goal in itself – cannot address the other structural and functional objections cited above.

In addition, data scientists can shape AI tools in different ways in the design and training phases, so that were AI tools to become an obligatory part of the decision-making process, governments selecting the tools to be used by the courts could potentially indirectly interfere with the independence of the judges.

This risk is not eliminated by the fact that the judge remains free to disregard AI decisions, providing a specific motivation. Although human oversight is an important element[100], its effective impact may be undermined by the psychological or utilitarian (cost-efficient) propensity of the human decision-maker to take advantage of the solution provided by AI[101].

---

[96] *Id.*

[97] See Council of Europe, *Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems*, cit. Appendix, par. 11. See also F. Pasquale, G. Cashwell, *Prediction, Persuasion, and the Jurisprudence of Behaviourism*, cit.

[98] See also CEPEJ, European Commission for the Efficiency of Justice, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*, cit. Appendix I, par. 138.

[99] *Id.* at par. 131.

[100] M. Zalnieriute, L. Bennett Moses, G. Williams, *The Rule of Law and Automation of Government Decision-Making*, 82 *Modern L. Rev*. 425, 2019. In the case of administrative decisions, this propensity may be reinforced by the threat of potential sanctions for taking a decision that ignores results produced by analytics; Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data (Convention 108), *Guidelines on artificial intelligence and data protection*, cit. par. 3.4.

[101] See also D. K. Citron, R. Calo, *The Automated Administrative State: A Crisis of Legitimacy*, cit.; Mantelero, *Artificial Intelligence and Data Protection: Challenges and Possible Remedies*. Report on Artificial Intelligence. Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of

3.2. *Crime Prevention*

The complexity of crime detection and prevention has stimulated research in AI applications to facilitate human activities. In recent years, several solutions[102] and a growing literature have been developed in the field of predictive policing, which is a proactive data-driven approach to crime prevention. Essentially, the available solutions pursue two different goals: to predict where and when crimes might occur or to predict who might commit a crime[103].

These two purposes have a distinct potential impact on human rights and freedom, which is more pronounced when AI is used for individual predictions. However, in both cases, we can repeat here the considerations about the general challenges related to AI (obscurity, intellectual property rights, large-scale data collection[104], etc.) discussed in the previous sections and partially addressed by transparency, data quality, data protection, auditing and the other measures[105]. It is worth noting that the role of transparency in the judicial context could be limited so as not to frustrate the deterrent effect of these tools[106]. Full transparency could therefore be replaced by auditing and oversight by independent authorities.

Leaving aside the organisational aspects regarding the limitation of police officers' self-determination in the performance of their duties, the main issues with regard to the use of AI to predict crime on geographic

personal data, cit. R. Brauneis, E. P. Goodman, *Algorithmic Transparency for the Smart City*, cit. p. 127.

[102] A. Završnik, *Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings*, in *European Journal of Criminology* 1, 2019, doi:10.1177/1477370819876762; European Union Agency for Fundamental Rights, *#BigData: Discrimination in Data-Supported Decision Making*, 2018, pp. 98-100; O. A. Osoba, W. Welser, *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*, cit.

[103] For a taxonomy of predictive methods, see W.L. Perry et al., *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, 2013, available at <https://www.rand.org/pubs/research_ reports/RR233.html>, accessed on Mar. 30, 2020.

[104] See also Council of Europe, *Recommendation CM/Rec(2011)7 on a new notion of media*, cit. Appendix, par. 42.

[105] See also R. Richardson, J. M. Schultz, K. Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 *N. Y.U. L. Rev*. 42, 2019.

[106] See also M. Oswald, *Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power*, cit.; L. Barrett, *Reasonably Suspicious Algorithms: Predictive Policing at the United States Border*, 41 *N.Y. U. Rev. L. & Soc. Change* 327, 2017, pp. 361-2.

and temporal basis concern the impact of these tools on the right to non-discrimination[107]. Self-fulfilling bias, community bias[108] and historical bias[109] can produce forms of stigmatisation for certain groups and the areas where they typically live.

Where data analysis is used to classify crimes and infer evidence on criminal networks, proprietary solutions raise issues in terms of respect for the principles of fair trial and of equality of arms with regard to the collection and use of evidence. Moreover, if the daily operations of policy departments are guided by predictive software, this raises a problem of accountability of the strategies adopted, as they are partially determined by software and hence by software developer companies, rather than the police.

A sharper conflict with human rights arises in the area of predictive policing tools that use profiling to support individual forecasting. Quite apart from the question of data processing and profiling[110], these solutions can also adversely affect the principle of presumption of innocence, procedural fairness, and the right to non-discrimination[111].

While non-discrimination issues could be partially addressed, the remaining conflicts seem to be more difficult to resolve. From a human rights standpoint and in terms of proportionality (including the right to respect for private and family life)[112], the risk of prejudice to these principles seems high and not adequately countered by the evidence of benefits for individual and collective rights and freedoms[113]. In the light of future AI regulation, this should urge careful consideration of these issues,

[107] See European Union Agency for Fundamental Rights, *#BigData: Discrimination in Data-Supported Decision Making*, cit. p. 10.

[108] Barrett, *Reasonably Suspicious Algorithms: Predictive Policing at the United States Border*, cit. pp. 358-9.

[109] Zalnieriute, Bennett Moses, Williams, *The Rule of Law and Automation of Government Decision-Making*, cit.

[110] See O. Lynskey, *Criminal Justice Profiling and EU Data Protection Law: Precarious Protection from Predictive Policing*, 15 *Int'l J. L. Context* 162, 2019; A. Mantelero, G. Vaciago, *Data Protection in a Big Data Society. Ideas for a Future Regulation*, in *Digital Investigation* 15, 2015, p. 104.

[111] See Council of Europe, *Recommendation CM/Rec(2001)10 on the European Code of Police Ethics*, 2001.

[112] See R. van Brakel, P. De Hert, *Policing, Surveillance and Law in a Pre-Crime Society: Understanding the Consequences of Technology Based Strategies*, in *Cahiers Politiestudies*, 3, 2011, p. 183.

[113] See A. Meijer, M. Wessels, *Predictive Policing: Review of Benefits and Drawbacks*, 42 *Int'l J, Pub. Admin*. 1031, 2019.

taking into account the distinction between the technical possibilities of AI solutions and their concrete benefits in safeguarding and enhancing human rights and freedoms.

Finally, from a wider and comprehensive human rights perspective, the focus on crime by data-driven AI tools drives a short-term factual approach that underrates the social issues that are often crime-related and require long-term social strategies involving the effective enhancement of individual and social rights and freedoms[114].

4. *Conclusions*

The latest wave of AI development is having a growing transformative impact on society and raises new questions in several fields, from predictive medicine and media content moderation to the quantified self and judicial systems.

With a view to preserving the harmonisation of the existing legal framework in the field of human rights, this article sets out to contribute to the debate on future AI regulation by building on existing binding instruments, contextualising their principles and providing key regulatory guidance in the fields of electronic democracy and digital justice.

This approach is based on the assumption that all human activities, including innovation through AI, should be underpinned by the general international principles on human rights. Moreover, only the human rights framework can provide a universal reference for the regulation of AI, while other yardsticks (for example ethics) do not have the same global dimension, are more context-dependent and characterised by a variety of theoretical approaches.

The findings of this analysis show that a limited number of cases do share common principles (for example individual self-determination, non-discrimination, human oversight). This is due to several factors.

First, some principles are sector specific. This is the case, for instance, with the independence of judges or the principles of fair trial and equality of arms, which concern justice alone[115].

[114] See D. P. Rosenbaum, *The limits of hot spots policing*, in D. Weisburd, A. A. Braga (eds.), *Police innovation: contrasting perspectives*, 2006, pp. 245-66.

[115] See also the principles of equitable access and of beneficence in health sector, or the principles of non-interference by public authorities in the media to influence elections and the obligation to treat all political parties and candidates equally in electoral advertising.

Second, some guiding principles are shared by different areas, but with different nuances in each context. This is true for transparency, which is often regarded as pivotal in AI regulation, but takes on different meanings in different regulatory contexts.

Transparency, as a means to control the power over data in the hands of public and private entities, is crucial with regard to AI applications for democratic participation and good governance. In the context of justice, transparency has a more complex significance, being vital to safeguard fundamental rights and freedoms (e.g., use of AI in the courts), but also requiring limitation to avoid prejudicing competing interests (e.g., crime detection and prevention in predictive policing).

We can therefore conclude that transparency is a guiding principle, but we must go beyond a mere claim for transparency as a key principle for AI regulation. As with other key principles (such as participation, inclusion, democratic oversight, and openness), a proper contextualisation is needed, with provisions that take into account the different contexts in which they operate.

Third, some principles are different, but belong to the same conceptual area, assuming various nuances in the different contexts. This is the case with accountability and guiding principles on risk management in general. Here the level of detail and related requirements can be more or less elaborate. While, for instance, in the field of data protection there are several provisions implementing these principles with a significant degree of detail[116], in the case of democracy and justice these principles are less developed in data-intensive applications such as AI.

Finally, there are certain components of an AI regulatory strategy that are not principles, but operational approaches and solutions, common to the different areas though requiring context-based development. This is the case with the important role played by education and training.

Such considerations suggest only partial harmonisation is achievable. The framework of future international AI regulation should therefore be based on a legally binding instrument that includes both general provisions– focusing on common principles and operational solutions – and more specific and sectoral provisions, covering those principles that are relevant only in a given field or cases where the same principle is contextualised differently in the different fields.

The analysis carried out in the previous sections has also confirmed

[116] See Council of Europe, *Recommendation CM/Rec(2018)1 on media pluralism and transparency of media ownership*, cit.

that the existing framework based on human rights can provide an appropriate and common context for the development of more specific binding instruments to regulate AI, in line with the principles enshrined in the international legal instruments and capable of effectively addressing the issues raised by AI.

With a view to future regulation of AI, this study does not rule out a number of gaps, largely due to the fact that in broad areas, such as democracy and justice, differing options and interpretations are available, depending on the political and societal vision of the future relationship between humans and machines. Further investigation in the field of human rights and AI, as well as the ongoing debate at international and regional level, will contribute to bridging these gaps.

Alessandro Mantelero, Maria Samantha Esposito

An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems

Abstract: Different approaches have been adopted in addressing the challenges of Artificial Intelligence (AI), some centred on personal data and others on ethics, respectively narrowing and broadening the scope of AI regulation. This contribution aims to demonstrate that a third way is possible, starting from the acknowledgement of the role that human rights can play in regulating the impact of data-intensive systems. The focus on human rights is neither a paradigm shift nor a mere theoretical exercise. Through the analysis of more than 700 decisions and documents of the data protection authorities of six countries, we show that human rights already underpin the decisions in the field of data use.
Based on empirical analysis of this evidence, this work presents a methodology and a model for a Human Rights Impact Assessment (HRIA). The methodology and related assessment model are focused on AI applications, whose nature and scale require a proper contextualisation of HRIA methodology. Moreover, the proposed models provide a more measurable approach to risk assessment which is consistent with the regulatory proposals centred on risk thresholds. The proposed methodology is tested in concrete case-studies to prove its feasibility and effectiveness. The overall goal is to respond to the growing interest in HRIA, moving from a mere theoretical debate to a concrete and context-specific implementation in the field of data-intensive applications based on AI.

1. *Introduction*

The debate that has characterised the last few years on data and Artificial Intelligence (AI) represents an interesting arena in which to consider the theoretical evolution of the future approach in addressing the challenges posed by AI to human rights. This debate has been marked by an emphasis

* A. Mantelero coordinated this study and is author of all sections except 4.2. M.S. Esposito is author of section 4.2 and carried out the research described in that section. The authors would like to acknowledge the helpful feedback on the model design and multicriteria analysis received from Maria Franca Norese, Associate Professor of Operations Research at the Polytechnic University of Turin. We are also grateful to the anonymous reviewers for their helpful feedback.
** This article was published in *Computer Law and Security Review*, Volume 41, July 2021, https://doi.org/10.1016/j.clsr.2021.105561.

on the ethical dimension of the use of algorithms[1] and, in the legal domain, by a focus on potential bias and protection from dis- crimination.[2]

However, the emerging AI-driven society shows a variety of potential impacts on individual and collective rights and freedoms suggesting, on the one hand, a reaffirmation of the central role of the legal instruments, not replaceable by ethical guidelines, and, on the other hand, the need for a more com- prehensive analysis of the rights and freedoms concerned.

This contribution, after some considerations on the inter- play between ethical and legal approaches in the debate on AI regulation, focuses on the analysis of the impact of AI on human rights and the development and application of an AI- specific impact assessment model.

Empirical evidence on how data-intensive systems may affect rights and freedoms is drawn from the European context which is characterised by a long-standing and persistent focus on human rights–partly through the role of the European Court of Human Rights–and a theoretical approach that has intertwined data processing and human rights since the early data protection regulations.[3]

The 'case study' of Europe, which provides extensive jurisprudence on data processing developed by the data protection authorities, is used to achieve a broader global perspective in dealing with AI and human rights. Creating a methodological approach to impact assessment built

[1] See L. Floridi et al., *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds & Machine*, 2018, DOI: 10.1007/s11023-018-9482-5, accessed on Nov. 30, 2020; B. D. Mittelstadt et al., *The ethics of algorithms: Mapping the debate*, 3 *Big Data & Soc'y* 1–21, 2016.

[2] See S. Wachter, B. Mittelstadt, and C. Russell, *Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law*, in *West Virginia L. Rev*. (2021), forthcoming <https://papers.ssrn.com/abstract=3792772>, accessed on Feb. 27, 2021; Algorithm Watch, *Automating Society Report*, 2020, <https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report- 2020.pdf>, accessed on Jan. 23, 2021; S. Myers West, M. Whittaker & K. Crawford, *Discriminating Systems*, 2019, p. 33, <https://ainowinstitute.org/discriminatingsystems.pdf >, accessed on June 13, 2020; F.J. Zuiderveen Borgesius, *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, 24 Int'l J. Hum. Rights 1572-1593, 2020; M. Mann and T. Matzner, *Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination*, in *Big Data & Soc'y*, vol. 6, 2019, DOI: 10.1177/2053951719895805.

[3] See e.g., L. A. Bygrave, *Data Privacy Law: An International Perspective*, Oxford, 2014; G. Gonzalez Fuster, *The Emergence of Personal Data Protection as a Fundamental Right of the EU* (Springer 2014); V. Mayer-Schönberger, *Generational Development of Data Protection in Europe*, in Philip E. Agre and Marc Rotenberg (eds.) *Technology and Privacy: The New Landscape*, 1997, 219-241.

on international legal instruments, we provide a model that can be more easily used in different legal cultures, grounded on the universal nature of the rights and freedoms in question.[4]

The next two sections of this work highlight the need to turn our gaze to law after the focus on data ethics in recent years, suggesting a complementary role for ethics and law and stressing that only human rights can provide a uniform reference for regulating AI in different cultural, ethical and legal contexts.

Section 4 investigates how human rights can be impacted by data intensive systems, adopting an empirical evidence- based approach rather than a theoretical one. Instead of creating fictitious cases on human rights and AI, we conduct an empirical analysis of decided cases, identifying the interplay between the use of data and human rights. The European con- text was chosen as the case study, given the existence of a general regulation on data protection – not only the GDPR but also in the preceding decades, at national level and, after 1995, at EU level – with the availability of extensive case law and related records of practice.

This section does not take a European-centred perspective in addressing AI issues, but provides empirical evidence on situations where

[4] Referring to this universal nature, we are aware of the underlying tensions that characterise it, the process of contextualisation of these rights and freedoms (appropriation, colonisation, vernacularisation, etc.) and the theoretical debate on universalism and cultural relativism in human rights. See P. Levitt and S. Merry, *Vernacularization on the Ground: Local Uses of Global Women's Rights in Peru, China, India and the United States*, in *Global Networks*, vol. 9, 441-461, 2009; S. Benhabib, *The Legitimacy of Human Rights*, 137 *Daedalus* 94-104, 2008; S. Engle Merry, *Human rights and gender violence: translating international law into local justice*, Chicago, 2006. See also D. M Goldstein, *Human Rights as Culprit, Human Rights as Victim: Rights and Security in the State of Exception*, in M. Goodale and S. Engle Merry (eds.), *The Practice of Human Rights: Tracking Law between the Global and the Local*, Cambridge, 2007, pp. 49-77; L. Leve, *'Secularism Is a Human Right!': Double- Binds of Buddhism, Democracy, and Identity in Nepal*, *ibid*, pp. 78-114; T. Risse and S. C. Ropp, *International Human Rights Norms and Domestic Change: Conclusions*, in K. Sikkink, S. C. Ropp and T. Risse (eds.), *The Power of Human Rights: International Norms and Domestic Change*, Cambridge, 1999, pp. 234-278; D. O'sullivan, *The History of Human Rights across the Regions: Universalism vs Cultural Relativism*, 2 *Int'l J. Human Rights* 22-48, 1998. However, from a policy and regulatory perspective, we believe that the human rights framework, including its nuances, can provide a more widely applicable common framework than other context-specific proposals on the regulation of the impact of AI. Furthermore, the proposed methodology includes in its planning section (see Section 5.1) the analysis of the human rights background, with a contextualisation based on local jurisprudence and laws, as well as the identification and engagement of potential stakeholders who can contribute to a more context-specific characterisation of the human rights framework.

a clash between human rights and data intensive systems occurs. The goal is not to analyse the response of the European legal framework but to extract a list of potentially impacted rights and freedoms based on concrete evidence as opposed to theoretical cases and hypothesis. This evidence-based approach underpins the methodology of our analysis and the proposed assessment model.

This exercise also brings out the limitations of data protection regulations in addressing AI issues and the need for a more tailored approach, rather than broad notions such as fairness, to enlarge the scope of data protection regulation and encompass AI application and related matters.

Having defined the theoretical framework and gathered the empirical evidence, Sections 5 and 6 present the key product of the research, the development and testing of an AI-focused human rights impact assessment model. The structure and components of the model are described in Section 5, while Section 6 discusses its concrete application in two cases with very different impacts in terms of scale. The cases chosen, one global and another limited to Canada, refer to two different types of AI applications, a smart device and a smart city plan. Since both projects have now been concluded, contrafactual analysis of the available documentation was able to test the model and show its results and effects in those contexts.

The last section provides some concluding remarks and points out the potential benefit of adopting the proposed assessment model in terms of legal compliance, risk management and human rights-orientated development and deployment of AI.

2. *The debate on AI regulation*

While data processing regulation has been focused for decades on the law, including the interplay between data use and human rights, in recent years the debate on AI and the use of data-intensive systems has rapidly changed its trajectory, from law to ethics[5]. This is evident not only in the literature[6], but also in the political and institutional debate[7]. In this

[5] See C. D. Raab, *Information Privacy, Impact Assessment, and the Place of Ethics*, in *Computer L. & Security Rev.*, vol. 37, 2020, par. 3, DOI:10.1016/j.clsr.2020.105404.

[6] See e.g., L. Floridi and M. Taddeo, *What is data ethics?*, 374 *Phil. Trans. R. Soc'y A.*, 2016, DOI: 10.1098/rsta.2016.0360.

[7] In the context of the legal debate on computer law, at the be- ginning of this decade

regard, an important turning point was the EDPS initiative on digital ethics[8] which led to the creation of the Ethics Advisory Group[9].

As regards the debate on data ethics, it is interesting to con- sider its origins. We can identify three different and chrono- logically consecutive stages: the academic debate, institutional initiatives, and the proliferation of AI ethical codes[10]. These contributions to the debate are different and have given voice to different underlying interests.

The academic debate on the ethics of machines is part of the broader and older reflection on ethics and technology. It is rooted in known and framed theoretical models, mainly in the philosophical domain, and has a methodological maturity. In contrast, the institutional initiatives are more recent, have a non-academic nature and aim at moving the regulatory debate forward, including ethics in the sphere of data protection. The main reason for this emphasis on ethics in recent years has been the growing concern in society about the use of data and new data-intensive applications, such as Big Data[11] and, more recently, AI.

Although similar paths are known in other fields, the shift from the theoretical analysis to the political arena represents a major change. The political attention to these issues has necessarily reduced the level of analysis, ethics being seen as an issue to be flagged rather than developing a full-

only few authors focused on ethical impact of IT, see e.g., D. Wright, *A framework for the ethical impact assessment of information technology*, 13 *Ethics Inf. Technol*. 199–226, 2010. Although the reflection on ethics and technology is not new in itself, it has become deeper in the field data use where new technology development in the information society has shown its impact on society. See also P. P. Verbeek, *Moralizing Technology. Understanding and Designing the Morality of Things,* Chicago, 2011; S. Spiekermann, *Ethical IT Innovation: A Value-Based System Design Approach*, Chicago, 2016; J. Bohn et al., *Social, Economic, and Ethical Implications of Ambient Intelligence and Ubiquitous Computing*, in W. Weber, J. M. Rabaey and E. Aarts (eds.), *Ambient Intelligence*, Cham, 2005, pp. 5, 19-29.

[8] See EDPS, *Opinion 4/2015. Towards a new digital ethics: Data, dignity and technology*, Sept. 11, 2015.

[9] See EDPS, *Decision of 3 December 2015 establishing an external advisory group on the ethical dimensions of data protection ('the Ethics Advisory Group')*, 2016/C 33/01 OJEU.

[10] See M. Ienca and E. Vayena, *AI Ethics Guidelines: European and Global Perspectives*, in Council of Europe, *Towards regulation of AI systems. Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law* (Council of Europe, DGI (2020)16), pp. 42-64.

[11] See also Council of Europe, Consultative Committee of the Convention 108 (T-PD), *Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data*, Strasbourg, 2017, T-PD(2017)01.

blown strategy for ethically-orientated solutions. In a nutshell, the message of regulatory bodies to the technology environment was this: law is no longer enough, you should also consider ethics.

This remarkable step forward in considering the challenges of new paradigms had the implicit limitation of a more general and basic ethical framework, compared to the academic debate. In some cases, only general references to the need to consider ethical issues has been added to AI strategy documents, leaving the task of further investigation to the recipients of these documents. At other times, as in the case of the EDPS, a more ambitious goal of providing ethical guidance was pursued.

Methodologically, the latter goal has often been pursued by delegating the definition of guidelines to committees of experts, including some forms of wider consultation. As in the tradition of expert committees, a key element of this process is the selection of experts.

These committees were not only composed of ethicists or legal scholars but had a different or broader composition defined by the appointing bodies.[12] Their heterogeneous nature made them more similar to multi-stakeholder groups.

Another important element of these groups advising policymakers concerns their internal procedures: the actual amount of time given to their members to deliberate, the internal distribution of assigned tasks (in larger groups this might involve several sub-committees with segmentation of the analysis and interaction between sub-groups), and the selection of the rapporteurs. These are all elements that have an influence in framing the discussion and its results.

All these considerations clearly show the differences between the initial academic debate on ethics and the same debate as framed in the context of institutional initiatives. Moreover, this difference concerns not only structure and procedures, but also outcomes. The documents produced by the experts appointed by policymakers are often minimalist in terms of theoretical framework and focus mainly on the policy message concerning

[12] This is the case, for example, of the Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission, which brought together 52 experts, the majority (27) from industry and the rest from academia (15, including 3 with a legal background and 3 with an ethical background), civil society (6) and governmental or EU bodies (4). See also *Laying down the Law on AI: Ethics Done, Now the EU Must Focus on Human Rights*, Access Now, April 8, 2019, <https://www.accessnow.org/laying-down-the-law-on-ai-ethics-done-now-the-eu-must-focus-on-human-rights/>, accessed on April 7, 2021; M. Veale, *A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence*, in *European J. Risk Regulation*, vol. 11, 2020, 1-10. doi:10.1017/err.2019.65.

the relevance of the ethical dimension.

The variety of the ethical approaches, the lack of clear indications on the frame of reference or the reasons for preferring a certain ethical framework make it difficult to understand the key choices on the proposed ethical guidelines[13]. Moreover, the local perspective of the authors of these documents, in line with the context-dependant nature of ethical values, undermines the ambition to provide global standards or, where certain values are claimed to have general relevance, may betray a risk of ethical colonialism.

3. *Framing the ethical and the human rights-based approaches*

From the outset, the debate on data ethics has been characterised by an improper overlap between ethics and law, in particular with regard to human rights. In this sense, it has been suggested that ethical challenges should be addressed by "fostering the development and applications of data science while ensuring the respect of human rights and of the values shaping open, pluralistic and tolerant information societies".[14] We can summarise this approach as 'ethics first': ethics plays a central role in technology regulation because it is the root of any regulatory approach, the pre-legal humus that is more important than ever where existing rules do not address or only partially address technological challenges.

Another argument in favour of the central role of ethics comes out of what we might call the 'ethics after' approach[15].

In the concrete application of human rights we necessarily have to balance competing interests. This balance test is not based on the rights themselves but on the underlying ethical values, meaning that the human rights framework is largely incomplete without ethics.

Both these approaches are only partially correct. It is true that human rights have their roots in ethics. There is an extensive literature on the relationship between ethics and law, which over the years has been described by various authors as identification, separation, complementation, and

[13] See also Ienca and Vayena, *AI Ethics Guidelines: European and Global Perspectives*, cit.

[14] Floridi and Taddeo, *What is data ethics?*, cit.

[15] See Cansu Canca, *AI & Global Governance: Human Rights and AI Ethics—Why Ethics Cannot Be Replaced by the UDHR*, 2019, UN University Centre for Policy Research, <https://cpr.unu.edu/ai-global-governance-human-rights-and-ai-ethics- why-ethics-cannot-be-replaced-by-the-udhr.html>, accessed on April 30, 2020.

interweavement[16]. Similarly, the influence of ethical values and more in general of societal issues in court decisions and balancing tests is known and has been investigated by various disciplines, including sociology, law & economics, and psychology.

Here the point is not to cut off the ethical roots, but to recognise that rights and freedoms flourish on the basis of the shape given them by law provisions and case law. There is no conflict between ethical values and human rights, but the latter represent a specific crystallisation of these values that are circumscribed and contextualised by legal provisions and judicial decisions.

This reflection may lead to a broader discussion of the role of ethics in the legal realm, but this study takes a more pragmatic and concrete approach by reframing the interplay be- tween these two domains within the context of AI and focusing on the regulatory consequences of adopting an approach based on ethics rather than human rights.

The main question should be formulated as follows: what are the consequences of framing the regulatory debate around ethical issues? Four different consequences can be identified: (1) uncertainty, (2) heterogeneity, (3) context dependence, and (4) risks of overestimation and of a 'transplant' of ethical values.

As far as uncertainty is concerned, this is due to the im- proper overlap between law and ethics in ethical guidelines[17].

While it is true that these two realms are intertwined in various ways, from a regulatory perspective the distinction between ethical imperatives and binding provisions is important. Taking a pragmatic approach, relying on a framework of general ethical values (such as beneficence, non-maleficence, etc.), on codes of conduct and ethical boards is not the same as adopting technological solutions on the basis of binding rules.

This difference is not only due to the different levels of enforcement, but also to the more fundamental problem of uncertainty about specific requirements. Stating that "while many legal obligations reflect ethical principles, adherence to ethical principles goes beyond formal compliance with existing laws"[18] is not enough to clarify the added value

[16] See A. Cortina, *Legislation, Law and Ethics*, 3 Ethical Theory and Moral Practice 3-7, 2000.

[17] See Raab, *Information Privacy, Impact Assessment, and the Place of Ethics*, cit. ("The products in the 'turn' to ethics often look more like 'data protection-plus' than a different kind of encounter with some of the age-old issues and concepts in the study and practice of ethics, and how to embed them in practice").

[18] Independent High-Level Expert Group on Artificial Intelligence set up by

of the proposed ethical principles and their concrete additional regulatory impact[19].

Given the different levels of binding nature and enforcement, shifting the focus from law to ethics and reformulating legal requirements as ethical duties open the door to de-regulation and self-regulation. Rather than binding rules, business can therefore benefit from a more flexible framework based on corporate codes of ethics[20].

This generates uncertainty in the regulatory framework. When ethical guidelines refer to human oversight, safety, privacy, data governance, transparency, diversity, non- discrimination, fairness, and accountability as key principles, they largely refer to legal principles that already have their contextualisation in specific provisions in different fields. The added value of a new generalisation of these legal principles and their concrete applications is unclear and potentially dangerous: product safety and data governance, for instance, should not be perceived as mere ethical duties, but companies need to be aware of their binding nature and related legal consequences.

Moreover, ethical principles are characterised by an inherent heterogeneity due to the different ethical positions taken by philosophers over the centuries. Virtue ethics, deontological or consequentialist approaches[21] can lead to different conclusions on ethical issues. AI developers or manufacturers might opt for different ethical paradigms (note that those mentioned are limited to the Western tradition only), making harmonised regulation difficult.

Similarly, the context-dependence of ethical values entails their variability depending on the social context or social groups considered, as

---

the European Commission (hereinafter AI HLEG), *Ethics Guidelines for Trustworthy AI*, 2019, p. 12.

[19] See for example the principle of respect for human autonomy, AI HLEG, *Ethics Guidelines for Trustworthy AI*, cit., which is specified in self-determination, democratic process and human oversight. These are general categories that have an ethical origin but already have a concrete legal implementation – from personality rights to product safety – which can provide a better and more detailed framework for the elaboration of contextualised provisions concerning AI.

[20] See B. Wagner, *Ethics as an escape from regulation: From "ethics-washing" to ethics-shopping?*, in Emre Bayamlioglu et al. (eds.), *Being Profiled*, Amsterdam, 2018, pp. 84-89; L. Taylor and L. Dencik, *Constructing Commercial Data Ethics*, in *Technology and Regulation*, 2020, pp. 1-10, DOI: 10.26116/techreg.2020.001; Ienca and Vayena, *AI Ethics Guidelines: European and Global Perspectives*, cit.

[21] See Verbeek, *Understanding and Designing the Morality of Things*, cit. pp. 30-33 and 61-63.

well as the different ethical traditions.

By contrast, although the universal nature of human rights necessarily entails contextualised application through national laws, which partially create context dependency and can lead to a certain degree of heterogeneity, human rights seem to provide a more stable framework. The different charters, with their provisions, but also regional courts (such as the European Court of Human Rights), and a coherent legal doc trine based on international experience can all help to reduce this dependence on context.

This does not mean that human rights do not present con- textual differences, but compared with ethical values, they are clearer, better defined, and stable. From a regulatory perspec- tive, this facilitates a better harmonisation and reduces the risk of uncertainty.

A fourth important, and largely unaddressed issue in the current debate on AI and ethics concerns the methodological approach that we might call overestimation and 'transplant' of ethical values. In a context characterised by ethics guidelines that pop up "like woodland mushrooms in a wet Autumn",[22] a number of studies have attempted to identify the core values through a quantitative and text-based approach[23].

The limitations of these studies are not restricted to the use of grey literature, results from search engines, and linguistic biases. The main limitation affecting this quantitative text-based approach is the lack of a policy perspective and contextual analysis.

Differing sources are considered on the same level, with- out taking into account the difference between the guidelines adopted by governmental bodies, independent authorities, private or public ad hoc committees, big companies, NGOs, academia, intergovernmental bodies etc. The mere frequency of occurrence does not reveal the different impacts of the distribution of these values amongst the different categories. For instance, the fact that some values are shared by several intergovernmental documents may have a greater policy impact than the same frequency in a cluster of NGOs or academic documents. When the focus is on values for future regulations, albeit based on ethics, the varying relevance of the sources in terms of political impact is important.

[22] See Raab, *Information Privacy, Impact Assessment, and the Place of Ethics*, cit.

[23] See e.g. A. Jobin, M. Ienca and E. Vayena, *The Global Landscape of AI Ethics Guidelines*, 1 *Nature Machine Intelligence* 389–399 (2019). The authors identified ten key ethical values in a set of 84 policy documents with the following distribution: transparency 73/84; non-maleficence 60/84; responsibility 60/84; privacy 47/84; beneficence 41/84; freedom and autonomy 34/84; trust 28/84; sustainability 14/84; dignity 13/84, and solidarity 6/84.

Despite this limitation, seven values are present in most documents:[24] five of them are ethical values with a strong legal implementation (transparency, responsibility, privacy, freedom and autonomy) and only two come from the ethical discourse (non-maleficence and beneficence).

Another study[25] identified several guiding values and the top nine, with a frequency of 50% or more, are: privacy protection; fairness, non-discrimination and justice; accountability; transparency and openness; safety and cybersecurity; common good, sustainability and well-being; human oversight, control and auditing; solidarity, inclusion and social cohesion; explainability and interpretability. As in the previous study, the aggregating of these principles is necessarily influenced by the categories used by the authors to reduce the variety of principles. In this case, if we exclude values with a legal implementation, the key ethical values are limited to openness, the common good, well-being and solidarity.

If we take a qualitative approach, restricting the analysis to the document adopted by the main European organisations and to those documents with a general and non-sectoral perspective[26], we can better identify the key values that are most popular amongst rule makers.

Considering the four core principles (respect for human autonomy, prevention of harm, fairness, and explicability) identified by the High-Level Expert Group on Artificial Intelligence[27], respect for human autonomy and fairness are widely developed legal principles in the field of human rights and law in general, while explicability is more a technical requirement than a principle. Regarding the seven requirements identified by the HLGAI[28] on the basis of these principles, human agency and oversight are further specified as respect for fundamental rights, informed autonomous decisions, the right not to be subject to purely automated decisions, and adoption of over- sight mechanisms. These are all requirements already present in the law in various forms, especially with regard to data processing. The same applies to the remaining requirements (technical robustness and safety, privacy and

[24] *Id.*

[25] T. Hagendorff, *The Ethics of AI Ethics: An Evaluation of Guidelines*, 30 *Minds and Machines* 99, 2020, p. 102.

[26] See e.g. Council of Europe – CEPEJ, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment*, Strasbourg, 3-4 December 2018.

[27] AI HLEG, *Ethics Guidelines for Trustworthy AI*, cit.

[28] These requirements are: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental wellbeing; accountability.

data governance; transparency; diversity, non-discrimination and fairness; accountability; and environmental well-being).

Looking at the entire set of values provided by the HLGAI, the only two elements – as framed in the document – that are only partially considered by the law are the principle of harm prevention – where "harms can be individual or collective, and can include intangible harm to social, cultural and political environments" – and the broad requirement of societal well-being, which generally requires a social impact assessment.

Another important EU document identifies nine core ethical principles and democratic prerequisites[29]. Amongst them, four have a broader content that goes beyond the legal context (human dignity, autonomy, solidarity and sustainability). However, in the field of law and technology, human dignity and autonomy are two key values widely considered both in the human rights framework and in specific legal instruments.

Based on the results of these different analytical methodologies (quantitative, qualitative), we can identify two main groups of values that expand the legal framework. The first consists of broad principles derived from ethical and sociological theory (common good, well-being, solidarity). These principles can play a crucial role in addressing societal issues concerning the use of AI, but their broad nature might be a limitation if they are not properly investigated and contextualised. The most interesting group is the second one, which includes the principle of non-maleficence, the principle of beneficence[30], and the related general claim of harm prevention. These are not new and undefined principles, especially in the field of applied ethics and research and medical ethics.

[29] European Commission - European Group on Ethics in Science and New Technologies, *Statement on Artificial Intelligence, Robotics and "Autonomous" Systems* (2018), <https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf >, accessed on Mar. 11, 2018. The Ethical principles and democratic prerequisites identified are: human dignity; autonomy; responsibility; justice, equity, and solidarity; democracy; rule of law and accountability; security, safety, bodily and mental integrity; data protection and privacy; sustainability.

[30] Although in the final version of the AI HLEG, *Ethics Guidelines for Trustworthy AI*, cit., these two principles are not explicitly listed as key values, they do underpin the whole approach of the AI HLEG, as demonstrated by the text of the draft version of the guidelines used for the public consultation; see AI HLEG, *Draft Ethics Guidelines for Trustworthy AI* (2018) <https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai>, accessed on Dec. 18, 2018, ("Ensure that AI is human-centric: AI should be developed, deployed and used with an "ethical purpose", grounded in, and reflective of, fundamental rights, societal values and the ethical principles of Beneficence (do good), Non-Maleficence (do no harm), Autonomy of humans, Justice, and Explicability. This is crucial to work towards Trustworthy AI.").

In regard to these, we should therefore consider the potential risk of the 'transplant' of ethical values[31].

It is not surprising that both the AI HLEG and the EU Commission[32], when deciding to concretise the suggested ethical approach, moved from ethics to human rights. This is in line with the model suggested in the literature – and supported by NGOs active in the field of human rights – that considers human rights as the core of future AI regulation[33]. Compared with this doctrinal approach, the conclusion reached in Europe by the AI HLEG and the Commission still shows a partially improper overlap between these two areas[34].

The notion of solving the problems of the variability of ethical values and their contextual nature by linking them to human rights seems inappropriate. Ethics can play an important role in AI regulation, not as a backdrop to a human rights-based approach, creating confusion about the existing definition of these rights, but as a complementary element. Ethics can cover those issues that are not addressed, or not fully addressed, by the human rights framework, and revolve around a discretionary evaluation based on the socio-ethical values of a given community with respect to the various human rights-orientated options to be chosen[35].

---

[31] See also Z. M. Schrag, *Ethical Imperialism. Institutional Review Boards and the Social Sciences 1965-2009*, Baltimore, 2017.

[32] See AI HLEG, *Ethics Guidelines for Trustworthy AI*, cit.; European Commission, *White paper on Artificial Intelligence - A European approach to excellence and trust*, COM(2020) 65 final, Brussels, Feb. 19, 2020; European Parliament, *Framework of ethical aspects of artificial intelligence, robotics and related Technologies European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies* (2020/2012(INL))' P9_TA-PROV(2020)0275.

[33] See A. Mantelero, *Regulating AI within the Human Rights Framework: A Roadmapping Methodology*, in P. Czech et al. (eds.) *European Yearbook on Human Rights*, 2020, pp. 477-502; K. Yeung, A. Howes and G. Pogrebna, *AI Governance by Human Rights–Centered Design, Deliberation, and Oversight*, in M. D. Dubber, F. Pasquale, and S. Das (eds.), *The Oxford Handbook of Ethics of AI*, Oxford, 2020, pp. 77-106; L. McGregor, D. Murray and V. Ng, *International Human Rights Law as a Framework for Algorithmic Accountability*, 68 *Int'l and Comp. L. Quarterly* 309-343, 2019; *Human Rights in The Age of Artificial Intelligence*, Access Now, 2018, <https://www.accessnow.org/cms/ assets/uploads/2018/11/AI-and-Human-Rights.pdf>, accessed 23 November 2020; P. Nemitz, *Constitutional Democracy and Technology in the Age of Artificial Intelligence*, 376 *Phil. Trans. Royal Soc'y A*, 2018, DOI: /10.1098/rsta.2018.0089.

[34] This uncertain focus is still present in the recent European Parliament, *Framework of ethical aspects of artificial intelligence*, cit., where the core requirements are based on fundamental rights, but the framework refers to ethics, including on risk management.

[35] See Raab, *Information Privacy, Impact Assessment, and the Place of Ethics*, cit.; A.

For example, a heavily data-driven community (see Section 6) is not necessarily contrary to human rights, since data can be processed in compliance with the law. But it does raise questions as to how we wish technology to be used in society, to what extent we want a pervasive digitalisation of everything in the name of a promised greater efficiency. These are ethical and social quandaries that need to be addressed and go beyond the human rights framework.

Based on these considerations on the interplay between law and ethics in regulating AI, for the purposes of this contribution, we can conclude that a focus on the human rights framework is necessary and crucial for an effective development of a human-centric AI. In this regard, an ex ante Human Rights Impact Assessment (HRIA)[36] is a necessary requirement in the development and deployment of AI solutions to prevent any prejudice to human rights and fundamental freedoms, as well as to promote a human rights-orientated AI[37].

At the same time, a positive outcome of such an assessment does not exclude the presence of ethical issues related to the proposed solution, which should be further investigated[38]. The HRIA thus makes the role of human rights in technology development evident and, in this way, helps to avoid improper overlap with ethical issues.

Mantelero, *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*, 34 *Computer L. & Security Rev*. 754-772, 2018.

[36] HRIA differs from other assessment methodologies used in the field of technology, such as technology assessment (TA) or impact assessment (IA). TA and IA have different focuses: TA is centred on technology development (e.g. road mapping) while IA on policy planing. See also A. Grunwald, *Technology Assessment: Concepts and Methods*, in A. Meijers, *Phil. Technology and Engineering Sciences. Handbook of the Philosophy of Science*, vol. 9, 2009, pp.1103-1146; T. A. Tran and T. Daim, *A taxonomic review of methods and tools applied in technology assessment*, in *Technol. Forecast. Soc. Change*, vol. 75, 9, 2008, pp. 1396-1405; World Bank and Nordic Trust Fund, *Human Rights Impact Assessments: A Review of the Literature, Differences with other forms of Assessments and Relevance for Development*, Washington, 2013.

[37] See Council of Europe – Ad hoc Committee on Artificial Intelligence, *Feasibility Study*, Strasbourg, 2020, pp. 44, 50, <www.coe.int/cahai>, accessed on Jan. 23, 2021; European Union Agency for Fundamental Rights (FRA), *Getting the Future Right– Artificial Intelligence and Fundamental* Rights, 2020, 87-98, <https://fra.europa.eu/en/publication/2020/ artificial-intelligence-and-fundamental-rights>, accessed on Dec. 14. 2020.

[38] For this reason, models like the HRESIA (Mantelero, *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*, cit.) may be better suited to grasp the holistic definition of the relationship between humans and machines.

4. *Defining an operational approach to human rights assessment in AI*

In considering the impact of AI on human rights, the dominant approach in many documents is mainly centred on listing the rights and freedoms potentially impacted[39] rather than operationalising this potential impact and proposing assessment models.

Moreover, case-specific assessment is more effective in terms of risk prevention and mitigation than using risk presumptions based on an abstract classification of "high-risk sectors and high-risk uses or purposes"[40], where sectors, uses and purposes are very broad categories which include different kind of applications – some of them continuously evolving– with a variety of potential impacts on rights and freedoms that cannot be clustered ex ante on the basis of risk thresholds, but require a case-by-case impact assessment.

Similarly, the adoption of a centralised technology assessment carried out by national ad hoc supervisory authorities[41] can provide useful

[39] See F. Raso, H. Hilligoss, V. Krishnamurthy, C. Bavitz, L. Kim, *Artificial Intelligence & Human Rights Opportunities & Risks*, 2018, <https://cyber.harvard.edu/sites/default/files/2018-09/2018-09_AIHumanRightsSmall.pdf? subscribe=Download+the+Report>, accessed on Sept. 28, 2018; AI HLEG, *Ethics Guidelines for Trustworthy AI*, cit.; Council of Europe, Committee of Ministers, *Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems*; Council of Europe, *Algorithms and Human Rights. Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications*, 2018.

[40] See European Parliament (fn 32). The Annex (Exhaustive and cumulative list of high-risk sectors and of high-risk uses or purposes that entail a risk of breach of fundamental rights and safety rules) considers transport as a high-risk sector for AI application, but there are several AI applications in this field based on non-personal data and related to infrastructure management with no relevant impact on rights and freedoms. Similarly, energy production and distribution is considered as high-risk use/purpose, but several automated energy sharing and switching solutions used in this field have no relevant impacts on rights and freedoms. In addition, this proposal and its Annex combine the impact on fundamental rights and safety risks, which are different types of potential risks, based on different criteria and assessment methodologies, without clarifying which sector, use or purpose is relevant in terms of potential adverse impact on fundamental rights and where the impact is limited to safety only. Furthermore, a harm-based approach is taken ("cause injury or harm"), rather than a rights-based approach focusing on potential prejudice to fundamental rights and freedoms, centered on risk prevention and accountability, as in the human rights impact assessment and in the GDPR.

[41] See European Parliament, *Framework of ethical aspects of artificial intelligence*, cit., Article 14.2 ("the risk assessment of artificial intelligence, robotics and related technologies, including software, algorithms and data used or produced by such technologies, shall be carried out, in accordance with the objective criteria provided for

guidelines for technology development and can be used to fix red lines[42] but must necessarily be complemented by a case-specific assessment of the impact of each application developed.

For these reasons, a case specific impact assessment remains the main tool to ensure accountability and the safe- guarding of individual and collective rights and freedoms. In this regard, a solution to the problem could easily be drawn from the human rights impact assessment models already adopted in several fields.

However, these models are usually designed for different contexts than those of AI applications[43]. The latter are not necessarily large-scale projects involving entire regions with multiple social impacts. Although there are important data-intensive projects in the field of smart cities, regional services (e.g. smart mobility) or global services (e.g. online content moderation provided by big players in social media), the AI operating context for the coming years will be more fragmented and distributed in nature, given the business environment in many countries, often dominated by SMEs, and the variety of communities interested in setting-up AI-based projects. The growing number of data scientists and the decreasing cost of hardware and software solutions, as well as their delivery as a service, will facilitate this scenario characterised by many projects with a limited scale, but involving thousands of people in data-intensive experiments.

For such projects, the traditional HRIA models are too articulated and oversized, which is why it is important to provide a more tailored model of impact assessment, at the same time avoiding mere theoretical abstractions based on generic decontextualised notions of human rights. To address this challenge, we have chosen to build the proposed model on the experience of data protection authorities (hereafter DPAs) in Europe, taken as a case study to identify the impacted areas of data intensive systems in relation to human rights.

in paragraph 1 of this Article and in the exhaustive and cumulative list set out in the Annex to this Regulation, by the national supervisory authorities referred to in Article 18 under the coordination of the Commission and/or any other relevant institutions, bodies, offices and agencies of the Union that may be designated for this purpose in the context of their cooperation").

[42] See, on the debate on the adoption of specific red lines regarding the use of AI in the field of facial recognition, European Digital Rights (EDRi), *Civil Society Calls for AI Red Lines in the European Union's Artificial Intelligence Proposal*, 2021, <https://edri.org/our-work/civil-society-call-for-ai-red-lines-in-the-european-unions-artificial-intelligence-proposal/>, accessed on Mar. 15, 2021.

[43] See *infra* n. 158 and 250.

DPAs, more than any other supervisory or judicial body, have in the last few decades addressed crucial issues concerning the use of data-intensive and data-invasive systems. Moreover, the jurisprudence of these authorities, both at national and EU level, has traditionally been inspired by attention to respect for fundamental rights – also given the strict relationship existing in the European context between data protection and personality rights[44] – as confirmed by the nature of fundamental right recognised to data protection and the European Court of Human Rights jurisprudence on the protection of private life[45].

Against this background, the following sub-sections will investigate the jurisprudence of these authorities to figure out how data-intensive systems potentially affect human rights. Before passing to the empirical analysis providing solid evidence for a human rights assessment, it is worth briefly considering the role played by impact assessment tools with respect to the precautionary principle as an alternative way of dealing with the consequences of AI.

As in the case of potential technology-related risks, there are two different legal approaches to the challenges of AI: the precautionary approach and the risk assessment. These approaches are alternative, but not incompatible. Indeed, complex technologies with a plurality of different impacts might be better addressed though a mix of these two remedies[46].

As risk theory states, their alternative nature is related to the notion of uncertainty[47]. Where a new application of technology might produce

[44] See G. Brüggemeier, *Protection of personality rights in the law of delict/torts in Europe: mapping out paradigms*, in G. Brüggemeier, A. Colombi Ciacchi and P. O'Callaghan, *Personality Rights in European Tort Law*, Cambridge, 2010, pp. 5-37; S. Stromhölm, *Right of Privacy and Rights of Personality. A comparative Survey*, Stockholm, 1967, pp. 28–31.

[45] See European Union Agency for Fundamental Rights and Council of Europe, *Handbook on European Data Protection Law*, 2018, <http://fra.europa.eu/en/publication/2018/handbook-european-data-protection-law>, accessed on May 25, 2018.

[46] See also Council of Europe, Consultative Committee of the Convention 108 (T-PD), *Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data*, cit., Section IV, par. 1 and 2, where the precautionary approach is coordinated with an impact assessment that also includes ethical and social issues.

[47] On the distinction between the precautionary approach and the precautionary principle, see J. Peel, *Precaution - A Matter of Principle, Approach or Process?*, 5 *Melb. J. Int'l L.* 483, 2004, <http://www.austlii.edu.au/au/journals/ MelbJlIntLaw/2004/19.html>, accessed on Feb. 4, 2017 ("One way of conceptualising what might be meant by precaution as an approach […] is to say that it authorises or permits regulators to take precautionary measures in certain circumstances, without dictating a particular response

potential serious risks for individuals and society, which cannot be accurately calculated or quantified in advance, a precautionary approach should be taken[48].

In this case, the uncertainty associated with applications of a given technology makes it impossible to conduct a concrete risk assessment, which requires specific knowledge of the ex- tent of the negative consequences, albeit in specific classes of risks[49].

Where the potential consequences of AI cannot be fully envisaged, as in the case of the ongoing debate on facial recognitions and its applications, a proper impact assessment is impossible, but the potentially high impact on society justifies specific precautionary measures (e.g., a ban or restriction on the use of AI-based facial recognition technologies)[50].

in all cases. Rather than a principle creating an obligation to act to address potential harm whenever scientific uncertainty arises, an approach could give regulators greater flexibility to respond").

[48] Only few contributions in law literature take into account the application of the precautionary approach in the field of data protection, see L. Costa, *Privacy and the precautionary principle*, 28 *Computer L. & Security Rev*. 14–24 (2012), and M. E. Gonçalves, *The EU data protection reform and the challenges of big data: remaining uncertainties and ways forward*, 26 *Inform. Comm. Tech. L*. 90-115, 2017. See also W. Pieters, *Security and Privacy in the Clouds: A Bird's Eye View*, in S. Gutwirth, Y. Poullet, P. de Hert, R. Leenes (eds.), *Computers, Privacy and Data Protection: an Element of Choice*, 2011, p. 455 ("generalised to information technology, it can serve as a trigger for government to at last consider the social implications of IT developments. Whereas the traditional precautionary principle targets environmental sustainability, information precaution would target social sustainability"). On the precautionary approach in data protection, see also A. Narayanan, H. Joanna and E. W. Felten, *A Precautionary Approach to Big Data Privacy*, in S. Gutwirth, R. Leenes, P. De Hert (eds) *Data Protection on the Move*, 2016, pp. 357-385; C. Raab and D. Wright, *Surveillance: Extending the Limits of Privacy Impact Assessment*, in D. Wright and P. De Hert (eds), *Privacy Impact Assessment*, 2012, p. 364; O. Lynskey, *The Foundations of EU Data Protection Law*, Oxford, 2015, p. 83; C. Raab, *The future of privacy protection*, Cyber Trust & Crime Prevention Project, 2014, p. 15, <https://www.piawatch.eu/node/86>, accessed on April 28, 2017.

[49] See also J. Tosun, *How the EU Handles Uncertain Risks: Understanding the Role of the Precautionary Principle*, 20 *J. European Public Pol'y* 1517-1528, 2013; T. Aven, *On Different Types of Uncertainties in the Context of the Precautionary Principle*, 31 *Risk Analysis* 1515–1525, 2011; A. Stirling and D. Gee, *Science, precaution, and practice*, 117 *Public Health Reports* 521–533, 2002.

[50] See e.g. European Parliament - Committee on Civil Liberties, Justice and Home Affairs, *Opinion of the Committee on Civil Liberties, Justice and Home Affairs for the Committee on Legal Affairs on artificial intelligence: questions of interpretation and application of international law in so far as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice*, 2020/2013(INI) (2020), par. 14, 15

This does not mean limiting innovation, but investigating more closely its potentially adverse consequences and guiding the innovation process and research, including the mitigation measures (e.g. containment strategies, licensing, standards, labelling, liability rules, and compensation schemes).

On the other hand, where the level of uncertainty is not so high, the risk-assessment process is a valuable tool in tackling the risks stemming from technology applications. According to the general theory on the risk-based approach, the process consists of four separate stages: 1) identification of risks, 2) analysis of the potential impact of these risks, 3) selection and adoption of the measures to prevent or mitigate the risks, 4) periodic review of the effectiveness of these measures[51]. Furthermore, to enable subsequent monitoring of the effective level of compliance, duty bearers should document both the risk assessment and the measures adopted.

Since neither the precautionary principle nor the risk assessment are an empty list but rather focus on specific rights and freedoms to be safeguarded, they can be seen as two tools for developing a human rights-centred technology. While the uncertainty of some technology solutions will lead to the application of the precautionary principle, a better awareness and management of related risk will enable a proper assessment.

However, the relationship between risk assessment and the precautionary principle is rather complicated and cannot be reduced to a strict alterative. Indeed, when a precautionary approach suggests that a technology should not be used in a certain social context, this does not necessary entail halting its development. On the contrary, where there is no incompatibility with human rights (e.g. mass destruction harms) the technology can be developed further to reach a sufficient level of maturity that shows awareness of the related risks and the effective solutions.

This means that, in these cases, human rights can play an additional role in guiding development such that, once it reaches a level of awareness of the potential consequences that exclude uncertainty, will be subject to risk assessment.

---

and 20. See also Council of Europe, Consultative Committee of the Convention 108 (T-PD), *Guidelines on Facial Recognition*, 28 January 2021, T- PD(2020)03rev4, par. 1.1.

[51] See also R. Koivisto and D. Douglas, *Principles and Approaches in Ethics Assessment. Ethics and Risk*. Annex 1.h Ethical Assessment of Research and Innovation: A Comparative Analysis of Practices and Institutions in the EU and selected other countries. Project Stakeholders Acting Together on the Ethical Impact Assessment of Research and Innovation – SATORI. Deliverable 1.1', 2015, <http://satoriproject.eu/work_packages/comparative-analysis-of-ethics-assessment-practices/>, accessed on Feb. 15, 2017.

Under this reasoning, two different scenarios are possible. One in which the precautionary principle becomes an outright ban on a specific use of technology and the other in which it restricts the adoption of certain technologies but not their further development. In the latter case, a precautionary approach and a risk assessment are two different phases of the same approach rather than an alternative response.

4.1. *A methodological approach for an evidence-based model*

Having defined the importance of a human rights-orientated approach in AI data processing, there remains the methodological question of how to define the assessment benchmark.

Three different approaches are possible: (i) a top-down theoretical approach; (ii) an inferential approach, and (iii) a bottom-up empirical approach. The first was used in the analysis conducted by Raso et al.,[52] in which various potentially affected rights are analysed on the basis of abstract scenarios grouped by sector-specific applications (risk assessments in criminal justice, credit scoring, healthcare diagnostics, online content moderation, recruitment and hiring systems, essay scoring in education).

The second approach was adopted by Fjeld et al.,[53] inferring values from existing ethics and right-based documents on AI regulation. This approach is close to the empirical approach, but is dominated by a quantitative dimension, focusing on the frequency of certain principles, and overlooking the heterogeneity of the documents. As the documents are often declarations by governmental and non-governmental bodies, they have the nature of guidelines and directives rather than concrete descriptions of the existing state-of-the-art: they are more focused on To-Be rather than on As-Is. This means that the prevalence of certain principles and values does not necessarily demonstrate a concrete and effective implementation of them.

The third approach, used in this work, adopts an evidence- based methodology grounded on empirical analysis of cases decided by DPAs and guidelines provided by these authorities. More specifically, the idea is to move from the reasoning adopted by decision-makers in scrutinising

[52] Raso, Hilligoss, Krishnamurthy, Bavitz, Kim, *Artificial Intelligence & Human Rights Opportunities & Risks*, cit.

[53] Jessica Fjeld et al., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, Berkman Klein Center for Internet & Soc'y, 2020, <https://dash.harvard.edu/handle/1/42160420>, accessed on Jan. 15, 2020.

data-centred applications and use this experience to better understand which rights and freedoms are relevant in practice.

The advantage and complementarity nature of this approach is its consistency with the already existing practices in the field of data protection, where DPAs are the supervisory bodies that are addressing and will further address the largest number of issues related to AI context. Using an empirical legal research methodology, this analysis focuses on what already exists in concrete practice and can thus be extended to cases concerning AI. In this sense it differs from the first approach, in its bottom-up nature, and from the second, as here the link between the As-Is and To-Be is stronger.

A model based on the empirical evidence can be better understood and used both by supervisory authorities and operators already accustomed to existing DPAs' jurisprudence and practice. Moreover, from a regulatory point of view, this approach is consistent with the recent worldwide growth of evidence-based policies.

From a more specific data protection standpoint, the European proposals for a future AI regulation focus on risk assessment without providing concrete models for this assessment. In this regard, both the main European legally binding instruments, Convention 108+[54] and GDPR,[55] refer to "rights and fundamental freedoms" (Convention 108+) and to "rights and freedoms of natural persons" (GDPR), without specifically identifying them. Similarly, the DPIA templates adopted by DPAs do not provide enough detail on this point.

An evidence-based model has the advantage of not resting on its authors' abstract vision, which may be coloured by their individual theoretical or cultural standpoint but based on decision-makers' concrete practices. While they too are necessarily affected by cultural influences, these decision-makers are the ones who will address the new cases concerning AI. Adopting this culture-specific perspective does not undermine the expected outcome but puts it in context.

At the same time, the evidence-based analysis adopted here maintains a general approach to the issues it deals with, without a case-specific focus – such as in other empirical studies concerting AI[56] – that cannot be generalised.

[54] See Convention 108+, arts. 6 and 10.

[55] GDPR, Article 35.

[56] See e.g. K. Crawford and V. Joler, *Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources*, AI Now Institute and Share Lab, Sept. 7, 2018, <http://www.anatomyof.ai>, Oct. 20, 2018.

Regarding the potential limitations affecting this approach, it is true that almost all the cases decided by supervisory authorities do not directly concern AI, given the time lag between technology development and litigation, but the reasoning adopted in current data intensive cases can be considered as a useful proxy for the application of human rights in data-intensive systems based on AI.[57]

In addition, compared to the decision of international courts, and the European Court of Human Rights,[58] DPAs have implemented a broader set of actions, not circumscribed to case decisions, including guidelines and other documents that contribute to defining best practices in data processing and are of interest for the contextualisation of human rights and freedoms in the AI context.

Given the enormous number of decisions made by these bodies, selection favoured cases where data use might entail an impact on human rights. The geographical area of investigation was limited to Europe, focussing on six countries with a longer experience in regulating data processing: Belgium, France, Germany, Italy, United Kingdom, and Spain.

Given the influence that social contexts may have on hu- man rights, the selection took into account both authorities belonging to systems with a similar socio-cultural environment (e.g. Italy and Spain), and authorities belonging to systems with a distinct legal and social culture (UK). Since this necessarily entailed the exclusion of a significant area within the EU, the research also considered the opinions and documents adopted by the Article 29 Data Protection Working Party and the European Data Protection Board to have a more comprehensive overview.

The decision not to circumscribe the research to the documents adopted by national DPAs was also influenced by the relationship between data protection claims and technology development. Indeed, the potential harms resulting from the use of innovative technologies and data-intensive systems might not yet be known to data subjects, but the prejudices may be well perceived and discussed in the context of the activities carried out

[57] It is also worth noting that in the DPAs' decisions, a direct focus on human rights is less prominent and explicit in the motivations, as detailed below in Section 4.2.

[58] See e.g. the decision of the Court of Justice of the European Union on the so-called right to be forgotten and the further need to define specific policies for its implementation. See Court of Justice of the European Union, 13 May 2014, Case C-131/12, *Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*, OJ C 212, 07.07.2014, 4. Similarly, the decisions of the European Court of Human Rights (hereinafter ECtHR), albeit based on concrete cases, focus on rights and freedoms violations, but do not develop concrete solutions for data processing.

by supranational bodies. This also explains why the decisions of national DPAs examined rarely refer to new technological solutions.

More than 700 documents were analysed,[59] selected on the basis of their relevance to human rights and fundamental freedoms. The concepts used to extract the most relevant documents from the databases of the decisions adopted by DPAs took into account, inter alia, the nature of the devices used for data collection (e.g. video-surveillance systems, geolocation tools, IoT systems and personal devices), the places where data is collected (public or private spaces, the workplace, etc.) and the nature of the data (e.g. biometric data).

The number and nature of documents examined differ on national basis. As these documents cover a wide period (1994–2020), most of the materials examined are based on Directive 95/46/EC and a more limited number of decisions refer to the GDPR, given its relatively recent entry into force and the inevitable time lag between the first implementation of a new law and decided cases.

Regarding the materials based on Directive 95/46/EC, it is worth noticing the heterogeneity of the documents as a result of the different powers exercised by the national authorities before the GDPR came into force,[60] the different nature of their acts and their policy approaches[61]. This diversity has been mitigated by the advent of the GDPR, which provided for uniform regulation of the DPAs' powers, as reflected in the subset of decisions referring to the most recent cases.

[59] The analysis is based on the documentation made available on the official websites of the DPAs and the EDPB, see <https://www.garanteprivacy.it/> (Italian DPA); <https://www.cnil.fr> (French DPA); <https://www.autoriteprotectiondonnees.be/> (Belgian DPA); <https://www.bfdi.bund.de/DE/> (Federal German DPA); <https://ico.org.uk> (UK DPA); <https://edpb.europa.eu/edpb_en> (EDPB). The documents adopted by the Article 29- Data Protection Working Party are available at <https://ec.europa.eu/justice/article-29/documentation/index_en.htm>.

[60] See, Baker & Mc Kenzie, *Global Data Protection Enforcement Report - Enforcement by regulators: penalties, powers and risks*, 2016, <https://www.bakermckenzie.com/-/media/files/insight/events/2018/04/gdrp-enforcement.pdf?la=en>, accessed on Dec. 10, 2017.

[61] The documents from the Italian, Spanish and French authorities are mainly decisions on specific complaints, whereas the UK DPA documents tend to be guidelines, recommendations and information provided to various industries. The documents from Belgium are mainly recommendations. As regards Germany, the statements of the Federal DPA and the minutes of the meetings between the federal and the länder DPAs were taken into consideration. These meetings (Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder) adopt agreed resolutions which outlining the attitude of federal and länder privacy authorities with regard to technical, economic and legal issues concerning data processing.

From a methodological perspective, although the materials examined can be divided into these two subsets (before and after the GDPR), this does not affect the overall analysis for several reasons: (i) Directive 95/46/EC and the GDPR are grounded on the same core principles; (ii) human rights are often relevant in data processing activities and in this sense data protection is considered as an enabling right with a view to human rights protection; (iii) although the GDPR introduced a risk assessment approach focused on the rights and freedoms of natural persons, the Regulation does not provide specific rules on rights and freedoms other than data protection and even the data protection impact assessment enshrined in Article 35 needs specific implementation.

For these reasons, we examined these two clusters of data in a unified way, without distinguishing between the pre- and post-GDPR periods. Moreover, for the EU context, this analysis can provide a contribution in terms of developing concrete solutions to carry out the impact assessment required by the GDPR.

As for document selection, the collected materials were analysed in detail to identify the most significant cases and discard those concerning the same issues or adopting a simi- lar argumentative logic. At the end of this screening phase, 350 documents were taken into consideration for the purposes of this study (broken down as follows: Italy100, Spain 35, France 60, Belgium 40, Germany 20, United Kingdom 45, Article 29 Data Protection Working Party 50)[62].

[62] It is worth noting that this uneven distribution of cases among the countries examined is the result of a content-based approach and not of an intentional deeper investigation in one country rather than another. Several factors may have produced this (unintentional) geographical distribution due to their impact on publicly available decisions. For example, prior to the GDPR (the examined period is 1994-2020) some DPAs had no sanctioning power (UK, Belgium), some had concurrent competence with regional authorities (Germany), some DPAs (Belgium) had a model more favourable to ADRs, and some had less performing internal search engines. With regard to the latter, we conducted an initial selection based on keywords and, as there is no single search engine, but each authority has its own, the low performance of some of these search services may have provided fewer results for some DPAs. In addition, the issues investigated with respect to human rights are more present with respect to some topics (e.g. video surveillance) than others, with a potential different distribution among DPAs according to the concrete demands addressed. The fact that some DPAs have mainly issued recommendations also contributes to reducing the number of documents available for some DPAs, compared to others that have a large number of decided cases. Finally, even when DPAs have adopted many decisions, there are large clusters of cases concerning aspects (e.g. access rights or lack of informed consent) that have little or no relevance from a human rights perspective.

4.2. *Human rights and data use in the DPAs' jurisprudence*

Despite the authorities considered belonging to different legal and cultural traditions, the analysis of the documents did find common ground between them in their approach to human rights and freedoms.

It worth noting that, whereas the importance of these interests is clearly stated, in several cases the analysis of their relevance is not properly developed and in others emerges only indirectly in the DPAs' observations. In fact, DPAs often prefer to refer to principles such as proportionality, necessity or transparency set forth in the data protection regulations to safeguard interests other than privacy and data protection, without a further elaboration.

However, we identified a special attention to the possible risks for individual rights and freedoms, with specific reference to human rights principles and to several human rights and freedoms. The following sub-sections will analyse this evidence.

4.2.1. *Respect for human dignity*

A first core value underlying the DPA's decisions is human dignity, recognised as crucial in many legal systems and widely protected in European[63] and international frameworks[64]. Despite the difficulties in

[63] See Article 1, EU Charter of Fundamental Rights. See also, e.g., Court of Justice of the European Union, 12 November 2019, Case C-233/18, *Zubair Haqbin v. Federaal agentschap voor de opvang van asielzoekers*. Although the European Convention on Human Rights does not explicitly refer to the notion of human dignity, there is no doubt that, implicitly, this document affirms the value of respect for human dignity. A confirmation of this can be found in the jurisprudence of the European Court of Human Rights, where it is clear that human dignity is implicated in the Convention's protective regime (see e.g. ECtHR, 22 November 1995, *S.W. v. the United Kingdom*, App. No. 20166/92, para 44). See, among others, R. Brownsword, *Human dignity from a legal perspective*, in M. Düwell, J. Braarvig, R. Brownsword and D. Mieth (eds), *The Cambridge Handbook of Human Dignity. Interdisciplinary Perspectives*, Cambridge, 2014, p. 1.

[64] See UDHR, Preamble, which refers to dignity as an inherent value of each human being, simply as an innate consequence of human existence. In particular, dignity is seen as the core value that underpins human rights to which three basic values refer: liberty, equality and solidarity. See B. De Gaay Fortman, *Equal dignity in international human rights*, in Düwell et al. cit., p. 356. See also, among others, J. Mårtenson, *The Preamble of the Universal Declaration of Human Rights and the UN Human Rights Programme*, in E. Asbjørn et al. (eds), *The Universal Declaration of Human Rights: A Commentary*, 1992, p. 17.

determining the precise meaning of this concept[65], it is a key notion in human rights law and a guiding principle that underpins and grounds all other principles in human rights[66], even in the context of data processing[67].

We also found this broad notion of human dignity in the decisions of the DPAs, according to which human dignity encompasses various aspects of the individual sphere and is an important factor in many different contexts.

For instance, the DPAs recognise that negative outcomes for individual dignity may result from continuous and invasive monitoring, such as video surveillance or other monitoring technologies[68] or data-intensive systems collecting mobility data and driving behavioural information (e.g. GPS; Wi-Fi tracking devices; RFID technologies; Intelligent Transport Systems and "event data recorder" devices)[69]. They regard these practices as

[65] See, among others, C. Dupré, *Art. 1 – Human dignity*, in S. Peers, T. Hervey, J. Kenner, and A. Ward (eds), *The EU Charter of Fundamental Rights. A Commentary*, 2014, p. 18. See also R. Dworkin, *Taking Rights Seriously*, 1978, pp. 198-199.

[66] See G. Den Hartogh, *Is human dignity the ground of human rights?*, in Düwell et al. cit., p. 200; De Gaay Fortman, *Equal dignity in international human rights*, cit., p. 356; Dupré, *Art. 1 – Human dignity*, cit., p. 24. See also *Explanation on Article 1 — Human Dignity (Explanations Relating to the Charter of Fundamental Rights)*, in O. J. European Union C 303/17 - 14.12.2007, <https://eur-lex.europa.eu/legal- content/EN/TXT/PDF/?uri=OJ:C:2007:303:FULL&from=EN>, accessed on Sept. 2, 2019 ("The dignity of the human person is not only a fundamental right in itself but constitutes the real basis of fundamental rights").

[67] See Council of Europe, *Convention 108+*, Preamble.

[68] With reference to video surveillance of workers, see Information Commissioner's Office (hereinafter ICO), *The employment practices code*, 2011, Part. 3; Garante per la protezione dei dati personali (hereinafter GPDP), April 4, 2013, doc. web n. 2439178; GPDP, Oct. 30, 2013, doc. web n. 2851973; Commission de la protection de la vie privée (hereinafter CPVP), avis, n. 8/2006, 12 April 2006; Commission Nationale de l'Informatique et des Libertés (hereinafter CNIL) n. 2014-307, July 17, 2014; Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (hereinafter BFDI), *Videoüberwachung am Arbeitsplatz*. In relation to invasive monitoring activities see also GPDP, Jan. 25, 2018, doc. web n. 7810766 (monitoring of patients, within a health-care facility, through the use of wearable devices); Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder (hereinafter DSK), *„Videoüberwachungsverbesserungsgesetz" zurückziehen!*, Nov. 9, 2016.

[69] See GPDP, Sept. 8, 2016, doc. web n. 5497522; GPDP, Nov. 7, 2013, n. 499, doc. web n. 2911484; CPVP, avis n. 12/2005, Sept. 7, 2005; CPVP, recommandation n. 01/2010, 17 March 2010; CNIL n. 2010-096, April 8, 2010; Agencia Española de Protección de Datos (hereinafter AEPD), resolución R/01208/2014; DSK, *Gesetzesentwurf zur Aufzeichnung von Fahrdaten ist völlig unzureichend!*, Mar. 16, 2017; ICO, *Data Protection Technical Guidance Radio Frequency Identification*, 2006; ICO, *Wi-fi location analytics*, 2016; Article 29-Data Protection Working Party (hereinafter

potentially oppressive or demeaning, if associated risks are not mitigated.

Human dignity also plays a role in DPAs' decisions on monitoring activities of a private or intimate nature which might create discomfort for individuals (e.g. monitoring of employees' electronic communications or Internet use[70]). According to the DPAs, human dignity is also particularly relevant when video-surveillance or other monitoring tools are used in environments characterised by a high privacy expectation (e.g. restrooms or changing rooms)[71].

Regarding the nature of the information used, DPAs see sensitive data as more closely linked to human dignity. This is evidenced by cases of invasive information requests by employers (e.g. health conditions, religious beliefs, criminal records, and drug and alcohol use)[72], biometric data collection[73], and the use of wearable and IoT devices to gather sensitive data (e.g. health data) or profiling information[74].

ART29WP), *Working document on data protection issues related to RFID technology*, WP 105 (2005); ART29WP, *Opinion 03/2017 on Processing personal data in the context of Cooperative Intelligent Transport Systems (C-ITS)*, WP 252 (2017); ART29WP, *Opinion 13/2011 on Geolocation services on smart mobile devices*, WP 185 (2011).

[70] See GPDP, July 9, 2020, doc. web n. 9474649; GPDP, Dec. 4, 2019, doc. web. n. 9215890; ICO, *The employment practices code*, cit., Part. 3; CPVP, avis n. 10/2000, April 3, 2000; BFDI, *Internet-und E-Mail Nutzung am Arbeitsplatz*; ART29WP, *Working document on the surveillance of electronic communications in the workplace*, WP 55 (2002).

[71] See GPDP, Dec. 4, 2008, doc. web n. 1576125; GPDP, 24 February 2010, doc. web n. 1705070 (use of written coupons to authorise workers to leave their workstation to go to the toilet); AEPD, expediente n. E/01760/2017; AEPD, expediente n. E/01769/2017; CNIL, décision n. 2013-029, 12 July 2013; BFDI (fn 68); ICO, *Installing CCTV? Things you need to do first*; ICO, *In the picture: A data protection code of practice for surveillance cameras and personal information*, 2017; ICO *The employment practices code*, cit., Part. 3; ICO, *Wi-fi location analytics*, 2016; ART29WP, *Opinion 4/2004 on the Processing of Personal Data by means of Video Surveillance*, WP 89 (2004). See also AEPD, procedimiento n. A/00109/2017, on the use of video surveillance systems by a hotel to monitor customers in relaxation areas.

[72] See GPDP, July 21, 2011, doc. web. n. 1825852; ICO, *The employment practices code*, cit., Parts 1 and 4; Hamburg Commissioner for Data Protection and Freedom of Information, Oct. 1, 2020, <https://datenschutz-hamburg.de/ assets/pdf/2020-10-01-press-release-h+m-fine.pdf >, accessed on Jan. 20, 2021. See also GPDP, Jan. 11, 2007, doc. web n. 1381620 (collection of sensitive information (e.g. sexual habits) by a real estate agency).

[73] See GPDP, Aug. 1, 2013, n. 384, doc. web n. 2578547; AEPD, Gabinete Jurídico, informe 0392/2011; CNIL n. 2008-492, Dec. 11, 2008; CPVP, avis n. 17/2008, April 9, 2008; ART29WP, *Opinion 3/2012 on developments in biometric technologies*, WP193 (2012).

[74] See ART29WP, *Opinion 8/2014 on the on Recent Developments on the Internet of Things*, WP 223, 2014; ART29WP, *Opinion 13/2011 on Geolocation services on smart*

Finally, DPAs pointed out how human dignity can also be affected by public disclosure of personal information, such as evaluation judgments (e.g. publication of exam results by schools[75] or employee evaluation ratings[76]; use of services of the so-called reputation economy[77]) or personal debt situations[78], which may cause distress and embarrassment to individuals.

4.2.2. *Freedom from discrimination*

According to DPAs, discriminatory practices[79] may occur in many contexts and in relation to different types of personal data processing. Negative consequences may result, for example, from automated decision-making and profiling activities, which may perpetuate existing stereotypes and social segregation[80].

With regard to AI-related applications, DPAs have focused on the risks of perpetuating discriminatory practices through automated profiling[81]. Moreover, as the criteria and functioning of algorithms are often opaque,

mobile devices, WP 185, cit. With regard to the collection of sensitive data likely to cause embarrassment and discomfort to the data subject, see also ART29WP, *Opinion 2/2010 on online behavioural advertising*, WP 171 (2010).

[75] See ICO, *Publication of exam results by schools*, 2014. In this regard, particular concerns were also expressed by ART29WP, *Opinion 2/2009 on the protection of children's personal data (General Guidelines and the special case of schools)*, WP 160, 2009; GPDP, *Scuola: Privacy, pubblicazione voti online è invasiva. Ammissione non sull'albo ma in piattaforme che evitino rischi*, doc. web n. 9367295 (2020).

[76] See GPDP, Dec. 13, 2018, n. 500, doc. web n. 9068983; BFDI, *Notenspiegel im Intranet*.

[77] For instance, platforms which display and manage product and service reviews, as well as tax or criminal information. See GPDP, Nov. 24, 2016, n. 488, doc. web n. 5796783.

[78] See GPDP, May 28, 2015, n. 319, doc. web n. 4131145; AEPD, procedimiento n. A/00104/2017.

[79] See, FRA, *Handbook on European non-discrimination law*, 2011, pp. 22-29, <https://fra.europa.eu/en/publication/2011/handbook-european-non-discrimination-law-2011-edition>, accessed on Jan. 11, 2021; ECtHR, 13 November 2007, *D.H. and Others v. the Czech Republic*, App. No. 57325/00, para. 175. With reference to the difference between direct and indirect discrimination see also C. Kilpatrick, *Art. 21 – Non-discrimination*, in Peers, Hervey, Kenner, and Ward, p. 592.

[80] See ICO, *Big data, artificial intelligence, machine learning and data protection*, 2017; ICO, *Guidance on AI and data protection*; CNIL, *Comment permettre à l'homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*, 2017; ART29WP, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, WP 251, 2017.

[81] See ICO, *Big data, artificial intelligence, machine learning and data protection*, 2017.

individuals might not know that they are being profiled or not understand the potential consequences. In this context, DPAs have considered, inter alia, the risk of bias that may arise from online behavioural advertising[82] or IoT-based profiling[83]. Likewise, DPAs have referred to the use of data-intensive systems in the context of police services and law enforcement, such as predictive policing[84].

Adverse discriminatory impacts may also result from the use of sensitive data to prevent or limit access to certain services or benefits. This is the case when sensitive information is requested by the employer (e.g. medical information), during the period of employment[85] or even at the time of recruitment[86], or by real estate agencies in order to meet the discriminatory requirements of property owners[87]. Sensitive data for discriminatory purposes may also be used by insurance companies, which may collect genetic data to calculate insurance costs on the basis of foreseeable individual health conditions[88].

Finally, in the case of discrimination too, monitoring and video-surveillance systems can have a negative impact on individuals and groups. This has led DPAs to highlight the unlawfulness of surveillance based exclusively on, inter alia, racial origin, religious or political opinions, membership in trade unions, or sexual orientation, without a justified reason[89].

---

[82] *Opinion 2/2010 on online behavioural advertising*, cit.; ICO, *Big data, artificial intelligence, machine learning and data protection*, cit.

[83] ART29WP, *Opinion 13/2011 on Geolocation services on smart mobile devices*, cit.

[84] See DSK, *Big Data zur Gefahrenabwehr und Strafverfolgung: Risiken und Nebenwirkungen beachten*, Mar. 18-19, 2015. See also GPDP, *Uomini e Macchine. Protezione dati per un'etica del digitale*, doc. web n. 7598686, 2018; ICO, *Guidance on AI and data protection*, cit.

[85] See ICO, *The employment practices code*, cit., Part 4; ICO, *The Employment Practices Code. Supplementary Guidance*, Part. 4, 2005; GPDP, June 5, 2019, n. 146, doc. web n. 9124510.

[86] See ICO, *The employment practices code*, cit., Part 1; CNIL, *Les opérations de recrutement*, 2013; GPDP, doc. web n. 9124510, cit.; ART29WP, *Working Document on Genetic Data*, WP 91, 2004.

[87] See GPDP, Jan. 11, 2007, doc. web n. 1381620; CPVP, recommendation n. 01/2009, Mar. 18, 2009.

[88] ART29WP, *Working Document on Genetic Data*, cit., ART29WP, *Opinion 3/2012 on developments in biometric technologies*, cit.; ART29WP, *Opinion 3/2012 on developments in biometric technologies*, cit. (on the potential discriminatory effects that may arise from the use of biometric data for targeting and profiling purposes).

[89] ART29WP, *Opinion 4/2004 on the Processing of Personal Data by means of Video Surveillance*, cit.

4.2.3. *Physical, psychological, and social identity*

Both the international and European legal frameworks consider personal identity in the broader context of individual privacy[90]. The personal aspects of an individual's identity traditionally cover different dimensions – physical, psychological, and social identity – and a range of data (e.g. name, image, reputation, family and ethnic heritage, gender identification, sexual, political and religious orientation)[91].

The notion of personal identity can thus encompass two different meanings: (i) the body of information that unequivocally identifies a person, distinguishing him or her from any other; (ii) information concerning the individual's projection in the social community.

The documents examined considered only the first meaning, with regard to data processing operations for personal identification.

One example is the use of biometric data to control access to certain areas (e.g. preventing outsiders from entering schools, tracking employees' whereabouts)[92] or the use of genetic data[93].

The need to protect personal identity also emerges from the considerations expressed by the authorities in relation to the identification information collected in the social media environment[94] and through RFID systems[95].

In the context of personal identity, the potential unauthorised or unlawful use of identification information is of particular concern to the

[90] See, article 12 UDHR; article 7 EUCFR; article 8 ECHR.

[91] See Jens Vested-Hansen, *Article 7 – Respect for Private and Family Life (Private Life, Home and Communications)*, in Peers et al., cit. p. 161; I. Roagna, *Protecting the right to respect for private and family life under the European Convention on Human Rights*, Council of Europe, Strasbourg, 2012, p. 12 <https://rm.coe.int/16806f1554>, accessed on Feb. 2, 2021; G. Brüggemeier, A. Colombi Ciacchi and P. O'Callaghan, *A common core of personality protection*, in G. Brüggemeier, A. Colombi Ciacchi and P. O'Callaghan, *Personality Rights in European Tort Law*, 2010, pp. 573-574. See also ECtHR, Jul. 20, 2010 (Final 20 October 2010), *Dadouch v. Malta*, App. No. 38816/07, pars. 47-48.

[92] See, among others, CNIL n. 2016-017, Jan. 28, 2016; GPDP, June 15, 2006, doc. web n. 1306098; AEPD, Gabinet Juridico, informe 0392/2011; ICO, *The use of biometrics in schools*, <https://schools.essex.gov.uk/data/information-governance/ Documents/ biometrics.pdf >, accessed on Feb. 17, 2021. See also CPVP, avis n. 17/2008, April 9, 2008; ART29WP, ART29WP, *Opinion 3/2012 on developments in biometric technologies*, cit.; GPDP, June 18, 2015, n. 360, doc. web n. 4170232 (use of a facial recognition system by a company operating in the cruise travel sector).

[93] See ART29WP, *Working Document on Genetic Data*, cit.

[94] See, Art. 29WP, *Opinion 5/2009 on online social networking*, WP163, 2009.

[95] See GPDP, 9 March 2005, doc. web n. 1109493; ART29WP, *Working document on data protection issues related to RFID technology*, cit.

DPAs as this information cannot be modified (biometric and genetic data) or cannot be easily modified. Given the importance of this information, the DPAs have emphasised the need to limit its use and to ensure an adequate level of security to prevent identity theft and other offences.

4.2.4.*Physical, psychological, and moral integrity and the intimate sphere*

Personal integrity is protected at European and international level as an aspect of an individual's private life[96], and comprises the individual's physical, psychological and moral integrity[97]. In this sense, a natural person must be free from any interference, both in relation to the body and the mind. Respect for the intimate sphere of the data subject is also an important aspect of safeguarding individuals' integrity, referring to its moral dimension[98].

The importance of the individual's physical integrity and intimate sphere is confirmed in DPA jurisprudence and opinions. Regarding the data subject's physical integrity, the DPAs mainly considered invasive data processing, such as that using implanted RFID devices (e.g. subcutaneous microchips) to collect and process personal information, including identification data, credit card number or health information[99].

The DPAs limit their use to situations where they are strictly necessary and there are no less intrusive alternatives, giving data subject the right to ask for their removal at any time.

With regard to the individual's intimate sphere, DPAs have focused on monitoring tools, including video-surveillance in environments where privacy expectations are high[100], as well as on the collection of biometric

[96] See article 12 UDHR; article 7 EUCFR ; article 8 ECHR. See also ECtHR, 24 July 2012 (Final 24 October 2012), *Dordevic v. Croatia*, App. No. 41526/10, para 97; ECtHR, 26 March 1985, *X and Y v. Netherlands*, App. No. 8978/80, para 22; ECtHR, 22 October 1996, *Stubbings and Others v. the United Kingdom*, App. No. 22083/93 and 22095/93, par. 61. Serious matters relating to physical and mental integrity fall under Articles 3 EUCFR and 3 ECHR. See also Roagna, *Protecting the right to respect for private and family life under the European Convention on Human Rights*, cit., p. 24; S. Michalowski, *Article 3 - Right to the Integrity of the Person*, in Peers et al. cit., p. 42.

[97] See Rehof, *Article 12*, in Asbjørn et al. cit., p. 187; Roagna, *Protecting the right to respect for private and family life under the European Convention on Human Rights*, cit., p.12; Vested-Hansen, *Article 7 – Respect for Private and Family Life (Private Life, Home and Communications)*, in Peers et al., cit., p.156.

[98] See ECtHR, May 28, 2015, *Y. v. Slovenia*, App. No. 41107/10.

[99] See AEPD, Gabinete Jurídico, informe 0292/2010; GPDP, Mar. 9, 2005, doc. web n. 1109493.

[100] Several cases concern restrooms and changing rooms and other places where privacy

data, given their invasive nature[101], which could interfere with the data subject's intimate sphere in an excessive way.

The last group of cases concerns data processing operations carried out using IoT wearable devices or other devices used in close vicinity to the human body in daily life (e.g. smartphones and smart home devices)[102]. The DPAs pointed out that the use of such devices is likely to cause significant interference to the individual's intimate sphere, by gathering information on health condition, behaviour, location, intimacy and many other aspects of the data subject.

4.2.5. *Self-determination and personal autonomy*

Individual self-determination and personal autonomy are widely recognised at both international and European level[103].

Personal autonomy is protected as an aspect of individual private life[104] and safeguards individuals against a wide range of external interference[105].

Individual self-determination and personal autonomy necessarily entail the ability to freely take decisions and have them respected by others[106]. According to international and European human rights

expectations are high: ICO, *In the picture: A data protection code of practice for surveillance cameras and personal information*, 2017; ICO, *The employment practices code*, cit., Part. 3; ICO, *Wi-fi location analytics*, 2016; CNIL n. 2014-307, July 17, 2014; CNIL, décision n. 2013-029, July 12, 2013; GPDP, July 10, 2014, doc. web n. 3325380; AEPD, procedimiento n. A/00109/2017; ART29WP, *Opinion 2/2009 on the protection of children's personal data (General Guidelines and the special case of schools)*, cit.; ART29WP, *Opinion 4/2004 on the Processing of Personal Data by means of Video Surveillance*, cit.

[101] See, among others, CNIL n. 2008-492, Dec. 11, 2008; GPDP, Jan. 31, 2013, doc. web n. 2304669; GPDP, May 30, 2013, doc. web n. 2502951; AEPD, Gabinete Jurídico, informe 0065/2015; ART29WP, *Opinion 3/2012 on developments in biometric technologies*, cit.

[102] See ART29WP, *Opinion 8/2014 on the on Recent Developments on the Internet of Things*, cit.; ART29WP, *Opinion 12/2011 on smart metering*, WP 183, 2011; ICO, *Privacy in mobile apps. Guidance for app developers*, 2013.

[103] See, in particular: article 12 UDHR; article 7 EUCFR; article 8 ECHR.

[104] See, among others, Roagna, *Protecting the right to respect for private and family life under the European Convention on Human Rights*, cit., p. 12; Vested-Hansen, *Article 7 – Respect for Private and Family Life (Private Life, Home and Communications)*, in Peers et al., cit., p. 156. See also ECtHR, April 29, 2002 (Final July 29, 2002), *Pretty v. the United Kingdom*, App. No. 2346/02, par. 61.

[105] The individual personal autonomy is protected both in the private (e.g. home and workplace) and in the public context (e.g. against interferences from public authorities). See Rehof, *Article 12*, in Asbjørn et al. cit., p. 187-201.

[106] See Commissioner for Human Rights, Human Rights and Disability, *Equal*

jurisprudence, individual personal autonomy also covers a further range of human behaviours, amongst which are the right to develop one's own personality and the right to establish and develop relationships with other people[107], the right to pursue one's own aspirations and to control one's own information[108].

Individual self-determination and personal autonomy also represent foundational principles in data protection, which is why it is not surprising that DPAs often refer to them in a broad sense[109]. These aspects emerge in both individual and relational contexts and involve freedom of choice, including freedom of movement and action, the free development of human personality and the right to informational self-determination.

Regarding freedom of choice, which encompasses freedom of movement and action, the DPAs have paid particular attention to the possible adverse effects of continuous and invasive monitoring. For instance, they considered cases of data processing carried out using video surveillance systems in workplaces[110] and schools[111] or in public spaces[112] (e.g. through

rights for all, Strasbourg, 20 October 2008, par. 5.2. <https://rm.coe.int/16806dabe6>, accessed on June 5, 2018. Moreover, this interest is also protected in relation to communications; see Council of Europe, *Guide on Article 8 of the European Convention on Human Rights. Right to respect for private and family life, home and correspondence*, updated on Aug. 31, 2019, <https://www. echr.coe.int/Documents/Guide_Art_8_ENG.pdf >, accessed on Jan. 13 2021.

[107] See ECtHR, May 18, 1976, *X v. Iceland*, App. No. 6825/74.

[108] See E. Fialová, *Data Portability and Informational Self- Determination*, in *Masaryk University Journal of Law and Technology*, 2014, (8), pp. 45-55; E. J. Eberle, *The Right to Information Self-Determination*, 4 *Utah L. Rev*. 965-1016, 2001.

[109] See also T. L. Beauchamp, J. F. Childress, *Principles of Biomedical Ethics*, 2001, p. 63 ("to respect an autonomous agent is, at a minimum, to acknowledge that person's right to hold views, to make choices, and to take actions based on personal values and beliefs").

[110] See CNIL n. 2010-112, April 22, 2010; GPDP, 30 October 2013, n. 484, doc. web n. 2908871; ICO, *The employment practices code*, cit., Part. 3; BFDI, *Videoüberwachung am Arbeitsplatz*, cit.; ART29WP, *Opinion 2/2017 on data processing at work*, WP 249, 2017; ART29WP, *Opinion 4/2004 on the Processing of Personal Data by means of Video Surveillance*, cit.

[111] GPDP, May 8, 2013, n. 230, doc. web n. 2433401; CPVP, avis, n. 8/2006, April 12, 2006; ART29WP, *Opinion 2/2009 on the protection of children's personal data (General Guidelines and the special case of schools)*, cit.

[112] CNIL, deliberation n. 94-056, 21 June 1994; ICO, *CCTV code of practice. Draft for consultation*, 2014; DSK, *Einsatz von Videokameras zur biometrischen Gesichtserkennung birgt erhebliche Risiken*, Mar. 30, 2017; DSK, *„Videoüberwachungsverbesserungsgesetz" zurückziehen!*, cit.); ART29WP, *Opinion 01/2015 on Privacy and Data Protection Issues relating to the Utilisation of Drones*, WP 231, 2015; ART29WP, *Opinion 4/2004 on the Processing of Personal Data by means of Video Surveillance*, cit. See also DSK,

the use of drones). The authorities also focus attention on the potentially negative outcomes arising from the use of devices such as wearable devices and smart meters[113]. Particular concerns were expressed in relation to monitoring activities through the use of mobile applications[114], as mobiles are strictly personal and are almost always on. Similarly, the authorities considered systems that collect mobility data, such as GPS, Wi-Fi tracking devices and RFID technologies[115].

Individual freedom of choice could be also undermined by communications monitoring[116], which might limit and influence individuals with respect to content and the decision to communicate it.

The DPAs clearly recognise that all these activities can limit or influence individual self-determination. Constant monitoring can have adverse consequences in terms of kerbing the data subject's behaviour in such a way as to comply with the controller's wishes[117]. Invasive

Gesetzesentwurf zur Aufzeichnung von Fahrdaten ist völlig unzureichend!, cit.

[113] See ART29WP, *Opinion 8/2014 on the on Recent Developments on the Internet of Things*, cit.; ART29WP, *Opinion 12/2011 on smart metering*, cit.

[114] See ICO, *Privacy in mobile apps. Guidance for app developers*, cit.

[115] With reference to the monitoring outside the workplace, see CNIL n. 2014-294, July 22, 2014; CNIL n. 2005-278, Nov. 17, 2005; GPDP, Nov. 7, 2013, n. 499, doc. web n. 2911484; GPDP, Mar. 9, 2005, doc. web n. 1109493; CPVP, avis n. 27/2009, Oct. 28, 2009; CPVP, recommendation n. 01/2010, Mar. 17, 2010; DSK, *Keine PKW-Maut auf Kosten des Datenschutzes!*, Nov. 14, 2014; DSK, *Datenschutz im Kraftfahrzeug – Automobilindustrie ist gefordert*, Oct. 8-9, 2014; ICO, *Data Protection Technical Guidance Radio Frequency Identification*, 2006; ICO, *Wi-fi location analytics*, 2016; ART29WP, *Working Party 29 Opinion on the use of location data with a view to providing value-added services*, WP 115, 2005; ART29WP, *Opinion 13/2011 on Geolocation services on smart mobile devices*, cit.; ART29WP, *Opinion 03/2017 on Processing personal data in the context of Cooperative Intelligent Transport Systems (C-ITS)*, cit.; ART29WP, *Working document on data protection issues related to RFID technology*, cit.. With regard to the collection of mobility data within the work context see GPDP, June 28, 2018, n. 396, doc. web n. 9023246; GPDP, May 18, 2016, n. 226, doc. web n. 5217175; CPVP, avis n. 12/2005, Sept. 7, 2005; CPVP, recommendation n. 03/2013, 24 April 2013; CNIL n. 2013-366, Nov. 23, 2013; CNIL n. 2006-066, March 16, 2006; ICO, *The employment practices code*, cit., Part. 3.

[116] See GPDP, Feb. 1, 2018, doc. web n. 8159221; GPDP, Mar. 8, 2018, n. 139, doc. web n. 8163433; AEPD, Gabinete Jurídico, informe 0464/2013; CPVP, avis n. 39/2001, Oct. 8, 2001; ICO, *The employment practices code*, cit., Part. 3; ICO, *The employment Practices Code. Supplementary Guidance*, 2005, Part. 3; BFDI, *Videoüberwachung am Arbeitsplatz*, cit.; ART29WP, *Working document on the surveillance of electronic communications in the workplace*, cit.; ART29WP, *Opinion 04/2014 on surveillance of electronic communications for intelligence and national security purposes*, WP 215, 2014.

[117] See, among others, J. W. Penney, *Chilling Effects: On-line Surveillance and Wikipedia*

monitoring can be more serious depending on the context in which it is carried out. This is the case in the workplace where there is an imbalance of power between the employer and the employee[118].

Self-determination and autonomy can also be affected by the aforementioned services of the so-called 'reputation economy'[119], which limit or influence the choices and behaviours of those who want to avoid negative opinions, and by profiling[120], which can lock data subjects into a specific category and restrict them to their suggested preferences[121].

From a different perspective, DPAs use the broad notion of individual self-determination and autonomy to safeguard the free and full development of individual personality and the right to establish and develop relationships with other people. This happens with regard to monitoring communications and online behaviour[122] or the use of video-surveillance systems[123], as they can affect the data subject's freedom to establish and develop relationships with other people[124].

Similarly, negative consequences may also derive from data processing operations involving special categories of individuals, such as minors, or sensitive data. Here, for example, the DPAs mention the publication of exam results by schools, which can cause embarrassment and inhibit free relationships amongst students[125], and the publication by the media of

Use, 3 *Berkeley Technol. L.J.* 117-182, 2016; R. Clarke, *The regulation of civilian drones' impacts on behavioural privacy*, 30 *Computer L. Security Rev.*, 263-285 (2014); D.J. Solove, *A Taxonomy of Privacy*, 154 *U. Pa. L. Rev.* 477-560, 2006.

[118] DPAs have identified a set of limits to employers' monitoring, see, among others, GPDP, Feb. 24, 2010, doc. web n. 1705070; CNIL n. 2009-201, April 16, 2009; ICO, *The employment practices code*, cit., Part. 3; CPVP, avis, n. 8/2006, cit.

[119] See GPDP, Nov. 24, 2016, n. 488, doc. web n. 5796783.

[120] See ART29WP, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, cit.

[121] See also E. Pariser, *The filter bubble. What the Internet is Hiding from You*, 2011.

[122] See GPDP, Dec. 4, 2019, doc. web. n. 9215890; GPDP, June 4, 2015, n. 345, doc. web n. 4211000; GPDP, Feb. 2, 2006, doc. web n. 1229854; ICO, *The employment practices code*, cit., Part. 3.

[123] See ART29WP, *Opinion 2/2009 on the protection of children's personal data (General Guidelines and the special case of schools)*, cit.

[124] With reference to the protection of this right in the work context: see ECtHR, Dec. 16, 1992, *Niemietz v. Germany*, App. No. 13710/88, par. 29; ECtHR, Sept. 5, 2017, *Barbulescu v. Romania*, App. No. 61496/08, pars. 70-73.

[125] See ICO, *The employment practices code*, cit.; See ART29WP, *Opinion 2/2009 on the protection of children's personal data (General Guidelines and the special case of schools)*, cit. On the same grounds the DPAs considered the disclosure of employees' level of performance and evaluation marks, see BFDI, *Notenspiegel im Intranet*.

information concerning children who are victims of violence and abuse[126].

Finally, DPAs consider the role of individual self- determination and autonomy to safeguard the data subject's right to informational self-determination and in order to protect use of their own personal data. This is the case, for example, of mandatory consent to access services or to access them on more advantageous terms (e.g. access to social networks[127], access to certain services and features of IoT devices[128], transport services[129] or energy services[130]).[131]

4.2.6. *Freedom of expression and freedom of thought, conscience and religion*

As in European and international legal systems[132], the DPAs take freedom of expression to include various elements, such as the freedom to hold an opinion[133], or to impart and receive information and ideas[134].

[126] See GPDP, July 10, 2008, doc. web n. 1536583. See also GPDP, Nov. 15, 2001, doc. web n. 39596.

[127] See ART29WP, *Opinion 15/2011 on the definition of consent*, WP 187, 2011.

[128] ART29WP, *Opinion 8/2014 on the on Recent Developments on the Internet of Things*, cit.

[129] CNIL n. 2009-002, Jan. 20, 2009.

[130] See GPDP, Oct. 27, 2016, n. 439, doc. web n. 5687770. With reference to compulsory consent in the context of other services see, for example, DSK, *Novellierung des Personalausweisgesetzes - Änderungen müssen bürger- und datenschutzfreundlich realisiert werden!*, Jan. 24, 2017; DSK, *Wearables und Gesundheits-Apps – Sensible Gesundheitsdaten effektiv schützen!*, April 6-7, 2016; GPDP, July 20, 2017, doc. web n. 6955363; GPDP, May 13, 2015 n. 291, doc. web n. 4337465; ICO, *Direct marketing*, 2018.

[131] Negative consequences may also occur when consent is provided in a situation of power imbalance, such as in the workplace. See CPVP (fn111). See also DSK, *Wearables und Gesundheits-Apps – Sensible Gesundheitsdaten effektiv schützen!,* cit.; ICO, *The employment practices code*, cit., Part. 3; ART29WP, *Opinion 15/2011 on the definition of consent*, cit.; ART29WP, *Opinion 8/2001 on the processing of personal data in the employment context*, WP 48, 2001.

[132] See, among others, D. Bychawska-Siniarska, *Protecting the right to freedom of expression under the European Convention on Human Rights*, Council of Europe, Strasbourg, 2017; Hannikainen and Myntti, *Article 19*, in Asbjørn et al., cit., p. 275; L. Woods, *Article 11 – Freedom of Expression and Information*, in Peers et al. cit., p. 311.

[133] The freedom to hold an opinion concerns the forum internum of a person and protects their thoughts. Individuals must not be indoctrinated by states or other actors. Promoting one-sided information can also be an unacceptable obstacle to the freedom to hold opinions.

[134] Individuals have the freedom to express information and ideas as well as the right to disseminate them. They also have the right to receive any information, opinion, report, or news made public. According to the courts, individual freedom to receive information and ideas also includes the right to be adequately informed, in particular on matters of

According to the DPAs, data subjects' freedom of expression may be constrained, for example, by use of targeted AI systems in political campaigns[135], to influence voters and manipulate outcomes. Similarly, the DPAs pay attention to the interference with freedom that may occur when social net- works automatically block access to a political group's page on the grounds of unverified complaints[136]. Moreover, the DPAs stress the need to safeguard the individual's freedom of expression in relation to fake news and online disinformation, underlining how misleading or false information may influence the public's political opinions[137].

Negative consequences for freedom of expression and, in particular, for the freedom to receive information, may also arise from the publication of incorrect, obsolete information[138], or unreal news by media[139]. In the same way, the DPAs consider individual freedom to receive information in assessing the legitimacy of data subjects' requests to remove or conceal information relating to them because they were unlaw- fully acquired[140].

Finally, the DPAs take into account the potential prejudice to the data subject's freedom of thought, conscience and religion[141] that may derive

public interest. Access to the Internet is seen as included in the freedom of expression by the courts, as a key vehicle for the transmission and reception of information and ideas, see ECtHR, Dec. 1, 2015 (Final Mar. 1, 2016), *Cengiz and Others v. Turkey*, App. No. 48226/10 and 14027/11.

135 See ICO, *Democracy disrupted? Personal Information and political influence*, 2018. See also GPDP, April 18, 2019, doc. web n. 9105201; CNIL n. 2012-021, Jan. 21, 2012.

136 See GPDP, Nov. 26, 2019, doc. web n. 9195349. See also GPDP, Sept. 16, 2019, doc. web n. 9138934.

137 See also EDPS, *Opinion 3/2018 EDPS opinion on online manipulation and personal data*, 2018, <https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf>, accessed on Jan. 18, 2021; Konferenz der Informationsfreiheitsbeauftragten, *Mit Transparenz gegen „Fake-News",* June 13, 2017.

138 See also ECtHR, July 16, 2013 (Final Oct. 16, 2013), *Węgrzynowski and Smolczewski v. Poland*, App. No. 33846/07 (Internet archives were considered to be covered by freedom of expression, as a fundamental source for education and research).

139 See GPDP Sept. 15, 2016, doc. web n. 5515910; GPDP Jan. 24, 2013, doc. web n. 2286820; Autorité de protection des données (Belgian DPA, former Commission de la protection de la vie privée, CPVP), July 14, 2020 (n. DOS-2019-03780); ART29WP, *Guidelines on the implementation of the Court of Justice of the European Union Judgment on "Google Spain and INC V. Agencia Española de Protección de datos (AEPD) and Mario Costeja Gonzàlez"* C-131/12, 2014.

140 See, among others, GPDP, Nov. 26, 2020, doc. web n. 9509558; GPDP, July 10, 2014, doc. web n. 3352396.

141 See article 9 ECHR; article 10 EUCFR; article 18 UDHR. The right to freedom of thought, conscience and religion protects an individual's fundamental beliefs and the right to manifest those beliefs both individually and with others, in both the private

from certain data processing operations. This is the case, for example, of the use of video surveillance systems in places of worship without security purposes or alternative measures[142], as this may condition and limit individuals' activity.

4.2.7. *Freedom of assembly and association*

Though only in a limited number of decisions, DPAs also consider the need to safeguard the data subject's freedom of assembly and association[143]. They recognise that negative outcomes for the data subject's freedom of assembly may result for example from the gathering of identification data of participants in a trade union rally[144], as this may discourage some from taking part in it[145]. Similarly, negative effects on individuals' freedom of assembly may arise from the use of drones by the police and other law enforcement authorities to monitor public demonstrations or similar gatherings[146].

and public sphere. Non-religious viewpoints are also taken into consideration. Thus, the imposition upon individual actions or practices contrary to personal beliefs, as well as restrictions on individual actions or behaviours imposed by belief, will fall within the scope of the guarantee. See J. Murdoch, *Protecting the right to freedom of thought, conscience and religion under the European Convention on Human Rights*, Council of Europe, Strasbourg, 2012, p. 7. See also R. McCrea, *Art. 10 – Right to Freedom of Thought, Conscience and Religion*, in Peers et al., cit., pp. 300-302; M. Scheinin, *Article 18*, in Asbjørn et al., cit., pp. 263-274.

[142] AEPD, expediente n. E/03614/2017; GPDP, Feb. 23, 2017, doc. web n. 6040861. See also ICO, *Wi-fi location analytics*, 2016.

[143] See Council of Europe, *Guide on Article 11 of the European Convention on Human Rights*, 2020, <https://www.echr.coe.int/ Documents/Guide_Art_11_ENG.pdf >, accessed on Feb. 10, 2021.

[144] GPDP, Nov. 29, 2012, doc. web n. 2192643.

[145] See also ECtHR, Feb. 14, 2006 (Final May 14, 2006), *Christian Democratic People's Party v. Moldova*, App. No. 28793/02, par. 77.

[146] ART29WP, *Opinion 01/2015 on Privacy and Data Protection Issues relating to the Utilisation of Drones*, cit., p. 11.

4.2.8. *The right to the confidentiality of communications*

According to DPA jurisprudence, this right[147] is relevant, for example, in cases of firms monitoring their employees' electronic communications (telephone conversations, e-mails, and social media[148]), where the consequences of any breach of confidentiality might affect not only the workers, but also others, such as the worker's family members and the company's customers. Here the authorities stress the need to balance the workers' right to secrecy of correspondence (and that of the other individual involved) with the legitimate rights and interests of the employer.

Furthermore, the DPAs consider the right to confidential communications in relation to monitoring electronic communications and traffic data retention for national security purposes, where any interference with this fundamental right is allowed only if it is strictly necessary in the interests of national security[149].

[147] The right to respect for communications is protected in the international and European regulatory framework as an important aspect of private life, see Article 12 UDHR, Article 7 EUCFR, and Article 8 ECHR. In this context, the protection of communications includes not only correspondence of a personal nature but also that with professional and commercial content. See e.g ECtHR, Sept. 5, 2017, *Bărbulescu v. Romania*, cit., pars. 70-73; ECtHR, May 22, 2008 (Final Aug. 22, 2008), *Ilya Stefanov v. Bulgaria*, App. No. 65755/01.

[148] GPDP, Feb. 1, 2018, doc. web n. 8159221; GPDP, June 4, 2015, n. 345, doc. web n. 4211000; BFDI, *Internet-und E-Mail Nutzung am Arbeitsplatz*; AEPD, Gabinete Jurídico, informe 0464/2013; CPVP, Recommandation n. 08/2012, May 2, 2012; CNIL, *Le contrôle de l'utilisation d'internet et de la messagerie électronique*, 2015; ICO, *The employment practices code*, cit., Part. 3; ART29WP, *Working document on the surveillance of electronic communications in the workplace*, cit.

[149] ART29WP, *Opinion 04/2014 on surveillance of electronic communications for intelligence and national security purposes*, cit.; ART29WP, *Opinion 4/2005 on the Proposal for a Directive of the European Parliament and of the Council on the Retention of Data Processed in Connection with the Provision of Public Electronic Communication Services and Amending Directive 2002/58/EC (COM(2005)438 final of 21.09.2005)*, WP 113, adopted on Oct. 21, 2005.

5. *A proposal for an HRIA model*

The analysis described in Section 4 and the evidence provided by DPAs' decisions and practice show that the issues concerning the use of data-intensive applications are not circumscribed to the debated topic of bias and discrimination but have a broader impact on several human rights and freedoms. For this reason, a comprehensive HRIA model is needed.

It is worth noting that traditional HRIAs are often territory-based considering the impact of business activities in a given local area and community, whereas in the case of AI applications this link with a territorial context may be less significant.

There are two different scenarios: cases characterised by use of AI in territorial contexts with a high-impact on social dynamics (e.g. smart cities plans, regional smart mobility plans, predictive crime programmes) and those where AI solutions have a more limited impact as they are embedded in globally distributed products/services (e.g. AI virtual assistants, autonomous cars, recruiting AI-based software, etc.) and do not focus on a given socio-territorial community. While in the first case the context is very close to the traditional HRIA cases, where large-scale projects affect whole communities and the potential impacts cover a wide range of human rights, the second case is characterised by a more limited social impact, centred more on individuals rather than on society at large[150]. This difference has a direct effect on the structure and complexity of the model, as well as the tool employed.

Criteria such as the AAAQ framework[151], for example, or issues concerning property and lands, can be used in assessing a smart city plan, but are unnecessary or disproportionate in the case of an AI-based recruitment software. Similarly, a large-scale mobility plan may require a significant monitoring of needs through interviews of right holders and stakeholders, while in the case of an AI-based personal IoT device this phase can be much reduced.

Regarding the first and more complex scenarios, we provide only a limited contribution, as the existing HRIA models can be used in those cases. Here, a greater focus on data- intensive systems leads to a reflection

[150] This does not mean that the collective dimension does not plays an important role and should be adequately considered in the assessment process, see A. Mantelero, *Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection*, 32 *Computer L. & Sec. Rev*. 238-255, 2016.

[151] See Danish Institute for Human Rights, *The AAAQ Framework and the Right to Water: International indicators for availability, accessibility, acceptability and quality*, 2014.

on the challenges that large-scale poses with regard to multi-factor scenarios (Section 6.2).

We expect our model to make more significant contribution in the second scenario, where the traditional eighth/twelve-month HRIA should be scaled down to a more manageable size for small-scale projects, and focused on quantifiable criteria to be applied in AI product and service development.

The model described here consists of two main building blocks, which are examined in detail in the following subsections: planning and scoping, and data collection and analysis.

5.1. *Planning and scoring*

The first stage deals with definition of the HRIA target, identifying the main features of the product/service and the context in which it will be placed, in line with the context-dependant nature of the HRIA. Three are the main areas to consider at this stage:

- description and analysis of the type of product/service, including data flows and data processing purposes
- the human rights context (contextualisation on the basis of local jurisprudence and laws)
- identification of relevant stakeholders.

The table below (Table 1) provides a non-exhaustive list of potential questions for HRIA planning and scoping. The extent and content of these questions will depend on the specific nature of the product/service and the scale and complexity of its development and deployment[152]. This list is therefore likely to be further supplemented with project-specific questions[153].

[152] See e.g. The Danish Institute for Human Rights, *Scoping practitioner supplement. Human rights impact assessment guidance and toolbox*, Copenhagen, 2020; The Danish Institute for Human Rights, *Guidance on HRIA of Digital Activities. Phase 1: Planning and scoping*, Copenhagen, 2020, <https://www.humanrights.dk/sites/humanrights.dk/files/media/document/Phase%201_%20Planning%20and%20Scoping_n.pdf >, accessed on 20 February 2021.

[153] For similar questionnaires, see e.g. p. 23-30 The Danish Institute for Human Rights, *Human Rights Impact Assessment Guidance and Toolbox*, Copenhagen, 2016, <https://www.humanrights.dk/business/tools/human-rights-impact-assessment-guidance-and-toolbox>, accessed on June 18, 2020.

Table 1 – Planning & scoping	
Description and analysis of the type of product/service, including related data flows and data processing purposes	What are the main features of the product/service? • In which countries will the product/service be offered? • Identification of rights-holders: who are the target-users of the product/service? • What types of data are collected (personal, non-personal, special categories)? • What are the main purposes of data processing? • Identification of the duty-bearers: which subjects are involved in data management and what is their role in data processing?
Human rights context (contextualisation based on local jurisprudence and laws)	• Which human rights are potentially affected by the product/service? • Which international/regional legal instruments have been implemented at an operational level? • Which are the most relevant courts or authoritative bodies in the field of human rights in the context? • What are the relevant decisions and provisions in the field of human rights?
Controls in place	• What policies and procedures are in place to assess the potential impact on human rights, including stakeholder engagement? • Has an impact assessment been carried out, developed and implemented in relation to specific issues or some features of the product/service (e.g. use of biometrics)?
Stakeholder engagement	• Which are the main groups or communities potentially affected by the service/product, including its development? • What other stakeholders should be involved, in addition to affected community and groups, (e.g. civil society and international originations, experts, industry associations, journalists)? • Are there any other duty-bearers to be involved, apart from the product/service developer (e.g. national authorities, governmental agencies)? • Were business partners, including suppliers (e.g. subcontractors in AI systems and datasets) involved in the assessment process? • Has the developer conducted an assessment of its supply chain to identify whether the activities of suppliers/contractors involved in product/service development might contribute to adverse human rights impacts? Has the developer promoted human rights standards or audits to ensure respect for human rights amongst suppliers? • Do the product/service developers publicly communicate the potential impacts on human rights of the service/product? • Does the developer provide training on human rights standards for relevant management and procurement staff?

factors must be considered: risk identification, likelihood (L), and severity (S). As regards the first, the focus on human rights and freedoms already defines the potentially affected categories and the case specific analysis identifies those concretely affected, depending on the technologies used and their purposes. Since this is a rights-based model, risk concerns the prejudice to rights and freedoms, in terms of unlawful limitations and restrictions, regardless of material damage.

The expected impact of the identified risks is assessed by considering both the likelihood and the severity of the expected consequences, using a four-step scale (low, medium, high, very high) to avoid any risk of average positioning.

Likelihood is the combination of two elements: the probability of adverse consequences and the exposure. The former concerns the probability that adverse consequences of a certain risk might occur (Table 2) and the latter the potential number of people at risk (Table 3). In considering the potential impact on human rights, it is important not only to consider the probability of the impact, but also its extension in terms of potentially affected people.

Table 2 – Probability.

	Probability	
Low	The risk of prejudice is improbable or highly improbable	1
Medium	The risk may occur	2
High	There is a high probability that the risk occurs	3
Very high	The risk is highly likely to occur	4

Table 3 – Exposure.

	Exposure	
Low	Few or very few of the identified population of rights-holders are potentially affected	1
Medium	Some of the identified population are potentially affected	2
High	The majority of the identified population is potentially affected	3
Very high	Almost the entire identified population is potentially affected	4

assessment or to estimate the overall impact. Moreover, the model is used for an ex post comparative analysis, rather than for iterative design-based product/service development, as does the model we present here. In this sense, by providing two fictitious basic cases, Janssen tests her model through a comparative analysis (one case against the other) and without a clear analysis of the different risk components, in terms of individual impact and probability, with regard to each potentially affected right or freedom (e.g. "given that the monitor sensor captures every noise in its vicinity in situation (1), it probably has a high impact on a number of privacy rights, including that of intimacy of the home, communication privacy and chilling effects on the freedom of speech of (other) dwellers in the home"), and without a clear description of the assessment of their cumulative effect and overall impact. With a focus on the GDPR, see M. E Kaminski and G. Malgieri, *Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations*, in *Int'l Data Privacy L.*, 2020, DOI: 10.1093/idpl/ipaa020. See also D. Reisman and others, *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, 2018, <https://ainowinstitute.org/aiareport2018.pdf >, accessed on June 29, 2018.

Both these variables must be assessed on a contextual basis, considering the nature and features of the product and service, the application scenario, previous similar cases and applications, and any measures taken to prevent adverse consequences. Here, the engagement of relevant shareholders can help to better understand and contextualise these aspects, alongside the expertise of those carrying out the impact assessment.

These two variables are combined in the combinatorial table (Table 4) using a cardinal scale to estimate the overall likelihood level (L). This table can be further modified on the basis of the context-specific nature of assessed AI systems and feedback received from experts and stakeholders.

Table 4 – Likelihood table (L).

Exposure	Probability 1	2	3	4
1	1	2	3	4
2	2	3	5	9
3	3	5	9	12
4	4	7	12	15

Likelihood	
Low	1
Medium	2
High	3
Very high	4

Table 5 – Gravity of the prejudice.

	Gravity of the prejudice	
Low	Affected individuals and groups may encounter only minor prejudices in the exercise of their rights and freedoms.	1
Medium	Affected individuals and groups may encounter significant prejudices.	2
High	Affected individuals and groups may encounter serious prejudices.	3
Very high	Affected individuals and groups may encounter serious or even irreversible prejudices.	4

Table 6 – Effort to overcome the prejudice and to reverse adverse effects.

	Effort	
Low	Suffered prejudice can be overcome without any problem (e.g. time spent amending information, annoyances, irritations, etc.)	1
Medium	Suffered prejudice can be overcome despite a few difficulties (e.g. extra costs, fear, lack of understanding, stress, minor physical ailments, etc.).	2
High	Suffered prejudice can be overcome albeit with serious difficulties (e.g. economic loss, property damage, worsening of health, etc.).	3
Very high	Suffered prejudice may not be overcome (e.g. long-term psychological or physical ailments, death, etc.).	4

Table 7 – Severity table (S).

		Gravity			
		1	2	3	4
Effort	1	1	2	4	6
	2	2	3	5	8
	3	3	5	8	10
	4	5	8	10	12

Severity	
Low	1
Medium	2
High	3
Very high	4

Table 8 – Table of envisaged risks.

	L	S	Overall impact
R1			
R2			
….			
Rn			

The severity of the expected consequences (S) is estimated by considering the nature of potential prejudice in the exercise of rights and freedoms and their consequences. This is done by considering the gravity of the prejudice (gravity, Table 5), and the effort to overcome it and to reverse adverse effects (effort, Table 6).

As in the case of likelihood, these two variables are combined in the Severity table (Table 7) using a cardinal scale to estimate the severity level (S).

A table (Table 8) for the overall assessment charts both variables – likelihood (L) and severity (S) of the expected consequences – against each envisaged risk to rights and freedoms (R1, R2,… Rn).

The overall impact for each examined risk, taking into consideration the L and S values, is determined using further table (Table 9). The colours represent the overall impact, which is very high in the dark red sector, high in the red sector, medium in the yellow sector and is low in the green sector.

Table 9 – Overall risk impact table

		Severity [impacted right/freedom]			
		Low	Medium	High	Very high
Likelihood	Low				
	Medium				
	High				
	Very high				

Once the potentially adverse impact has been assessed for each of the rights and freedoms considered, a radial graph is charted to represent the overall impact on them (Figure 1). This graph is then used to decide the priority of intervention in altering the characteristics of the product/service to reduce the expected adverse impacts[156].

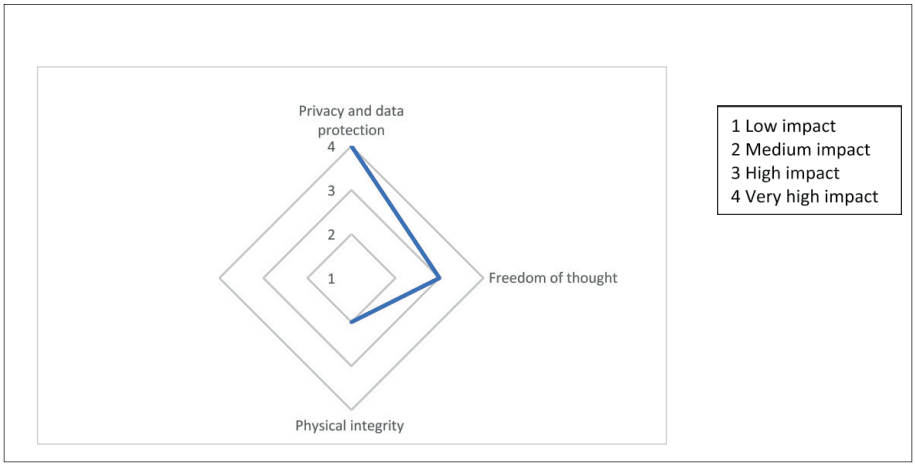


Fig. 1 – Radial graph (impact) example

To reduce the envisaged impacts, factors that can exclude the risk from a legal perspective (EFs) – such as the mandatory nature of certain impacting features or the prevalence of competing interests recognised by law – and those that can reduce the risk by means of appropriate mitigation measures (MMs) should be considered.

After the first adoption of the appropriate measures to mitigate the risk, further rounds of assessment can be conducted according to the level of residual risk and its acceptability, enriching the initial table with new columns (Table 10).

Table 10 – Comparative risk impact analysis table (before/after mitigation measures and excluding factors).

	L	S	Overall impact	EFs	MMs	rL	rS	Final impact
R1								
R2								
…								
Rn								

[156] This approach is also in line with the adoption of the Agile methodology in software development.

The first two new columns show any risk excluding factors (EFs) and mitigation measures (MMs), while the following two columns show the residual likelihood (rL) and severity (rS) of the expected consequences, after accounting for excluding and mitigation factors. The last column gives the final overall impact, using rL and rS values and the overall impact table (Table 9); this result can also be represented in a new radial graph.

Note that it is possible to estimate overall impact, as an average of the impacts on all the areas analysed. But this necessarily treats all the different impacted areas (i.e. rights and freedoms) as having the same importance and is therefore a somewhat imprecise synthesis.

In terms of actual effects on operations, the radial graph is therefore the best tool to represent the outcome of the HRIA, showing graphically the changes after introducing mitigation measures. However, an estimation of overall impact could also be made in future since several legislative proposals on AI refer to an overall impact of each AI-based solution, using a single risk scale covering all potential consequences.

6. *Testing the HRIA*

The next two sub-sections examine two possible applications of the proposed model, with two different scales of data use. The first case, an Internet-connected doll equipped with AI, shows how the impact of AI is not limited to adverse effects on discrimination, but has a wider range of consequences (privacy and data protection, education, freedom of thought and diversity, etc.), given the innovative nature of the application and its interaction with humans.

This highlights the way in which AI does not merely concern data and data quality but more broadly the transformation of human-machine interaction by data-intensive systems. This is even more evident in the case of the smart cities, where the interaction is replicated on large scale affecting a whole variety of human behaviours by individuals, groups and communities.

The first case study (an AI-powered doll) shows in detail how the HRIA methodology can be applied in a real-life scenario. In the second case (a smart city project) we do not repeat the exercise for all the various data-intensive components, because a full HRIA would require extensive information

collection, stakeholder engagement, and supply-chain analysis[157], which go beyond the scope of this work[158]. But above all, the purpose of this second case study is different: to shed light on the dynamics of the HRIA in multi-factor scenarios where many different AI systems are combined.

Indeed, a smart city environment is not a single device, but encompasses a variety of technical solutions based on data and algorithms. The cumulative effect of integrating many layers results in a whole system that is greater and more complicated than the sum of its parts.

This explains why the assessment of potential risks to human rights and freedoms cannot be limited to a fragmented case-by-case analysis of each application. Rather, it requires an integrated approach that looks at the whole system and the interaction amongst its various components, which may have a wider impact than each component taken separately.

Scale and complexity, plus the dominant role of one or a few actors, can produce a cumulative effect which may entail multiple and increased impacts on rights and freedoms, requiring an additional integrated HRIA to give an overall assessment of the large-scale project and its impacts.

6.1. *Testing HRIA on a small scale: the Hello Barbie case*

Hello Barbie was an interactive doll produced by Mattel for the English-speaking market, equipped with speech recognition systems and AI-based learning features, operating as an IoT device. The doll was able to interact with users but did not interact with other IoT devices[159].

The design goal was to provide a two-way conversation be- tween the doll and the children playing with it, including capabilities that make

[157] See also K. Crawford and V. Joler, *Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources*, AI Now Institute and Share Lab, 2018, <http://www.anatomyof.ai>, accessed on Dec. 27, 2019.

[158] A proper HRIA would require a multidisciplinary team working locally for a significant period of time. For example, the human rights impact assessment of the Bisha Mine in Eritrea, which started in July 2013, issued its final HRIA report in February 2014, followed by an auditing procedure in 2015. See LKL International Consulting Inc., *Human Rights Impact Assessment of the Bisha Mine in Eritrea*, 2014, <https://media.business-humanrights.org/media/documents/ files/documents/Nevsun_HRIA_Full_ReportApril_2014_.pdf >, accessed Oct. 26, 2020; LKL International Consulting Inc., *Human Rights Impact Assessment of the Bisha Mine in Eritrea 2015 Audit*, 2015, <https://media.business- humanrights.org/ media/documents/files/documents/Bisha-HRIA- Audit- 2015.pdf >, accessed on Oct. 26, 2020.

[159] See Mattel, *Hello Barbie FAQ*, Version 2, 2015, <http://hellobarbiefaq.mattel.com/faq/>, accessed on Nov. 12, 2020.

the doll able to learn from this interaction, e.g. tailoring responses to the child's play history and remembering past conversations to suggest new games and topics[160].

The doll is no longer marketed by Mattel due to several concerns about system and device security[161].

This section discusses the hypothetical case, imagining how the proposed assessment model[162] could have been used by manufactures and developers and the results that might have been achieved.

6.1.1. *Planning and scoring*

Starting with the questions listed in Table 1 above and information on the case examined, the planning and scoping phase would summarise the key product characteristics as follows:

a) A connected toy with four main features: (i) programmed with more than 8000 lines of dialogue[163] hosted in the cloud, enabling the doll to talk with the user about "friends, school, dreams and fashion"[164]; (ii) speech recognition technology[165] activated by a push-and-hold but- ton on the doll's belt buckle; (iii) equipped with a microphone, speaker and two tricolour LEOs embedded in the doll's necklace, which light up when the device is active; (iv) a Wi-

[160] *Id.*

[161] See also S. Shasha et al., *Playing With Danger: A Taxonomy and Evaluation of Threats to Smart Toys*, in *IEEE Internet of Things Journal*, vol. 6, 2019, pp. 2986-3002 (with regard to Hello Barbie see Appendix A, par. A.3).

[162] On the safeguard of human rights and the use of HRIA in the business context see United Nations, *Guiding Principles on Business and Human Rights*, 2011, <https://www.ohchr.org/ Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf >, accessed on Dec. 8, 2020 ("The State duty to protect is a standard of conduct. Therefore, States are not per se responsible for human rights abuse by private actors. However, States may breach their international human rights law obligations where such abuse can be attributed to them, or where they fail to take appropriate steps to prevent, investigate, punish and redress private actors' abuse") and more specifically Principles 13, 18 and 19. See also The Danish Institute for Human Rights, *Human Rights Impact Assessment Guidance and Toolbox*, cit., pp. 6-7.

[163] The comprehensive list of all the lines Hello Barbie says as of Nov. 17, 2015 is available here: <http://hellobarbiefaq.mattel.com/wp- content/uploads/2015/11/hellobarbie-lines-v2.pdf >, accessed on Nov. 28, 2020.

[164] Mattel, *Hello Barbie FAQ*, cit. Cloud service was provided by ToyTalk, see the following footnote.

[165] This technology and services were provided by ToyTalk, a Mattel's partner.

Fi connection to provide for two-way conversation[166].

b) The target-user is an English-speaking child (minor). Theoretically the product could be marketed worldwide in many countries, but the language barrier represents a limitation.

c) The right-holders can be divided into three categories: direct users (minors), supervisory users (parents, who have partial remote control over the doll and the doll/user interaction) and third parties (e.g. friends of the user or re-users of the doll).

d) Regarding data processing, the doll collects and stores voice-recording tracks based on dialogues between the doll and the user; this information may include personal data[167] and sensitive information[168].

e) The main purpose of the data processing and AI is to create human-robot interaction (HRI) by using machine learning (ML) to build on the dialogue between the doll and its young users. There are also additional purposes: (i) educational; (ii) parental control and surveillance[169] (parents can listen, store and re-use recorded

[166] Mattel, *Hello Barbie FAQ*, cit.

[167] *Id.* ("Q: Can Hello Barbie say a child's name? No. Hello Barbie does not ask for a child's name and is not scripted to respond with a child's name, so she will not be able to recite a child's name back to them"). But see M. Leta Jones, *Your New Best Frenemy: Hello Barbie and Privacy Without Screens*, in *Engaging Science, Technology, and Society*, vol. 2, 2016, pp. 242, 245, who reports this reply in the dialogue with the doll: "Barbie: Sometimes I get a little nervous when I tell people my middle name. But I'm really glad I told you! What's your middle name?".

[168] Mattel, *Hello Barbie FAQ*, cit. ("Although Hello Barbie was designed not to ask questions which are intended to elicit answers that might contain personal information, we cannot control whether a child volunteers such information without prompting. Parents who are concerned about this can monitor their child's use of Hello Barbie, and parents have the power to review and delete any conversation their child has with Hello Barbie, whether the conversations contain personal information or not. If we become aware of any such personal information captured in recordings, it is our policy to delete such information, and we contractually require our Service Providers to do the same. This personal information is not used for any purpose").

[169] *Id.* ("Hello Barbie only requires a parent's email address to set up an account. This is necessary so that parents can give permission to activate the speech recognition technology in the doll. Other information, such as a daughter's birthday, can be provided to help personalize the experience but are not required"). See also note 170.

conversations)[170]; (iii) direct advertising to parents[171]; (iv) testing and service improvement[172].

f) The chief duty-bearer is the producer, but in connected toys other partners – such as ToyTalk in this case – may be involved in the provision of ML, cloud and marketing services.

Another important set of data to be collected at this stage concerns the potential interplay with human rights and the reference framework, including main international/regional legal instruments, relevant courts or other authoritative bodies, and relevant decisions and provisions (see Table 1, the human rights context).

As regards the rights potentially affected, depending on the product's features and purposes, data protection and the right to privacy are the most relevant due to the possible content of the dialogue between the doll and the user, and the parental monitoring. Here the legal framework is represented by a variety of regulations at different levels. Compliance

[170] *Id.* ("Hello Barbie recording and storing conversations girls have with the doll? Yes. Hello Barbie has conversations with girls, and these conversations are recorded. These audio recordings are used to understand what is being said to Hello Barbie so she can respond appropriately and also to improve speech recognition for children and to make the service better. These conversations are stored securely on ToyTalk's server infrastructure and parents have the power to listen to, share, and/or delete stored recordings any time").

[171] *Id.* ("Q. Are conversations used to market to children? No. The conversations captured by Hello Barbie will not be used to contact children or advertise to them." This was confirmed by the analysis carried out by Shasha et al., *Playing With Danger: A Taxonomy and Evaluation of Threats to Smart Toys*, cit. Regarding the advertising directs to parents, this is the answer provided in the FAQ: "Q: Your Privacy Policy says that you will use personal information to provide consumers with news and information about events, activities, promotions, special offers, etc. That sounds like consumers could be bombarded with marketing messages. Can parents elect not to receive those communications? Yes. Opting out of receiving promotional emails will be an option during the set up process and you can opt out at any time by following the instruction in those emails. Note that marketing messages will not be conveyed via the doll itself").

[172] *Id.* ("Conversations between Hello Barbie and consumers are not monitored in real time, and no person routinely reviews those conversations. Upon occasion a human may review certain conversations, such as in order to test, improve, or change the technology used in Hello Barbie, or due to support requests from parents. If in connection with such a review we come across a conversation that raises concern about the safety of a child or others, we will cooperate with law enforcement agencies and legal processes as required to do so or as we deem appropriate on a case-by-case basis").

with the US COPPA[173] and the EU GDPR[174] can cover large parts of the potential market of this product and international guiding principles[175] can facilitate the adoption of global policies and solutions.

Moreover, in relation to data processing and individual freedom of choice, the potential effects of marketing strategies can also be considered as forms of freedom of expression[176] and freedom to conduct a business.

Given the broad interaction between the doll and the user and the behavioural, cultural and educational influence that the doll may have on young users[177], further concerns relate to freedom of thought and diversity[178].

---

[173] See Federal Trade Commission, *Enforcement Policy Statement Regarding the Applicability of the COPPA Rule to the Collection and Use of Voice Recordings*, Oct. 23, 2017, <https://www.ftc.gov/public-statements/2017/10/federal-trade-commission-enforcement-policy-statement-regarding>, accessed Nov. 28, 2020. See also E. Haber, *Toying with Privacy: Regulating the Internet of Toys*, 80 *Ohio State L. J.* 399, 2019.

[174] See also ICO, *Age appropriate design: a code of practice for online services*, section 14. Connected toys and devices 2020, <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services/>, accessed Feb. 20, 2021.

[175] See e.g. Council of Europe, *Convention 108+*. See also Council of Europe, *Recommendation CM/Rec(2018)7 of the Committee of Ministers. Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment*, <https://rm.coe.int/guidelines-to-respect-protect-and-fulfil-the-rights-of-the-child-in-th/16808d881a>, accessed on Nov. 28, 2020, par. 36 ("With respect to connected or smart devices, including those incorporated in toys and clothes, States should take particular care to ensure that data-protection principles, rules and rights are also respected when such products are directed principally at children or are likely to be regularly used by or in physical proximity to children"). See also A. Mantelero, *The future of data protection: Gold standard vs. global standard*, in *Computer L. & Security Rev.*, vol. 40, 2021, DOI: 10.1016/j.clsr.2020.105500.

[176] See *Universal Declaration of Human Rights*, article 19, and *International Covenant on Civil and Political Rights*, article 19(2). See also Human Rights Committee, *General Comment no. 34 (CCPR/C/GC/34)*, par. 11; UNICEF et. al., *Children's Rights and Business Principles*, 2012, <https://d306pr3pise04h. cloudfront.net/docs/issues_doc%2Fhuman_rights%2FCRBP%2FChildrens_Rights_and_Business_Principles.pdf>, accessed on Nov. 30, 2020, principle 6 (Use marketing and advertising that respect and support children's rights).

[177] See P. Mertala, *How Connectivity Affects Otherwise Traditional Toys? A Functional Analysis of Hello Barbie*, in *Int. J. Child. Comput. Interact.*, vol. 25, 2020, 25DOI: 10.1016/j.ijcci.2020.100186 ("As Hello Barbie is able to speak, the child no longer performs the role through the doll, but in relation to the doll. This changes the nature of the performative element from dominantly transitive to dominantly performative, in which the child occupies and embodies a role in relation to the toy"). See also the following statement included in the list of all the lines Hello Barbie says as of Nov. 17, 2015, *supra* note 163) "It's so cool that you want to be a mom someday".

[178] See Mattel, *Hello Barbie FAQ*, cit. ("The doll's conversation tree has been designed

In the event of cyberattack and data theft or transmission of inappropriate content to the user through the doll, safety issues also arise and may impact on the right to psychological and physical safety and health.

With the potentially global distribution of the toy, the possible impacts need to be further contextualised within each legal framework, taking into consideration local case law and that of regional supranational bodies like the European Court of Human rights. In this regard, it is necessary during the scoping phase to identify the significant provisions and decisions in the countries/regions where the product is distributed.

The last aspect to be considered in planning and scoping HRIA concerns the identification and engagement of potential stakeholders. In the case of connected toys, the most important stakeholders are likely to be parents' associations, educational bodies, professional associations (e.g. psychologists and educators), child, consumer and data protection supervisory bodies, as well as trade associations. Stakeholders may also include the suppliers involved in product/service development. In the latter case, the HRIA must also assess the activities by these suppliers and may benefit from an auditing procedure or the adoption of standards.

The following sections describe an iterative assessment process, starting from the basic idea of the connected AI-equipped toy with its pre-set functionality and moving on to a further assessment considering additional measures to mitigate unaddressed, or only partially addressed, concerns.

6.1.2. *Initial risk analysis and assessment*

The basic idea of the toy is an interactive doll, equipped with speech recognition and learning features, operating as an IoT device. The main component is a human-robot voice interaction feature based on AI and enabled by Internet connection and cloud services.

The rights potentially impacted are data protection and privacy, freedom

to re-direct inappropriate conversations. For example, Hello Barbie will not repeat curse words. Instead, she will respond by asking a new question"). However, besides the example given, there is no clear description of what is considered appropriate or not, and this category (appropriateness) is significantly influenced by the cultural component and potentially also by corporate ethics that may create forms of censorship or oriented behavior and thinking in the young user. Even when the FAQs refer to "school age appropriate content" ("All comments made by Hello Barbie are scripted with school age appropriate content"), they implicitly refer to a benchmark dependent the educational standards of developed economies.

of thought and diversity, and psychological and physical safety and health[179].

6.1.2.1. *Data protection and the right to privacy*

While these are two distinct rights, for the purpose of this case study we considered them together[180]. Given the main product features, the impact analysis is based on following questions[181]:

- Does the device collect personal information? If yes, what kind of data is collected, and what are the main features of data processing? Can the data be shared with other entities/persons?
- Can the connected toy intrude into the users' private sphere?
- Can the connected toy be used for monitoring and surveillance purposes? If yes, is this monitoring continuous or can the user stop it?
- Do users belong to vulnerable categories (e.g. minors, elderly people, etc.)?
- Are third parties involved in the data processing?
- Are transborder data flows part of the processing operations?

Taking into account the product's nature, features and settings (i.e. companion toy, dialogue recording, personal information collection, potential data sharing by parents) the likelihood of prejudice can be considered very high (Table 4). The extent and largely unsupervised nature of the dialogue be- tween the doll and the user, as well as the extent of data collection and retention make the probability high (Table 2). In addition, given its default features and settings, the exposure is very high (Table 3) since all the doll's users are potentially exposed to this risk.

Regarding risk severity, the gravity of the prejudice (Table 5) is high, given the subjects involved (young children and minors), the processing of personal data in several main areas, including sensitive information[182],

[179] See E. Keymolen and S. Van der Hof, C*an I still trust you, my dear doll? A philosophical and legal exploration of smart toys and trust*, in *Journal of Cyber Policy*, vol. 4, 2019, pp. 143-159 ("Smart toys come in different forms but they have one thing in common. The development of these toys is not just a feature of ongoing technological developments; their emergence also reflects an increasing commercialisation of children's everyday lives").

[180] See also *UN Convention on the Rights of the Child*, Article 16; *European Convention on Human Rights*, Article 8.

[181] For a more extensive list of guiding questions, see e.g. UNICEF, Children's Online Privacy and Freedom of Expression, 2018, <https://www.unicef.org/csr/files/UNICEF_Childrens_Online_Privacy_and_Freedom_of_Expression(1).pdf>, accessed on Dec. 18, 2020.

[182] Pre-recorded sentences containing references to, for instance, religion and ethical

and the extent of data collection. In addition, unexpected findings may emerge in the dialogue between the user and the doll, as the harmless topics prevalent in the AI-processed sentences can lead young users

to provide personal and sensitive information. Furthermore, the data processing also involves third parties and transborder data flows, which add other potential risks.

The effort to overcome potential prejudice or to reverse adverse effects (Table 6) can be considered as medium, due to the potential parental supervision and remote control, the nature of the doll's pre-selected answers and the adoption of standard data security measures that help to overcome suffered prejudice with a few difficulties (e.g. data erasure, dialogue with the minor in case of unexpected findings). Combining high gravity and medium effort, the resulting severity (Table 7) is medium.

If the likelihood of prejudice can be considered very high and the severity medium, the overall impact according to Table 9 is high.

6.1.2.2.*Freedom of thought, parental guidance and the best interest of the child*

- Based on the main features of the product, the following questions can be used for this analysis:
- Is the device able to transmit content to the user?
- Which kind of relationships is the device able to create with the user?
- Does the device share any value-orientated messages with the user?
    - If yes, what kind of values are communicated?
    - Are these values customisable by users (including parents) or on the basis of user interaction? If so, what range of alterative value sets is provided?
    - Are these values the result of work by a design team characterised by diversity?

Here the case study reveals the critical impact of AI on HRI owing to the potential content imparted through the device. This is even more critical in the context of toys where the interactive nature of AI-powered dolls changes the traditional interaction into a relational experience[183].

groups. See the full list of all lines for Hello Barbie, *supra* note 163, (e.g. "Sorry, I didn't catch that. Was that a yes or a no to talking about Kwanzaa?").

[183] Mertala, *How Connectivity Affects Otherwise Traditional Toys? A Functional Analysis*

In the model considered (Hello Barbie), AI creates a dialogue with the young user by selecting the most appropriate sentence from the more than 8000 lines of dialogue available in its database. On the one hand, this enables the AI to express opinions which may also include value-laden messages, as in this sentence: "It's so cool that you want to be a mom someday"[184]. On the other, some value-based considerations are needed to address educational issues concerning "inappropriate questions"[185] where the problem is not the AI reaction (Hello Barbie responds "by asking a new question"[186]), as previously, but the notion of appropriateness, which necessarily involves a value-orientated content classification by the AI system.

As these value-laden features of AI are inevitably defined during the design process, the composition of the design team, its awareness of cultural diversity and pluralism are key elements that impact on freedom of thought, in terms of default values proposed and the availability of alternative settings. In addition, the decision to provide only one option or several user-customisable options in the case of value-orientated content is another aspect of the design phase that can limit parents' freedom to ensure the moral and religious education of their children in accordance with their own beliefs.

This aspect highlights the paradigm shift brought by AI to freedom of thought and the related parental guidance in supporting the exercise by children of their rights[187]. This is even more evident when comparing AI-equipped toys with traditional educational products, such as books, serious games etc., whose contents can be examined in advance by parents[188].

The AI-equipped doll is different. It delivers messages to young users, which may include educational content and information, but no parent will read all the 8000 lines the doll can use or ask to have access to the

of Hello Barbie, cit.

[184] See *supra* note 163. On gender stereotypes in smart toys, see Norwegian Consumer Council, *#Toyfail An analysis of consumer and privacy issues in three internet-connected toys*, 2016, <https://fil.forbrukerradet.no/wp- content/uploads/2016/12/toyfail- report-desember2016.pdf >, accessed on Dec. 14, 2020.

[185] See Mattel, *Hello Barbie FAQ*, cit. *supra* note 178.

[186] See Mattel, *Hello Barbie FAQ*, cit. *supra* note 159.

[187] See *UN Convention on the Rights of the Child*, Articles 5, 14, and 18. See also See UNICEF, *Children's Online Privacy and Freedom of Expression*, 2018, p. 9, <https://www.unicef.org/csr/files/UNICEF_ Childrens_Online_Privacy_and_Freedom_of _ Expression(1).pdf>, accessed on Dec. 18, 2020; J. Murdoch, *Protecting the Right to Freedom of Thought, Conscience and Religion under the European Convention on Human Rights*, Council of Europe 2012, p. 13.

[188] See also *UN Convention on the Rights of the Child*, Articles 17(e) and 18.

logic used to match them with children's statements.

As AI-based devices interact autonomously with children and convey their own cultural values[189], this impacts on the rights and duties of parents to provide, in a manner consistent with the evolving capacities of the child, appropriate direction and guidance in the child's freedom of thought, including aspects concerning cultural diversity.

In terms of risk assessment, the probability (Table 2) is medium, considering the limited number of sentences involving a value-orientated statement and the exposure (Table 3) is medium, due to their alignment with values commonly accepted in many cultural contexts. The likelihood is therefore medium (Table 4).

Taking into account the nature of the product and its main features (i.e. some value-laden sentences used in dialogue with the young user)[190] the gravity of prejudice (Table 5) can be considered low in the case in question, as the value-laden sentences concern cultural questions that are not particularly controversial. The effort (Table 6) can also be considered low, as talking with children can mitigate potential harm. Combining these two values, the severity is therefore low (Table 7).

Note that this assessment would be completely altered if the dialogue content were not pre-selected, but generated by AI on the basis of information resulting from web searches[191], where the potential risk would be much higher. Similarly, the inclusion in the pre-recorded database of a greater number of value-laden sentences would directly increase the risk.

Considering the likelihood as medium and the severity of the prejudice as low, the overall impact (Table 9) is medium.

6.1.2.3. *Right to psychological and physical safety*

Connected toys may raise concerns about a range of psychological and physical harms deriving from their use, including access to data and remote

[189] See e.g. Norwegian Consumer Council, *#Toyfail An analysis of consumer and privacy issues in three internet-connected toys*, cit., referring to the connected doll Cayla ("Norwegian version of the apps has banned the Norwegian words for "homosexual", "bisexual", "lesbian", "atheism", and "LGBT" […]" "Other censored words include 'menstruation', 'scientology-member', 'violence', 'abortion', 'religion', and 'incest' ").

[190] See V. Steeves, *A dialogic analysis of Hello Barbie's conversations with children*, in *Big Data & Soc'y*, vol 7, n.1, 2020, DOI:10.1177/2053951720919151.

[191] In the case examined, the content provided by means of the doll was handcrafted by the writing team at Mattel and ToyTalk, not derived from open web search. See Mattel, *Hello Barbie FAQ*, cit.

control of the toy[192]. Based on the main features of the product examined, the following questions can be used for this analysis:

- Can the device put psychological or physical safety at risk?
- Does the device have adequate data security and cybersecurity measures in place?
- Can third parties perpetrate malicious attacks that pose a risk to the psychological or physical safety of the user?

As regards the probability, considering the third-party origin of the prejudices and the limited interest in malicious attacks (no business interest, distributed and generic target), but also how easy it is to hack the toy, the probability (Table 2) of an adverse impact is medium. Exposure (Table 3) is low, given the prevalent use of the device in a supposedly safe environment, such as schools and home, where malicious access and control of the doll is difficult and adult monitoring is more frequent. The likelihood (Table 4) is therefore low.

Taking into account the nature of the product examined, the young age of the user, and the potential safety and security risks[193], the gravity of prejudice (Table 5) can be considered medium. This is because malicious attacks can only be carried out by speech, and no images are collected. Nor can the toy – given its size and characteristics – directly cause physical harm to the user. The effort (Table 6) can be considered medium since parent-child dialogue and technical solutions can combat the potential prejudice. The severity (Table 7) is therefore medium.

Considering the likelihood as low and the severity of the prejudice as medium, the overall impact is medium (Table 9).

[192] See e.g. O. de Paula Albuquerque et al., *Privacy in smart toys: Risks and proposed solutions*, in *Electronic Commerce Research and Applications*, vol. 39, 2020, DOI: 10.1016/j.elerap.2019.100922, whose authors refer to harassment, stalking, grooming, sexual abuse, exploitation, pedophilia and other types of violence blackmail, insults, confidence loss, trust loss and bullying; Shasha et al., *Playing With Danger: A Taxonomy and Evaluation of Threats to Smart Toys*, cit. See also Federal Bureau of Investigation, *Consumer Notice: Internet-Connected Toys Could Present Privacy and Contact Concerns for Children*, Alert Number I-071717 (Revised)-PSA (2017), <https://www.ic3.gov/Media/Y2017/PSA170717>, accessed on Dec. 15, 2020.

[193] See Shasha et al., *Playing With Danger: A Taxonomy and Evaluation of Threats to Smart Toys*, cit.

6.1.2.4. *Results of the initial assessment*

Table 11 shows the results of the assessment carried out on the initial idea of the connected AI-equipped doll described above.

Based on this table, we can plot a radial graph representing the overall impact on all the affected rights and freedoms (Figure 2). The graph shows the priority of mitigating potentially adverse impacts on privacy and data protection, followed by risks related to physical integrity and freedom of thought.

This outcome is confirmed by the history of the actual product, where the biggest concerns of parents and the main reasons for its withdrawal related to personal data and hacking[194].

Table 11 – Table of envisaged risks for the examined case (L: low, M: medium; H: high; VH: very high).

Risk	L	S	Overall impact
Impact on privacy and data protection	VH	M	H
Impact on freedom of thought	M	L	M
Impact on the right to psychological and physical safety	L	M	M

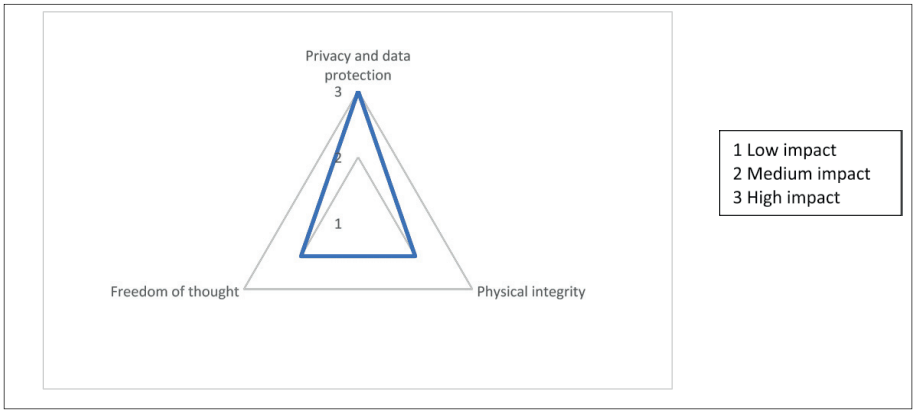


Fig. 2 – Radial graph (impact) of the examined case

[194] See S. Gibbs, *Hackers can hijack Wi-Fi Hello Barbie to spy on your children*, in *The Guardian*, Nov. 26, 2015, <https://www.theguardian.com/technology/2015/nov/26/hackers-can-hijack-wi-fi-hello-barbie-to-spy-on-your-children>, accessed on Nov. 12, 2020.

6.1.3 *Mitigation measures and re-assessment*

Following the iterative assessment, we can imagine that after this initial evaluation of the general idea, further measures are introduced to mitigate the potential risks found. At this stage, the potential stakeholders (users, parents associations, educational bodies, data protection authorities etc.) can make a valuable contribution to better defining the risks and how to tackle them.

While the role of the stakeholders cannot be directly assessed in this analysis, we can assume that their participation would have shown great concern for risks relating to communications privacy and security. This conclusion is supported by the available documentation on the reactions of parents and supervisory authorities[195].

After the first assessment and given the evidence on stakeholders' requests, the following mitigation measures and by-design solutions could have been adopted with respect to the initial prototype.

A) Data protection and the right to privacy

Firstly, the product must comply with the data protection regulation of the countries in which it is distributed[196]. Given the product's design, we cannot exclude the processing of personal data. The limited number of sentences provided for use by AI, as in the case of Hello Barbie, does not exclude the pro- vision of unexpected content by the user, including personal information[197].

Risk mitigation should therefore focus on the topics of conversation between the doll and the young user, and the safeguards in processing information collected from the user.

As regards the first aspect, an effective way to limit the potential risks would be to use a closed set of sentences, excluding phrases and questions

[195] See e.g. BEUC, *Connected Toys Do Not Meet Consumer Protection Standard*. Letter to Mr Giovanni Buttarelli, Euro- pean Data Protection Supervisor, Dec. 6, 2016, <https://www.beuc.eu/publications/beuc-x-2016-136_mgo_letter_to_giovanni_buttarelli_- _edps_- _connected_toys.pdf >, accessed on Nov. 12, 2020; ABA Journal, *Moms Sue Mattel, Saying "Hello Barbie" Doll Violates Privacy*, Dec. 9, 2015, <https://www.abajournal.com/news/article/hello_barbie_violates_ privacy_of _doll_owners_playmates_moms_say_in_lawsuit>, accessed on Mar. 20, 2021; E. McReynolds et al., *Toys That Listen: A Study of Parents, Children, and Internet-Connected Toys*, in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (ACM 2017)*, <https://dl.acm.org/doi/10.1145/3025453.3025735>, accessed on Nov. 12, 2020.

[196] In this regard Hello Barbie was certified as compliant with the US COPPA, see Mattel, *Hello Barbie FAQ*, cit.

[197] See Mattel, *Hello Barbie FAQ*, cit. ("we cannot control whether a child volunteers such information without prompting").

that might induce the user to disclose personal information, and making it possible to modify these phrases and questions by the owner of the toy[198].

Regarding the processing of personal data, the doll's AI-based information processing functions should be deactivated by default, giving the parents control over its activation[199]. In addition, to reduce the risk of constant monitoring, deliberate action by the child should be required to activate the doll's AI- equipped dialogue functions[200]. This would also help to make users more aware of their interaction with the system and related privacy issues[201].

Ex post remedies can also be adopted, such as speech detection to remove personal information in recorded data[202].

Conversations are not monitored, except to support requests from parents. To reduce the impact on the right to privacy and data protection, human review of conversations – to test, improve, or change the technology used – should be avoided, even if specific policies for unexpected findings have been adopted[203]. Individual testing phases or experiments can be carried out in a laboratory setting or on the basis of user requests (e.g. unexpected reactions and dialogues). This more restrictive approach helps

[198] In this case, the conditions are largely present, although there is evidence of minor issues. See e.g. *Id.* ("Hello Barbie does not ask for a child's name and is not scripted to respond with a child's name, so she will not be able to recite a child's name back to them"), but see the interaction reported in M. Leta Jones, *Your New Best Frenemy: Hello Barbie and Privacy Without Screens*, cit., p. 245 ("Barbie: Sometimes I get a little nervous when I tell people my middle name. But I'm really glad I told you! What's your middle name?! !"). The Mattel, *Hello Barbie FAQ*, cit. also points out the privacy-oriented design of the product with regard to dialogue content: "Although Hello Barbie was designed not to ask questions which are intended to elicit answers that might contain personal information".

[199] See Mattel, *Hello Barbie FAQ*, cit. ("Hello Barbie only requires a parent's email address to set up an account. This is necessary so that parents can give permission to activate the speech recognition technology in the doll. Other information, such as a daughter's birthday, can be provided to help personalize the experience but are not required […] If we discover that, in violation of our terms of service, an account was created by a child, we will terminate the account and delete all data and recordings associated with it.").

[200] In the Hello Barbie case, the doll was not always on but it was activated by pressing the belt buckle.

[201] In the examined case this was also emphasized because the two tri-color LEOs embedded in the doll's necklace lighted up to indicate she was active.

[202] See Mattel, *Hello Barbie FAQ*, cit. ("If we become aware of any such personal information captured in recordings, it is our policy to delete such information, and we contractually require our Service Providers to do the same. This personal information is not used for any purpose").

[203] See *supra* note 168.

to reduce the impact with respect to the initial design.

Further issues, regarding the information processing architecture and its compliance with data protection principles, concern data storage. This should be minimised, and giving parents the possibility to delete stored information[204].

With regard to the use of collected data, while access to, and sharing of, this information by parents[205] are not per se against the interest of the child, caution should be exercised in using this information for marketing purposes. Given the early age of the users and the potentially large amount of information they may provide in their conversation with the doll, plus the lack of active and continuous parental control, the best solution would be not to use child-doll conversations for marketing[206].

The complexity of data processing activities in the interaction between a child and an AI-equipped doll inevitably affects the form and content of the privacy policies and the options offered to users, as provided by many existing legislations.

A suitable notice and consent mechanism, clear and accessible and legally compliant, is therefore required[207], but meeting this obligation is not so simple in the case in question. The nature of the connected toy and the absence of any interface limits awareness of the policies and distances them from direct interaction with the device. This accentuates

[204] See Mattel, *Hello Barbie FAQ,* cit. ("Parents who are concerned about this can monitor their child's use of Hello Barbie, and parents have the power to review and delete any conversation their child has with Hello Barbie, whether the conversations contain personal information or not"). Considering the young age of the user this seems not to be a disproportionate monitoring with regard to their activities and right to privacy. This does not exclude a socio-ethical relevance of this behavior, see e.g. M. Leta Jones and K. Meurer, *Can (and Should) Hello Barbie Keep a Secret?*, in *IEEE International Symposium on Ethics in Engineering, Science and Technology (ETHICS)*, 2016, doi:10.1109/ETHICS.2016.7560047 ("the passive nature of Barbie's recording capabilities could prove per- haps more devastating to a child who may have placed an implicit trust in the doll. In order to determine the extent of the parent's involvement in their child's recordings, we extended our analysis to include the adult oversight capabilities").

[205] See above note 170.

[206] This was the option adopted in the Hello Barbie case, see *supra* note 171. But see Steeves, *A dialogic analysis of Hello Barbie's conversations with children*, cit. on the sentences used by Hello Barbie to indirectly reinforce the brand identity and encourage the child to adopt that identity for his/her own.

[207] In the case examined, one of the main weakness claimed with regard to Hello Barbie concerned the privacy policies adopted, the interplay between the different entities involved in data processing, and the design of these policies and access to them, which were considered cumbersome. See Leta Jones and Meurer, *Can (and Should) Hello Barbie Keep a Secret?*, cit.

the perception of the notice and consent mechanism as a mere formality to be completed to access the product.

The last crucial area concerns data security. This entails a negative impact that goes beyond personal data protection and, as such, is also analysed below under impact on the right to psychological and physical safety.

As the AI-based services are hosted by the service provider, data security issues concern both device-service communications and malicious attacks to the server and the device. Encrypted communications, secure communication solutions, and system security requirements for data hosted and processed on the server can minimise potential risks, as in the case study, which also considered access to data when the doll's user changes[208].

None of these measures prevent the risks of hacking to the device or the local Wi-Fi connection, which are higher when the doll is used outdoors[209]. This was the chief weakness noted in the case in question and in IoT devices more generally. They are often designed with poor inherent data security and cybersecure features for cost reasons. To reduce this risk, stronger authentication and encryption solutions have been proposed in the literature[210].

Taking into account the initial impact assessment plus all the measures described above, the exposure is reduced to low, since users are thus exposed to potential prejudices only in special circumstances, primarily malicious attack. Probability also becomes low, as the proposed measures mitigate the risks relating to dialogue between doll and user, data collection and retention. Likelihood (Table 4) is therefore reduced to low.

Regarding severity of prejudice, gravity can be lowered to at least medium by effect of the mitigation measures, but effort remains medium, given the potential risk of hacking. Severity is therefore lowered somewhat (from 5 to 3 in Table 7), though remaining medium.

If the severity and the likelihood are medium in Table 9, the overall impact is lowered from high to medium.

---

[208] See also Mattel, *Hello Barbie FAQ*, cit. ("Conversations and other information are not stored on the doll itself, but rather in the associated parent account. So, if other users are using a different WiFi network and using their own account, Hello Barbie would not re- member anything from the prior conversations. New users would need to set up their own account to enable conversations with Barbie").

[209] See Leta Jones, *Your New Best Frenemy: Hello Barbie and Privacy Without Screens*, cit., p. 244.

[210] See also below letter C).

B) Impact on freedom of thought

As described in Section 6.1.2.2, the impact on freedom of thought is related to the values conveyed by the doll in dialogue with the user. Here the main issue concerns the nature of the messages addressed to the user, their sources and their interplay with the rights and duties of parents to provide appropriate direction and guidance in the child's exercise of freedom of thought, including issues of cultural diversity.

A system based on Natural Language Processing allows AI various degrees of autonomy in identifying the best response or sentence in the human-machine interaction. Given the issues considered here (the nature of the values shared by the doll with its young user) the two main options are to use a closed set of possible sentences or search for potential answers in a large database, such as the Internet. A variety of solutions can also be found between these two extremes.

Since the main problem is content control, the preferable option is the first, and this was indeed the solution adopted in the Hello Barbie case[211]. Content can thus be fine-tuned to the education level of the user, given the age range of the children[212]. This reduces the risk of unexpected and inadequate content and, where full lines of dialogue are available (this was the case with Hello Barbie), parents are able to get an idea of the content offered to their children.

Some residual risks remain however, due to intentional or unintentional cultural models or values, including the difference between appropriate and inappropriate content[213]. This is due to the special relationship the toy generates[214] and the only limited mitigation provided by transparency on pre- recorded lines of dialogue.

To address these issues, concerning both freedom of thought and diversity, the AI system should embed a certain degree of flexibility (user-customizable content) and avoid stereotyping by default. To achieve this, the team working on pre-recorded sentences and dialogues should be characterised by diversity, adopting a by-design approach and bearing in mind the target user of the product[215].

Moreover, taking into account the parents' point of view, mere

[211] Mattel, *Hello Barbie FAQ*, cit.

[212] *Id.* ("All comments made by Hello Barbie are scripted with school age appropriate content").

[213] See *supra* note 178.

[214] See *supra* note 177.

[215] Steeves, *A dialogic analysis of Hello Barbie's conversations with children*, cit., on the different attitude in prerecorded sentences with regard to different religious topics.

transparency, i.e. access to the whole body of sentences used by the doll, is not enough. As is demonstrated extensively in the field of data protection, information on processing is often disregarded by the user and it is hard to imagine parents reading 8000 lines of dialogue before buying a doll.

To increase transparency and user awareness, therefore, forms of visualisation of these values through logic and content maps could be useful to easily represent the content used. In addition, it would be important to give parents the opportunity to partially shape the AI reactions, customising the values and content, providing other options relating to the most critical areas in terms of education and freedom of thought.

With regard to the effects of these measures, they mitigate both the potentially adverse consequences of product design and the lack of parental supervision of content, minimising the probability of an adverse impact on freedom of thought. The probability (Table 2) is therefore lowered to low.

Given the wide distribution of the product, the potential variety of cultural contexts and the need for an active role of parents to minimise the risk, the exposure remains medium, although the number of affected individuals is expected to decrease (Table 3).

If the probability is low and the exposure is medium, the likelihood (Table 4) is lowered to low after the adoption of the suggested mitigation measures and design solutions.

The gravity of prejudice and the effort were originally low and the additional measures described can further reduce gravity through a more responsible management of content which might support potentially conflicting cultural models or values. Severity therefore remains low.

Considering both likelihood and severity as low, the overall impact (Table 9) is reduced from medium to low, compared with the original design model.

C) Impact on the right to psychological and physical safety

The potential impact in this area is mainly related to malicious hacking activities[216] that might allow third parties to take control of the doll and use it to cause, psychological and physical harm to the user[217].

[216] See Gibbs, *Hackers can hijack Wi-Fi Hello Barbie to spy on your children*, cit.

[217] See V. Chang, Z. Li and M. Ramachandran, *A Review on Ethical Issues for Smart Connected Toys in the Context of Big Data*, in F. Firouzi, E. Estrada, V. Mendez Munoz, V. Chang, in *COMPLEXIS 2019 - Proceedings of the 4th International Conference on Complexity, Future Information Systems and Risk*, SciTePress, 2019, pp. 149-156 ("For example, the attackers can spread content through the audio system, which is adverse for

This was one of the most widely debated issues in the Hello Barbie case and one of the main reasons that led Mattel to stop producing this toy[218]. Possible mitigation measures are the exclusion of interaction with other IoT devices[219], strong authentication and data encryption[220].

As regards likelihood, considering the protection measures adopted and the low interest of third parties in this type of individual and context-specific malicious attack, the probability is low (Table 2). Although the suggested measures do not affect the exposure, this remains low due to the limited circumstances in which a malicious attack can be carried out (Table 3). The likelihood therefore remains low but is lowered (from 2 to 1 in Table 4).

Regarding severity, the proposed measures do not impact on the gravity of the prejudice (Table 5), or the effort (Table 6) which remain medium. Severity therefore remains medium (Table 7).

Since the final values of neither likelihood nor severity change, overall impact remains medium (Table 9), with malicious hacking being the most critical aspect of the product in terms of risk mitigation.

Table 12 – Comparative risk impact analysis table (examined case).

Risk	L	S	Overall impact	MMs	rL	rS	Final impact
Impact on privacy and data protection	VH	M	H	See above sub A)	M	M	M
Impact on freedom of thought	M	L	M	See above sub B)	L	L	L
Impact on the right to psychological and physical safety	L	M	M	See above sub C)	L	M	M
Overall impact (all impacted areas)	M/H		M/L				

Table 12 shows the assessment of the different impacts, comparing the results before and after the adoption of mitigation measures.

In the case in question, there is no Table 10 EF column since there are

children's growth through the built-in audio in the smart toys").

[218] Shasha et al., *Playing With Danger: A Taxonomy and Evaluation of Threats to Smart Toys*, cit.

[219] Doll's speech content was hand crafted by the writing team at Mattel and ToyTalk, not derived by open web search. Mattel, *Hello Barbie FAQ*, cit.

[220] See K. Demetzou, L. Böck and O. Hanteer, *Smart Bears don't talk to strangers: analysing privacy concerns and technical solutions in smart toys for children*, in *IET Conference Proceedings*; *Stevenage Stevenage: The Institution of Engineering & Technology* (2018) DOI:10.1049/cp.2018.0005; L. Gonçalves de Carvalho and M. Medeiros Eler, *Security Tests for Smart Toys*, in *Proceedings of the 20th International Conference on Enterprise Information Systems*, 2018, pp. 111-120, <http://www.scitepress. org/DigitalLibrary/Link. aspx?doi=10.5220/0006776101110120>, accessed on Dec. 23, 2020.

no factors that could exclude risk, such as certain mandatory impacting features or overriding competing interests recognised by law.

A radial graph (Figure 3) shows the concrete effect of the assessment (the blue line represents the initial impacts and the orange the impacts after adoption of the measures described above). It should be noted that the reduction of potential impact is limited as the Hello Barbie product already included several options and measures to mitigate adverse effects on rights and freedoms (pre-recorded sentences, no Internet access, data encryption, parental access to stored data, etc.). The effect would have been greater starting from a general AI-equipped doll using Natural Language Processing interacting with children, without mitigation measures.

In this regard, the HRIA model we propose is in line with a human rights-by design approach, where the design team is asked to consider human rights impact from the earliest product design stages, discarding those options that have an obvious negative impact on human rights. With this approach, there is no a HRIA where the proposed product is completely open to the riskiest scenarios (e.g. a connected doll equipped with unsupervised AI that uses all available web sources to dialogue with young users, with unencrypted doll-user communication sent to a central data centre where information is stored without a time limit and used for further purposes, including marketing communications direct to doll users).

In human rights-orientated design, HRIA thus becomes a tool to test, refine and improve adopted options that already entail a risk-aware approach. In this way, HRIA is a tool for testing and improving human rights-orientated design strategies.

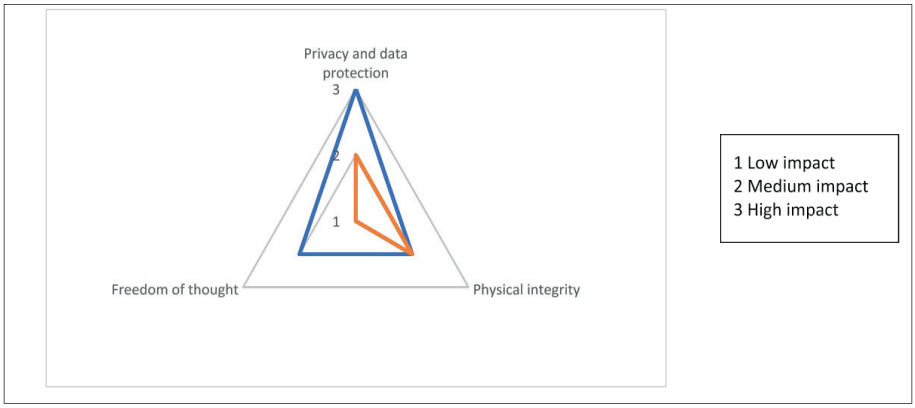


Fig. 3 – Final radial graph of the examined case. (Blue line: original impact. Orange line: final impact after adoption of mitigation measures and design solutions).

6.2. *HRIA in large-scale multi-factor scenarios: the sidewalk case*

Large-scale projects using data-intensive applications are characterised by a variety of potentially impacted areas concerning individual and groups. This produces a more complex and multifactor scenario which cannot be fully assessed by the mere aggregation of the results of HRIAs conducted for each component of data-intensive applications.

An example is provided by data-driven smart cities, where the overall effect of an integrated model including different layers affecting a variety of human activities means that the cumulative impact is greater than the sum of the impacts of each application.

In such cases, a HRIA for data-intensive systems also needs to consider the cumulative effect of data use and the AI strategies adopted, as already happens in HRIA practice with large-scale scenario cases. This is all the more important in the field of AI where large-scale projects often feature a unique or dominant technology partner who benefits from a general overview of all the different processing activities ('platformisation'[221]).

The Sidewalk project in Toronto is an example of this 'platformisation' effect and a case study in the consequent impacts on rights and freedoms. This concluded smart city project was widely debated[222] and raised several human rights-related issues common to other data-intensive projects. It also highlights how the universal nature of the benchmark framework proposed makes the assessment model suited to deployment in various jurisdictions, beyond European borders.

The case concerned a requalification project for the Quayside, a large urban area on Toronto's waterfront largely owned by Toronto Waterfront Revitalization Corporation. Based on an agreement between the City of Toronto and Toronto Waterfront[223], in 2017, through a competitive

[221] See E. Goodman & J. Powles, *Urbanism Under Google: Lessons from Sidewalk Toronto*, 88 *Fordham L. Rev.* 457–498, 2019.

[222] See C. Carr and M. Hesse, *When Alphabet Inc.Plans Toronto's Waterfront: New Post-Political Modes of Urban Governance*, 5 *Urban Planning* 69-83, 2020; A. Flynn and M. Valverde, *Where The Sidewalk Ends: The Governance Of Waterfront Toronto's Sidewalk Labs Deal*, 36 *Windsor Yearbook of Access to Justice* 263–283, 2019.

[223] The Waterfront Revitalization Corporation (which was renamed Waterfront Toronto) was a partnered not-for-profit corporation, created in 2003 by the City of Toronto, Province of Ontario and the Government of Canada (see also Province's Toronto Waterfront Revitalization Corporation Act) to oversee and deliver revitalization of Toronto's waterfront; further information are available here: <https://www.toronto.ca/city-government/accountability-operations-customer-service/city- administration/city-managers-office/agencies-corporations/corporations/waterfront-toronto/>, accessed

Request for Proposals, Waterfront Toronto hired Sidewalk Labs (a subsidiary of Alphabet Inc.) to develop a proposal for this area[224].

This proposal – the Master Innovation and Development Plan or MIDP[225] – outlined a vision for the Quayside site and suggested data-driven innovative solutions across the following areas: mobility and transportation; building forms and construction techniques; core infrastructure development and operations; social service delivery; environmental efficiency and carbon neutrality; climate mitigation strategies; optimisation of open space; data-driven decision making; governance and citizen participation; and regulatory and policy innovation[226].

This long list of topics shows how the data-intensive project went beyond mere urban requalification to embrace goals that are part of the traditional duties of a local administration, pursuing public interest purposes[227] with potential impacts on a variety of rights and freedoms.

---

on Dec. 30, 2020. See also Toronto Waterfront Revitalization, *Memorandum of Understanding between the City of Toronto, City of Toronto Economic Development Corporation and Toronto Waterfront Revitalization Corporation*, <https://www.toronto.ca/legdocs/2006/agendas/council/cc060131/pof1rpt/cl027.pdf >, accessed on Dec. 30, 2020; City of Toronto, Executive Committee, *Executive Committee consideration on January 24*, 2018.EX30. 9, 2018, <http://app.toronto.ca/tmmis/viewAgendaItemHistory.do?item=2018.EX30.9>, last visited Dec 30, 2020.

[224] Waterfront Toronto and Sidewalk Labs entered into a partnership Framework Agreement on October 16, 2017. The Framework Agreement was a confidential legal document, see City of Toronto, Executive Committee, cit. A summary of this agreement is available in City of Toronto, Executive Committee, *Executive Committee consideration on January 24, 2018, 2018.EX30. 9. Report and Attachments 1 and 2 from the Deputy City Manager, Cluster B on Sidewalk Toronto*, 2018, <https://www.toronto.ca/legdocs/mmis/2018/ex/bgrd/backgroundfile-110745.pdf>, accessed on Dec. 31, 2020, Comments, par. 2 and Attachment 2.

[225] Sidewalk Labs was charged with providing Waterfront Toronto with a MIDP for evaluation, including public and stakeholder consultation. Following the adoption of the MIDP by the Waterfront Toronto's Board of Directors, the City of Toronto was to complete an additional assessment programme focused on feasibility and legal compliance, including public consultation. See City of Toronto, Deputy City Manager, Infrastructure and Development, *Report for action. EX6.1*, 2019, <https://www.toronto.ca/legdocs/mmis/2019/ex/bgrd/backgroundfile-133867.pdf>, accessed on Dec. 30, 2020.

[226] See City of Toronto, Executive Committee, *Executive Committee consideration on January 24, 2018, 2018.EX30. 9*, cit.

[227] See also B. Wylie, *In Toronto, Google's Attempt to Privatize Government Fails—For Now*, in *Boston Rev.*, May 13, 2020; Goodman & Powles, *Urbanism Under Google: Lessons from Sidewalk Toronto*, cit.

Table 13 – Multi-factor scenario HRIA: main stages and tasks

Main stage	Sub-section	Main tasks
I. Planning and scoping	A. Preliminary analysis	1. Collection of information on the project, parties involved (including supply-chain), potential stakeholders, and territorial target areas (country, region)[1]. 2. Human rights reference framework: review of applicable binding and non-binding instruments, gap analysis.
	B. Scoping	1. Identification of main issues related to human rights to be examined. 2. Drafting of a questionnaire for HRIA interviews and main indicators.
II. Risk analysis and assessment	A. Fieldwork	1. Interviews with internal and external project stakeholders[2] and data collection[3]. 2. Understanding of contextual issues (political, economic, regulatory, and social).
	B. Analysis and assessment	1. Data verification and validation, comparing and combining fieldwork results and desk analysis. 2. Further interviews and analysis, if necessary 3. Impact analysis for each project branch and impacted rights and freedoms. 4. Integrated impact assessment report[4].
III. Mitigation and further Implementation	A. Mitigation	1. Recommendations. 2. Prioritisation of mitigation goals.
	B. Further implementation	1. Post-assessment monitoring. 2. Grievance mechanisms. 3. Ongoing stakeholder engagement.

The Sidewalk case[228] suggests several takeaways for the HRIA model. First, an integrated model, which combines the HRIAs of the different technologies and processes adopted within a multi-factor scenario (Table 13), is essential to properly address the overall impact, including a variety of socio-technical solutions and impacted areas.

Second, the criticism surrounding civic participation in the Sidewalk project reveals how the effective engagement of relevant stakeholders is central from the earliest stages of proposal design. Giving voice to potentially affected groups mitigates the risk of the development of top-down and merely technology driven solutions, which have a higher risk of

[228] For a more extensive discussion of this case, see also T. Scassa, *Designing Data Governance for Data Sharing: Lessons from Sidewalk Toronto*, in *Special Issue: Governing Data as a Resource, Technology and Regulation*, 2020, pp. 44-56; K. Morgan and B. Webb, *Googling the City: In Search of the Public Interest on Toronto's 'Smart' Waterfront*, 5 *Urban Planning* 84–95, 2020; A. Artyushina, *Is civic data governance the key to democratic smart cities? The role of the urban data trust in Sidewalk Toronto*, in *Telematics and Informatics*, vol. 55, 2020, DOI: 10.1016/j.tele.2020.101456; Flynn and Valverde (fn 222); K. Peel and E. Tretter, *Waterfront Toronto: Privacy or Piracy?*, 2019, <https://osf.io/xgz2s>, accessed on Dec. 28, 2020; C. Carr and M. Hesse, *Sidewalk Labs closed down – whither Google's smart city*, in *Regions*, 2020, <https://regions.regionalstudies.org/ezine/article/ sidewalk-labs-closed-down-whither-googles-smart-city/>, accessed on Dec. 28, 2020; Goodman & Powles, *Urbanism Under Google: Lessons from Sidewalk Toronto*, cit.

rejection and negative impact.

Third, the complexity and extent of large-scale integrated HRIA for multi-factor scenarios require a methodological approach that cannot be limited to an internal self-assessment but demand an independent third-party assessment by a multidisciplinary team of experts, as in traditional HRIA practice.

These elements suggest three key principles for large-scale HRIA: independence, transparency, and inclusivity. Independence requires third-party assessors with no legal or material relationship with the entities involved in the projects, including any potential stakeholders.

Transparency concerns both the assessment procedure, facilitating stakeholders' participation, and the public availability of the assessment outcome[229], using easily understandable language. In this sense, transparency is linked to inclusivity, which concerns the engagement of all the different stakeholders impacted by the activities examined.

An additional important contribution of the integrated HRIA is its ability to shed light on issues that do not emerge in assessing single components of large-scale AI systems, as the cumulative effect of such projects is key. Here, the human rights layer opens up to a broader perspective which includes the impact of socio-technical solutions on democratic participation and decisions.

The Urban Data Trust created by Sidewalk and its role in the Toronto project is an example in this sense. The Urban Data Trust was tasked with establishing "a set of RDU [Responsible Data Use] Guidelines that would apply to all entities seeking to collect or use urban data" and with implementing and managing "a four-step process for approving the responsible collection and use of urban data" and any entity that wishes to collect or use urban data in the district "would have to comply with UDT [Urban Data Trust] requirements, in addition to applicable Canadian privacy laws"[230].

This important oversight body was to be created by an agreement

[229] See also Mantelero, *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*, cit. p. 766, at footnote 94 ("It is possible to provide business-sensitive information in a separate annex to the impact assessment report, which is not publicly available, or publish a short version of the report without the sensitive content").

[230] 234 See Side Walk Labs, *Toronto Tomorrow. A new approach for inclusive growth*, in *MIDP*, Vol. 2, 2019, p. 419 and Vol. 3, p. 69. On the interplay the role of the Urban Data Trust in setting requirements for data processing and the legal framework into force in Canada and in Toronto, see Scassa, *Designing Data Governance for Data Sharing: Lessons from Sidewalk Toronto*, cit.

between Waterfront Toronto and Sidewalk Lab[231] and composed of a board of five members (a data governance, privacy, or intellectual property expert; a community representative; a public-sector representative; an academic representative; and a Canadian business industry representative) acting as a sort of internal review board and supported by a Chief Data Officer who, under the direction of the board, was to carry out crucial activities concerning data use[232]. In addition, the Urban Data Trust would have to enter into contracts with all entities authorised to collect or use urban data[233] in the district, and these data sharing agreements could also "potentially provide the entity with the right to enter onto property and remove sensors and other recording devices if breaches are identified"[234].

Although this model was later abandoned, due to the concerns raised by this solution[235], it shows the intention to create an additional layer of data governance, different from both the individual dimension of

[231] See also Scassa, *Designing Data Governance for Data Sharing: Lessons from Sidewalk Toronto*, cit., p. 55 ("in proposing the UDT, Sidewalk Labs chose a governance model developed unilaterally, and not as part of a collective process involving data stakeholders").

[232] See Side Walk Labs, *Toronto Tomorrow. A new approach for inclusive growth*, cit., vol. 2, p. 421 ("the Chief Data Officer would be responsible for developing the charter for the Urban Data Trust; promulgating RDU Guidelines that apply to all parties proposing to collect urban data, and that respect existing privacy laws and guidelines but also seek to apply additional guidelines for addressing the unique aspects of urban data […]; structuring oversight and review processes; determining how the entity would be staffed, operated, and funded; developing initial agreements that would govern the use and sharing of urban data; and coordinating with privacy regulators and other key stakeholders, as necessary").

[233] The notion of urban data is a novel category proposed by Side Walk, referring to "both personal information and information that is not connected to a particular individual […] it is collected in a physical space in the city and may be associated with practical challenges in obtaining meaningful consent […] Urban data would be broader than the definition of personal information and include personal, non-personal, aggregate, or de-identified data […] collected and used in physical or community spaces where meaningful consent prior to collection and use is hard, if not impossible, to obtain", see Side Walk Labs, *Toronto Tomorrow. A new approach for inclusive growth*, cit., vol. 2, p. 416. But see, for critical comments on this category and its use, Scassa, *Designing Data Governance for Data Sharing: Lessons from Sidewalk Toronto*, cit., pp. 51-54; Goodman & Powles, *Urbanism Under Google: Lessons from Sidewalk Toronto*, cit., p. 473.

[234] See Side Walk Labs, *Toronto Tomorrow. A new approach for inclusive growth*, cit., vol. 2, pp. 420-422.

[235] See Open Letter from Waterfront Toronto Board Chair, Oct. 31, 2019, <https://waterfrontoronto.ca/nbe/wcm/connect/ waterfront/waterfront_content_library/waterfront+home/news+ room/news+archive/news/2019/october/open+letter+from+waterfront+toronto+board+chair+-+october+31%2C+2019>, accessed on Mar. 8, 2021.

information self-determination and the collective dimension of public interest managed by public bodies, within a process of centralisation and privatisation of data governance regarding information generated within a community[236].

In this sense, the overall impact of AI applications in urban spaces and their coordination by a dominant player providing technological infrastructure raise important questions about the cumulative effect on potentially impacted rights, and even more concerning democracy and the socio-political dimension of the urban landscape[237], particularly in terms of the division of public and private responsibilities on matters of collective interest.

This privatisation of the democratic decision process, based on the 'platformisation' of the city, directly concerns the use of data, but is no longer just about data protection. In socio-technical contexts, data governance is about human rights in general, insofar as the use of data by different AI applications raises issues about a variety of potentially adverse effects on different rights and freedoms[238]. If data becomes a means of managing and governing society, its use necessarily has an impact on all the rights and freedoms of individuals and society. This impact is further exacerbated by the empowerment enabled by AI technologies (e.g. the use of facial recognition to replace traditional video-surveillance tools).

For these reasons, cumulative management of different data-intensive systems impacting on the social environment cannot be left to private service providers or an ad hoc associative structure, but should remain within the context of public law, centred on democratic participation in decision-making processes affecting general and public interest[239].

Large-scale data-intensive projects therefore suggest using the HRIA not only to assess the overall impact of all the various AI applications used,

[236] See also Artyushina, *Is civic data governance the key to democratic smart cities? The role of the urban data trust in Sidewalk Toronto*, cit.

[237] See also Carr and Hesse, *When Alphabet Inc.Plans Toronto's Waterfront: New Post-Political Modes of Urban Governance*, cit.

[238] Raso, et al., *Artificial Intelligence & Human Rights Opportunities & Risks*, cit.

[239] The right to participate in public affairs (Article 25 Covenant) is based on a broad concept of public affairs, which includes public debate and dialogue between citizens and their representatives, with close links to freedom of expression, assembly and association. See UN Human Rights Committee (HRC), *CCPR General Comment No. 25: The right to participate in public affairs, voting rights and the right of equal access to public service (Art. 25)*, CCPR/C/21/Rev.1/Add.7, 12 July 1996. See also UN Committee on Economic, Social and Cultural Rights (CESCR), *General Comment No. 1: Reporting by States Parties*, July 27, 1981, par. 5.

but also to go beyond the safeguarding of human rights and freedoms. The results of this assessment therefore become a starting point for a broader analysis and planning of democratic participation in the decision-making process on the use of AI, including democratic oversight on its application[240].

In line with the approach adopted by international human rights organisations, the human rights dimension should combine with the democratic dimension and the rule of law in guiding the development and deployment of AI projects from their earliest stages.

The findings of the HRIA will therefore also contribute to addressing the so-called 'Question Zero' about the desirability of using AI solutions in socio-technical systems. This concerns democratic participation and the freedom of individuals, which are even more important in the case of technological solutions in an urban context, where people often have no real opportunity to opt out due to the solutions being deeply embedded in the structure of the city and its essential services.

A key issue then for the democratic use of AI concerns architecture design and its impact on rights and freedoms. The active role of technology in co-shaping human experiences[241] necessarily leads us to focus on the values underpinning the technological infrastructure and how these values are trans- posed into society through technology[242]. The technology infrastructure cannot be viewed as neutral, but as the result of both the values, intentionally or unintentionally, embedded in the devices/services and the role of mediation played by the different technologies and their applications[243].

These considerations on the power of designers – which are widely discussed in the debate on technology design[244] – are accentuated in the

[240] See also A. Mantelero, *Analysis of international legally binding instruments*, in Council of Europe, *supra* note 10, pp. 82-88.

[241] See also N. Manders-Huits and J. van den Hoven, *The Need for a Value-Sensitive Design of Communication Infrastructures*, in Paul Sollie and Marcus Düwell (eds) *Evaluating New Technologies. Methodological Problems for the Ethical Assessment of Technology Developments*, 2009, pp. 55-56.

[242] See also D. Ihde, *Technology and the Lifeworld: from garden to earth*, 1990.

[243] See B. Latour and C. Venn, *Morality and Technology: The End of the Means*, in *Theory, Culture and Soc'y*, vol 19, 2022, pp. 47-60.

[244] See also L. Winner, *Do Artifacts Have Politics?*, 109 *Daedalus* 121–136 (1980); L. Winner, *Technē and Politeia: The Technical Constitution of Society*, in P. T. Durbin and F. Rapp (eds), *Philosophy and Technology*, 1983, p. 105 ("let us recognize that every technology of significance to us implies a set of political commitments that one can identify if one looks carefully enough. To state it more directly, what appear to be

context of smart cities and in many large-scale data-intensive systems. Here, the key role of service providers and the 'platformisation' of these environments[245] shed light on the part these providers play with respect to the overall impact of the AI systems they manage.

The HRIA can play an important role in assessing values and supporting a human rights-orientated design that also pays attention to participatory processes and democratic deliberation governing large-scale AI systems. This can facilitate the concrete development of a truly trustworthy AI, in which trust is based on respect for human rights, democracy and the rule of law.

7. *Conclusions*

The recent turn in the debate on AI regulation from ethics to law, the wide application of AI and the new challenges it poses in a variety of fields of human activities are urging legislators to find a paradigm of reference to assess the impacts of AI and to guide its development. This cannot only be done at a general level, on the basis of guiding principles and provisions, but the paradigm must be embedded into the development and deployment of each application.

With a view to providing a global approach in this field, human rights and fundamental freedoms can offer this reference paradigm for a truly human-centred AI. However, this growing interest in a human rights-focused approach needs to be turned into effective tools that can guide AI developers and key AI users, such as municipalities, governments, and

merely instrumental choices are better seen as choices about the form of the society we continually build, choices about the kinds of people we want to be"). See Verbeek, *Moralizing Technology. Understanding and Designing the Morality of Things*, pp.109, 129, and 164-165 ("Accompanying technological developments requires engagement with designers and users, identifying points of application fir moral reflection, and anticipating the social impact of technologies-in-design […] In order to develop responsible forms of use and design, we need to equip users ad designer with frameworks and methods to anticipate, assess, ad design the mediating role of technologies in people's lives and in the ways we organize society").

[245] See Council of Europe, Consultative Committee of the Convention 108 (T-PD), *Guidelines on Artificial Intelligence and Data Protection*, Strasbourg, Jan. 25, 2019, T-PD(2019)01; Council of Europe, Committee of Ministers, *Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine*, cit.

private companies.

To bridge this gap between theoretical thinking on the potential role of human rights in addressing and mitigating AI-related risks, this work has suggested an empirical evidence-based approach to developing a human rights impact assessment (HRIA) model for AI.

Using the results of an in-depth analysis of jurisprudence in the field of data processing in Europe, we have outlined how human rights and freedoms already play an important role in the assessment of data-intensive applications. However, there is the lack of a formal methodology to facilitate an ex-ante approach based on a human-orientated design of product/service development. Moreover, this empirical analysis has better clarified the interplay between human rights and data processing in data-intensive systems, facilitating the development of an evidence-based model that is easier to implement as it is based on existing case law rather than on an abstract theoretical evaluation of the potential impact of AI.

The core of our research is the proposed HRIA model for AI, which has been developed in line with the existing practices in HRIA, but in a way that better responds to the specific nature of AI applications, in terms of scale, impacted rights and freedoms, prior assessment of production design, and assessment of risk levels, as required by several proposals on AI regulation.

The result is a tool that can be easily used by entities involved in AI development from the outset in the design of new AI solutions and can follow the product/service throughout its lifecycle, providing specific, measurable and comparable evidence on potential impacts, their probability, extension, and severity, and facilitating comparison between alternative design options and an iterative approach to AI design, based on risk assessment and mitigation.

In this sense, the proposed model is no longer just an assessment tool but a human rights management tool, providing clear evidence for a human rights-orientated development of AI products and services and their risk management.

In addition, a more transparent and easy-to-understand impact assessment model facilitates a participatory approach to AI development by potential stakeholders, giving them clear and structured information about possible options and the effects of changes in AI design.

Finally, the proposed model can also be used by supervisory authorities and auditing bodies to monitor risk management in relation to the impact of data use on individual rights and freedoms.

Based on these results, several concluding remarks can be drawn. The

first general one is that conducting a HRIA should be seen not as a burden or a mere obligation, but as an opportunity. Given the nature of AI products/services and their features and scale, the proposed assessment model can significantly help companies and other entities to develop effective human-centric AI in challenging contexts.

The model can also contribute to a more formal and standardised assessment of AI solutions, facilitating comparison between different options and design approaches. Although HRIA has already been adopted in several contexts, large-scale projects are often assessed without using a formal evaluation of risk likelihood and severity[246]. Traditional HRIA reports often describe the risks found and their potential impact, but with no quantitative assessment, providing recommendations without grading the level of impact, leaving duty bearers to define a proper action plan.

This approach to HRIA is in line with voluntary and policy- based HRIA practice in the business sector. However, once HRIA becomes a legal tool – as suggested by the European Commission and the Council of Europe –, it is no longer merely a source of recommendations for better business policy. Future AI regulation will most likely bring specific legal obligations and sanctions for non-compliance in relation to risk assessment and management.

Analysis of potential impact will therefore become an element of regulatory compliance, with mandatory adoption of appropriate mitigation measures, and barriers in the event of high risk. A model that enables a graduation of risk can therefore facilitate compliance and reduce risks by preventing high- risk AI applications from being placed on the market.

With large-scale projects, such as smart cities, assessing each technological component using the proposed model and mitigating adverse effects is not sufficient. A more general overall analysis must be conducted in addition. Only an integrated assessment can consider the cumulative effect of a socio-technical system by measuring its broader impacts, including the consequences in terms of democratic participation and decision-making processes.

While the assessment of individual AI products/services might be

[246] See e.g. The Danish Institute for Human Rights, *Human Rights Impact Assessment – Durex and Enfa value chains in Thailand*, 2020, <https://www.humanrights.dk/publications/human-rights-impact-assessment-durex-enfa-value-chains-thailand>, accessed on Mar.2, 2021. But see K. Salcito and M. Wielga, *Kayelekera HRIA Moitorig Summary*, 2015, <http://nomogaia.org/wp-content/uploads/2015/10/KAYELEKERA-HRIA-MONITORING-SUMMARY-10-5-2015-Final.pdf >, accessed on Feb. 20, 2021.

carried out by the developing entity using the proposed model, large-scale multi-factor scenarios will require an additional layer: an integrated impact assessment conducted by external advisors. This integrated assessment, based on broader fieldwork, citizen engagement, and a co-design process, can evaluate the overall impact of an entire AI-based environment, in a way that is closer to traditional HRIA models.

In both cases, figures such as the human rights officer and tools like a HRIA management plan, containing action plans with timelines, responsibilities and indicators, can facilitate these processes[247], including the possibility of extending them to the supply chain and all potentially affected groups of people.

Finally, the proposed model with its more formalised assessment can facilitate the accountability and monitoring of AI products and services during their lifecycle[248], enabling changes in their impacts to be monitored through periodic reviews, audits, and progress reports on the implementation of the measures taken. It also makes it possible to incorporate more precise human rights indicators in internal reports and plans and make assessment results available to stakeholders clearly and understandably, facilitating their cooperation in a human rights-orientated approach to AI.

[247] See also D. Abrahams and Y. Wyss, *Guide to Human Rights Impact Assessment and Management (HRIAM)*, (The International Business Leaders Forum and the International Finance Corporation 2010) <https://d306pr3pise04h.cloudfront.net/docs/issues_doc%2Fhuman_rights%2FGuidetoHRIAM.pdf >, accessed on Oct. 26, 2020.

[248] See also The Danish Institute for Human Rights, *Guidance on HRIA of Digital Activities. Phase 4: Impact prevention, mitigation and remediation*, 2020, p. 25-33, <https://www.humanrights.dk/sites/humanrights.dk/files/media/document/Phase%204_%20Impact%20prevention%2C%20mitigation%20and%20remediation_n.pdf >, accessed on Feb. 20, 2021.

Nicoletta Rangone

Artificial Intelligence and Public Administrations: Addressing the Many Risks to Gain all the Benefits

Intelligenza Artificiale e pubbliche amministrazioni: affrontare i numerosi rischi per trarne tutti i vantaggi

Abstract: The paper aims at analysing the most relevant examples of the use of artificial intelligence within the Italian public administrations, in order to highlight related benefits and risks. It focuses on the internal organization, rulemaking and enforcement decisions. There appears to be a need for a general regulatory framework, as well as for specific procedural disciplines in each administration in order to allow real accessibility, comprehensibility and non-discrimination for the stakeholders. At the same time, these disciplines should introduce mechanisms and evaluation adequate to verify the appropriateness of the use of artificial intelligence instead of human intelligence, with a view to the effectiveness of organization and administrative action.

Abstract: Lo scritto parte dall'analisi delle più rilevanti esperienze di uso dell'intelligenza artificiale nelle pubbliche amministrazioni italiane, che attengono

* Questo articolo è stato originariamente pubblicato in *Biolaw Journal*, 2, 2022, pp. 473-488. Una prima versione del contributo è destinata alla pubblicazione in A. Lalli (a cura di), Le amministrazioni pubbliche nell'era digitale, Giappichelli in corso di pubblicazione. Le considerazioni svolte sono frutto della ricerca avviata nell'ambito del progetto PRIN 2017 "governance of/through Big Data: challenges for european law" e proseguita nel gruppo di ricerca ASTRID «Amministrazione e Intelligenza Artificiale» coordinato con B. Marchetti ed E. Chiti. L'autore è titolare della cattedra Jean Monnet on EU approach to Better regulation; il sostegno della Commissione europea alla produzione di questa pubblicazione non costituisce un'approvazione del contenuto, che riflette esclusivamente il punto di vista dell'autore, e la Commissione non può essere ritenuta responsabile per l'uso che può essere fatto delle informazioni ivi contenute. Contributo sottoposto a doppio referaggio anonimo.

all'assetto organizzativo, all'adozione di decisioni generali e all'attuazione amministrativa, per mettere in luce vantaggi e rischi. Ne emerge l'esigenza di un inquadramento normativo generale e di discipline procedurali delle singole amministrazioni, tali da assicurare reale accessibilità, comprensibilità e non discriminazione, oltre a meccanismi di controllo che consentano di verificare l'adeguatezza del ricorso all'intelligenza artificiale in luogo dell'intelligenza umana in un'ottica di effettività dell'organizzazione e dell'azione amministrativa.

1. *Introduzione*

Diverse e sempre più diffuse sono le applicazioni di intelligenza artificiale[1] nelle pubbliche amministrazioni italiane, a supporto di servizi al pubblico, di esigenze organizzative e conoscitive. Le nuove tecnologie sono inoltre utilizzate nei processi decisionali, o nella relativa fase prodromica, che si tratti dell'adozione di policies, di regole o di decisioni amministrative. Questi usi di nuove tecnologie sono all'origine di numerosi vantaggi per i pubblici poteri, che vanno dalla possibilità di comprimere i tempi per decidere e di razionalizzare l'impegno di risorse umane, alla elaborazione di indicazioni anche prospettiche, alla possibilità di evitare errori umani e di limitare le occasioni di corruzione[2]. L'intelligenza artificiale costituisce dunque uno strumento per l'effettività dell'organizzazione e dell'azione amministrativa, in ultima analisi di un

[1] In tal contesto, si può richiamare la definizione di intelligenza artificiale offerta dal Consiglio di Stato, per cui il machine learning «crea un sistema che non si limita solo ad applicare le regole software e i parametri preimpostati (come fa invece l'algoritmo "tradizionale") ma, al contrario, elabora costantemente nuovi criteri di inferenza tra dati e assume decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico» (punto 9.1, Consiglio di Stato, sez. III, 25 novembre 2021, n. 7891). Quanto alla definizione di intelligenza artificiale elaborata in seno alla proposta di Regolamento europeo sull'intelligenza artificiale (considerando 6 e art. 3), si rinvia a G. Avanzini, *Intelligenza artificiale, machine learning e istruttoria procedimentale: vantaggi, limiti ed esigenza di una specifica data governance*, in F. Donati, A. Pajno, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, Bologna, II, 2022, cap. 2, p. 47-48.

[2] Nella nota vicenda della mobilità docenti, il Consiglio di Stato (sez. VI, 4 febbraio 2020, n. 881) ha evidenziato che «l'utilizzo di una procedura informatica che conduca direttamente alla decisione finale non deve essere stigmatizzata, ma anzi, in linea di massima, incoraggiata: essa comporta infatti numerosi vantaggi quali, ad esempio, la notevole riduzione della tempistica procedimentale per operazioni meramente ripetitive e prive di discrezionalità, l'esclusione di interferenze dovute a negligenza (o peggio dolo) del funzionario (essere umano) e la conseguente maggior garanzia di imparzialità della decisione automatizzata».

diritto amministrativo che sia osservato, attuato e foriero di risultati coerenti con gli obiettivi per cui è stato previsto[3]. Un importante stimolo all'uso dell'intelligenza artificiale nell'organizzazione e azione amministrativa è svolto dagli incentivi europei di tipo economico[4] (da ultimo il Recovery and Resilience Facility)[5] e non economici che derivano dal confronto delle pratiche in sede internazionale[6] ed europea[7].

Gli indubbi vantaggi connessi all'uso delle nuove tecnologie non possono adombrare però i rischi, che vanno dall'opacità, all'errore, alla discriminazione[8]. Ed invero, se l'intelligenza artificiale nel processo decisionale consente di ridurre il «rumore» (*noise*), vale a dire la «variabilità» ingiustificata nei giudizi umani su situazioni identiche[9], può essa stessa essere il veicolo attraverso il quale sono perpetrate «le storture e le imperfezioni che caratterizzano tipicamente i processi cognitivi e le scelte compiute dagli esseri umani»[10] (*garbage in*, *garbage out*)[11]. Distorsioni possono anche essere legate alla scarsa qualità dei dati che alimentano l'intelligenza artificiale, ad esempio perché raccolti da soggetti in conflitto di interessi, basati su

[3] G. Corso, M. De Benedetto, N. Rangone, *Il diritto amministrativo effettivo. Una introduzione*, Bologna, 2022, p. 15. Il capitolo 3 di tale testo sviluppa una prima articolazione delle argomentazioni svolte in questo lavoro.

[4] Ad esempio, la *detection* algoritmica del linguaggio d'odio da parte dell'Autorità per le garanzie nelle comunicazioni ha ricevuto impulso anche da un progetto finanziato dalla Commissione europea (*Innovative Monitoring Systems and Prevention Policies of Online Hate Speech-IMSyP*) che ha portato l'autorità a lavorare nell'ambito di un consorzio di ricerca internazionale.

[5] Al rispetto e all'implementazione del principio, tra gli altri, del *digital by default* sono vincolate le riforme supportate dal *Recovery and Resilience Facility*.

[6] Ad esempio, il premio UNESCO al sistema INPS di smistamento delle PEC. Ed ancora, gli scrutini periodici di mutual evaluation realizzati a livello OCSE nel Gruppo intergovernativo d'Azione Finanziaria Internazionale-GAFI sono l'occasione per lo scambio di buone pratiche anche con riferimento all'uso dell'intelligenza artificiale nella vigilanza di Banca d'Italia. Questi esempi sono riportati da E. Chiti, B. Marchetti, N. Rangone, *L'impiego di sistemi di intelligenza artificiale nelle pubbliche amministrazioni italiane: prove generali*, in F. Donati, A. Pajno, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, cit., II, cap. 1.

[7] Ad esempio, in ambito ESMA.

[8] In tema, si veda F. Costantino, *Rischi o opportunità del ricorso delle amministrazioni alle predizioni dei big data*, in *Diritto pubblico*, 1, 2019, pp. 43 ss.

[9] «Whether the patent office grants or rejects a patent is significantly related to the happenstance of which examiner is assigned the application. This variability is obviously troublesome from the standpoint of equity» (D. Kahneman, O. Sibony, C.R. Sunstein, *Noise. A Flaw in human judgment*, New York, 2021, pp. 6-7).

[10] Consiglio di Stato, VI, 13 dicembre 2019 n. 8472.

[11] S.G. Mayson, *Bias In, Bias Out*, in *Yale L. J.*, vol. 128, 2019, p.2218 ss.

presunzioni non validate dal diretto interessato o esclusivamente su dati storici (che da un lato lasciano «sotto il radar» i nuovi entranti e nuovi rischi, dall'altro sono usati sulla base dell'assunto che il comportamento osservato non cambi in futuro[12]), raccolti per scopi diversi oppure tratti da «tracce» lasciate in rete dalle persone[13]. Ed ancora, un modello potrebbe risultare impreciso quando utilizzato nel mondo reale[14].

Vi sono poi problematiche specifiche, che attengono ai diversi momenti ed obiettivi per i quali i pubblici poteri fanno uso dell'intelligenza artificiale, su cui si concentra il presente contributo, che invece non affronta il tema delle competenze. Il paragrafo 2 è dedicato all'uso dell'intelligenza artificiale per il miglioramento delle prestazioni al pubblico e l'organizzazione interna, il paragrafo 3 alle nuove tecnologie nei procedimenti di regolazione e normativi, il paragrafo 4 all'attuazione amministrativa. Le conclusioni, di cui al paragrafo 5, sollevano l'esigenza di un inquadramento normativo per gli usi dell'intelligenza artificiale basato sulla trasparenza (in termini di effettiva conoscibilità), come regola minima e comune a tutte le applicazioni, ed esteso ai presidi della motivazione (come effettiva spiegabilità) e verificabilità dei sistemi per gli usi da parte delle pubbliche amministrazioni nei processi decisionali pubblici.

2. *Intelligenza artificiale per ottimizzare prestazioni e organizzazione interna delle amministrazioni pubbliche*

L'intelligenza artificiale viene utilizzata per ottimizzare le prestazioni al pubblico e l'organizzazione interna. Quanto ai servizi al pubblico, alcune applicazioni predisposte o in via di sperimentazione sono volte a facilitare il rapporto con i cittadini, rendendo più effettive le funzioni di prestazione.

Pur restando fondamentale la possibilità di interfacciarsi con una persona fisica[15], di grande utilità appare il miglioramento del supporto

[12] Definita «illusion of validity» su cui *infra*.

[13] T.B. Gillis, *The input fallacy*, in *Minnesota L. Rev*., vol. 106, 2022, p. 1175 ss.

[14] «First, a model can simply fail to fit any data — training or test — well. In such a scenario, even if the training and test data were perfectly representative of real-world data, the model would be inaccurate when deployed. Second, an algorithm can fit its training and, perhaps, test data well, but fail to generalize and perform equally well in real-world data» (D. Lehr, P. Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, in *University of California* vol.51, 2017, p. 711).

[15] C. Coglianese, *Administrative law in the automated State*, in *American Academy of Arts*

all'utenza che può essere offerto da chatbot "intelligenti" in grado di comprendere le richieste di cittadini diversi per preparazione e provenienza geografica (in uso ad esempio da parte di INPS e Agenzia delle entrate)[16]. Nella medesima ottica di facilitare il dialogo con l'amministrazione, muovono le applicazioni che attengono al miglioramento della funzionalità di motori di ricerca (anch'essi "intelligenti") di siti istituzionali, riducendo le barriere all'accesso e indirizzando l'utente verso pagine di effettiva utilità (come nella sperimentazione INPS). I sistemi che apprendono dalle ricerche fatte dai cittadini consentono, ad esempio, di selezionare le parole che consentono di raggiungere un determinato servizio all'interno del sito istituzionale (come può essere l'indicazione di "mensa scolastica" in luogo di "refezione"). Questi usi dell'intelligenza artificiale (come anche quelli che consentono la traduzione automatica da e verso varie lingue straniere o di filtrare messaggi di posta elettronica che costituiscono spam) potenziano i risultati ottenibili attraverso sistemi deduttivi alimentati attraverso un numero limitato di dati e presentano bassissimi livelli di rischio (in termini di gravità e probabilità) per la riservatezza dei dati raccolti dagli utenti e altri diritti fondamentali.

Una seconda tipologia di servizi al pubblico attiene alle numerose applicazioni che caratterizzano le smart cities[17], come la resa di informazioni raccolte attraverso l'internet delle cose e fornite agli utenti tramite app o siti istituzionali in ordine, ad esempio, alla disponibilità di parcheggi, alla densità di frequentazione di una determinata area di una città, a pericoli presenti sul manto stradale (in sperimentazione in alcuni comuni, come Padova, Trento e Venezia)[18]. Si tratta di funzionalità che potrebbero

and Sciences, 2021, 204 ss.; S. Ranchordás, *Empathy in the digital administrative law*, in *Duke L. J.* , vol. 71, 2022, p. 1341 ss.; J. Ponce Solé, *Inteligencia artificial, Detecho admiistrativo y reserve de humanidad: algoritmos y procedimiento administrative debito tecnologico*, in *Revista General de Derecho Administrativo*, 50, 2019.

[16] G. Buono, P. Bonanni, G. Del Mondo, A. Ciriello, *Intelligenza artificiale e amministrazioni centrali*, in *Biolaw Journal*, 1, 2022, p. 261 ss.

[17] E. Chiti, B. Marchetti, N. Rangone (a cura di), *L'uso dell'intelligenza artificiale nel sistema amministrativo italiano, rapporto 3/2022, SMART cities e intelligenza artificiale*, in *BioLaw Journal*, 1, 2022, p. 251 ss. Sull'uso dell'internet delle cose e l'intelligenza artificiale per offrire informazioni più puntuali, quando non personalizzate, agli utenti e adeguare l'organizzazione dell'offerta, si veda OCSE, *Shaping the future of regulators. The impact of emerging technologies on economic regulators*, 2020.

[18] Altre applicazioni possibili, in corso di realizzazione in altri ordinamenti giuridici, attengono alla robotica e consentono, ad esempio, la consegna di medicinali attraverso droni o la corrispondenza attraverso mezzi a guida autonoma (D.F. Engstrom, D.E. Ho, C.M. Sharkey, M.F. Cuellar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, Report submitted to the Administrative Conference of the

presentare rischi, soprattutto in termini di tutela privacy, là dove i dati siano raccolti attraverso telecamere e sensori acustici[19]. È dunque importante, in queste ipotesi, non solo assicurare la trasparenza degli usi, ma anche – ove possibile – evitare a monte questi rischi anonimizzando i dati proprio attraverso l'intelligenza artificiale e registrando solo il metadato (come nelle esperienze di Padova e Venezia), oppure creando un blind learning environment, che consente di evitare la condivisione di dati "in chiaro" all'interno di un medesimo soggetto pubblico (come nel caso di Banca d'Italia)[20].

Altri usi dell'intelligenza artificiale attengono all'organizzazione interna delle pubbliche amministrazioni e consentono di ottimizzare la distribuzione del lavoro tra uffici o la classificazione automatica della posta in entrata (come sperimentato dall'INPS)[21]. Si tratta di applicazioni che non sembrano presentare particolari rischi per i dipendenti o i soggetti terzi, se non in termini di una eventuale inefficiente organizzazione del lavoro o ritardati servizi all'utenza.

Diverso è il caso della razionalizzazione di attività che comportino la raccolta di informazioni che possono risultare prodromiche all'avvio eventuale di un procedimento di controllo, regolatorio o decisionale. Si pensi alla categorizzazione di esposti e denunce effettuata da Banca d'Italia[22] e alla vigilanza bancaria per la rilevazione di anomalie indicative

United States, pp. 65-69. Per un commento al report, si veda L. Parona, *"Government by algorithm" un contributo allo studio del ricorso all'intelligenza artificiale nell'esercizio di funzioni amministrative*, in *Giornale di diritto amministrativo*, 1, 2021, p. 10 ss.).

[19] Ed invero, la proposta di regolamento europeo sull'intelligenza artificiale elenca come ad alto rischio «i sistemi di IA destinati a essere utilizzati per l'identificazione biometrica remota "in tempo reale" e "a posteriori" delle persone fisiche» (allegato III).

[20] E. Chiti, B. Marchetti, N. Rangone, *L'impiego di sistemi di intelligenza artificiale nel sistema amministrativo italiano: prove generali*, cit.

[21] G. Buono, P. Bonanni, G. Del Mondo, A. Ciriello, *Intelligenza artificiale e amministrazioni centrali*, cit.

[22] Regolamento sul trattamento dei dati personali nella gestione degli esposti del 2022, pubblicato in gazzetta ufficiale e sul sito di Banca d'Italia. In base a tale regolamento, seppure la segnalazione non porti all'avvio un procedimento amministrativo, colui che presenta l'esposto e gli intermediari potenzialmente interessati devono sapere che le informazioni acquisite vengono trattate attraverso *machine learning*. Devono poi essere predisposte tecniche per garantire il rispetto della riservatezza dei dati personali, la sorveglianza umana del monitoraggio e dell'aggiornamento delle tecniche di *machine learning* (anche per assicurarne la spiegabilità), l'assenza di uso per decisioni automatizzate, profilazione o predizione di comportamenti. Questi gli aspetti particolarmente apprezzati dal Garante della *privacy,* nel parere del 24 febbraio 2022 alla Banca d'Italia sullo schema di regolamento concernente il trattamento dei dati personali effettuato nell'ambito della gestione degli esposti.

di possibili azioni di riciclaggio[23], alla classificazioni per grado di rischio di documenti oggetto di vigilanza (come nel caso documenti sintetici – KID – che illustrano le caratteristiche dei prodotti finanziari PRIIPs rivolti agli investitori al dettaglio oggetto di controllo da parte della Consob)[24], al monitoraggio dell'offerta abusiva on line di attività finanziarie riservate effettuato da Consob, o dell'uso del data mining e machine learning da parte di Agenzia delle entrate e INPS per rilevare i fenomeni fraudolenti nella lotta all'evasione contributiva. Nella valutazione di esposti e denunce (di cui si dirà nel paragrafo successivo) il rischio consiste in un possibile impoverimento del bagaglio informativo del decisore pubblico se non adeguatamente valutate/classificate, oltre a violazioni della privacy.

L'uso dell'intelligenza artificiale per la programmazione dei controlli informati al rischio (come nelle sperimentazioni supportate dall'OCSE e che coinvolgono la provincia autonoma di Trento e le regioni Lombardia e Campania)[25], che consente controlli più mirati anche in funzione prospettica, la eventuale limitata qualità dei dati utilizzati (come nel caso dell'uso di "data from the wild")[26], unitamente alla mancanza di trasparenza sul loro impiego, potrebbe non consentire ai soggetti interessati di contestare, ad esempio, l'erroneità del dato utilizzato. Inoltre, seppure un controllo non comporti necessariamente una conseguenza giuridica (in termini ad esempio di sanzione), esso rappresenta di per sé stesso un costo quando effettuato in loco; peraltro, la rilevazione di una anomalia anche non grave costituirà un precedente preso in considerazione nella mappatura del rischio che caratterizza un determinato individuo o impresa, comportando l'esposizione a nuovi controlli futuri. È dunque importante che i dati siano verificabili e trasparenti e che gli esiti dei controlli "algoritmici" siano costantemente

[23] E. Chiti, B. Marchetti, N. Rangone (a cura di), *L'uso dell'intelligenza artificiale nel sistema amministrativo italiano, rapporto 2/2021, L'impiego dell'intelligenza artificiale nell'attività di Banca d'Italia*, in *BioLaw Journal*, 4, 2021, p. 229 ss.

[24] E. Chiti, B. Marchetti, N. Rangone (a cura di), *L'uso dell'intelligenza artificiale nel sistema amministrativo italiano, rapporto 1/21, L'impiego dell'intelligenza artificiale nell'attività di CONSOB, Agcom e ARERA*, in *Biolaw Journal*, 4, 2021, p. 215.

[25] F. Blanc, M. Benedetti, C. Bertone, *L'informatica e il machine learning al servizio della semplificazione dei controlli sulle imprese: un equilibrio ancora da definire*, in *Studi parlamentari e di politica costituzionale*, 209, 2021, p. 121 ss.

[26] Si tratta dell'uso di dati raccolti a fini diversi o prodotti in esito ad altre attività che vanno ad alimentare sistemi di intelligenza artificiale (che da questi apprendono) senza che vi sia stato un riscontro esplicito e consapevole da parte del proprietario del dato o un monitoraggio da parte di terzi sulla qualità dello stesso (N. Cristianini, *Shortcuts to Artificial Intelligence*, in M. Pelillo, T. Scantamburlo (a cura di), *Machines We Trust: Perspectives on Dependable AI*, Cambridge, Massachusetts, London, England, 2021, pp. 15-17).

monitorati in modo da "allenare" ed ricalibrare l'algoritmo ove necessario. Il ricorso a controlli "algoritmici" dovrebbe essere reso noto al pubblico (trasparente) nelle forme che verranno evidenziate nel paragrafo 5.

L'intelligenza artificiale può, poi, supportare l'organizzazione di controlli in un'ottica di ordine pubblico, come nel caso di applicazioni utilizzate da alcune questure italiane che processano dati raccolti da telecamere dislocate sul territorio e dati in possesso della Polizia, del Ministero dell'interno o altri soggetti pubblici (in alcuni contesto anche social media) per aumentare l'effettività dei controlli, attraverso una loro programmazione in base al rischio. Oltre ai pericoli menzionati, le applicazioni di polizia predittiva, seppure nel nostro ordinamento giuridico limitate a tracciare «mappe» del rischio[27] e non «persone a rischio»[28], potrebbero riprodurre pregiudizi tipici della valutazione umana rispetto a certe zone e dunque persone che vi abitano[29] e, là dove diano un peso prevalente ai dati storici, possono portare a concentrare i controlli su certi tipi di crimini o aree perdendone di vista altre o nuovi rischi, se non a stigmatizzare determinate persone. Questi sistemi, pur non essendo considerati a livello europeo (forse paradossalmente) tra quelli ad alto rischio, devono comunque essere informati al principio della trasparenza in base alla proposta di regolamento sull'intelligenza artificiale. A ciò si aggiunga il monitoraggio dei risultati e la verifica del buon funzionamento dell'algoritmo.

3. *Intelligenza artificiale per il* rule-making

Nei procedimenti di regolazione e normativi, le nuove tecnologie sono utili per riorganizzare e analizzare i commenti raccolti nell'ambito di consultazioni con un elevato numero di partecipanti[30], per supportare la rac-

[27] Sulle esperienze di alcune questure italiane che utilizzano sistemi di polizia predittiva KeyCrime, X-Law, E-Security si rinvia a M.B. Armiento, *La polizia predittiva come strumento di attuazione amministrativa delle regole*, in *Diritto amministrativo*, 4, 2020, pp. 990-991 e, dello stesso autore, *Nuove tecnologie e "nuova" sicurezza delle città*, in *Studi parlamentari e di politica costituzionale*, 209, 2021, p. 95 ss.

[28] I sistemi che interessano le persone nell'ambito delle attività di contrasto di reati sono considerai ad alto rischio dalla proposta di regolamento europeo, allegato III.

[29] P.J. Brantingham, M. Valasik, G.O. Mohler, *Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial*, in *Statistics and Public Policy*, 5, 1, 2017, p. 1 ss.

[30] Ad esempio, nell'ambito delle consultazioni ralizzate per l'analisi preventive di impatto della proposta di direttiva europea su Energy performance of buildings (recast)

colta e l'elaborazione di dati di varia provenienza (come denunce e reclami) che possono far emergere l'esigenza di un intervento di regolazione[31], per svolgere attività automatizzata o semiautomatizzata di *drafting*.

Nel nostro ordinamento non sembrano esservi esperienze di uso del machine learning per la lettura dei risultati di consultazioni, fatta eccezione per due politiche del 2014, «La buona scuola»[32] e «Rivoluzione@ governo»[33]. I commenti ricevuti nella seconda consultazione sono stati il risultato di diverse campagne di mobilitazione promosse da soggetti organizzati per l'invio di commenti identici o sostanzialmente tali, di seguito definiti commenti di massa[34]. Nessuna applicazione in questo ambito si rileva invece, al momento, nelle autorità indipendenti italiane, che pure vantano le esperienze più avanzate in termini di consultazioni, presumibilmente per il numero contenuto di commenti normalmente ricevuto.

Quali considerazioni si possono svolgere sull'uso dell'intelligenza artificiale nelle consultazioni? A fronte di un prezioso supporto offerto ai decisori, l'uso in questa fase delle tecnologie più avanzate presenta rischi specifici, di cui occorre avere consapevolezza affinché possano essere adeguatamente affrontati dagli sviluppatori e dai decisori pubblici. Nella trattazione dei commenti di massa, l'intelligenza artificiale consente di identificare[35] e riorganizzare le osservazioni ricevute, generando un considerevole risparmio di tempo all'amministrazione che conduce la consultazione[36]. Questo sistema di razionalizzazione dell'analisi rischierebbe però di alterare il bagaglio informativo su cui si basa il processo decisionale

«the results of the feedback were analysed using Atlas.ti (text processing software)» (SWD(2021) 453 final, 142).

[31] M.A. Livermore, V. Eidelman, B. Grom, *Computational assisted regulatory participation*, in *Notre Dame L. Rev.*, 2018, vol. 93, 3, p. 977 ss.

[32] Nella consultazione, svoltasi dal 15 settembre al 15 novembre 2014, le 6.470.000 risposte strutturate ricevute sono state analizzate, per l'estrazione di concetti chiave, attraverso la linguistica computazionale con il supporto della Fondazione Bruno Kessler.

[33] Le 39.343 mail pervenute nell'ambito della consultazione che si è svolta dal 30 aprile – 30 maggio 2014 sono state analizzate attraverso strumenti di text mining per classificare i messaggi ricevuti secondo il grado di pertinenza con i 44 punti della riforma, con il supporto del Dipartimento di metodi e modelli per l'economia, il territorio e la finanza dell'Università Sapienza.

[34] «As a rule of thumb, the minimum threshold should be 10 or more identical responses (across all the closed questions) to count as a campaign» (European Commission, *Better Regulation Toolbox*, 2021, p. 473).

[35] Sui software dedicati si veda European Commission, *Better Regulation Toolbox*, 2021, p. 473.

[36] Administrative Conference of the United States, *Recommendation 2021-1, Managing Mass, Computer-Generated, and Falsely Attributed Comments*, 17 giugno 2021.

se supportasse l'automatica esclusione dei commenti di massa in quanto tali o la loro considerazione come un unico commento[37]. Lo stesso è da dirsi per l'uso delle nuove tecnologie per semplificare l'analisi di un'importante mole di dati frutto di segnalazioni, denunce o reclami (sperimentata da Consob e Banca d'Italia)[38], là dove ricevessero un'attenzione minore perché, ad esempio, contenenti errori ortografici o un gergo poco appropriato[39].

Nel rule-making, le nuove tecnologie consentono inoltre la traduzione in codici di una regola esistente, la scrittura in tale forma di nuove regole o la riforma di queste. Le norme potrebbero poi essere scritte ab origine in doppio formato, tradizionale e in codice, in attuazione del più generale principio del *digital by default*[40]. Ciò consente di facilitare l'attività di *drafting*, supportando, ad esempio, la rilevazione delle incongruenze o incompatibilità tra norme[41]. Al contempo, il "rules as code" agevola l'implementazione, consentendo l'automazione dell'attuazione

[37] Al riguardo la Comissione europea e la Corte dei Conti suggeriscono di effettuare un'analisi separata dei commenti di massa e di evidenziarne a parte i risultati (*Better Regulation Toolbox* 2021, 75, pp. 472-474, Court of Auditors, "*Have your say!": Commission public consultations engage citizens, but fall short of outreach activities*, special report 14/2019, pp. 39-40). Sul punto si vedano S.J. Balla, A.R. Beck, E. Meehan, A. Prasad, *Lost in the flood?: Agency responsiveness to mass comment campaigns in administrative rulemaking*, in *Regulation & Governance*, 2020 e la series of essays, *Mass Comments in Administrative Rulemaking*, in *The Regulatory Review*, pp. 13-21 dicembre 2021 (<https://www.theregreview.org/2021/12/13/mass-comments/ ultima consultazione>, 19/06/2022); S. Katzen, *Public input in rulemaking*, in *The Regulatory Review*, 7 marzo 2022 (<https://www.theregreview.org/2022/03/07/katzen-public-input-in-rulemaking/>, ultima consultazione Giugno 19, 2022).

[38] L'impiego dell'IA nell'attività di CONSOB, AGCOM e ARERA, cit., 215 ss.

[39] «Addressing the lack of participation by marginalized communities in regulatory decision-making is crucial, but there is another fundamental issue. The input of marginalized communities will not matter if agencies ignore or devalue it because these insights are not expressed using the standard narratives of policymaking» (S.A. Shapiro, *Marginalized Groups and the Multiple Languages of Regulatory Decision-Making*, in *The Regulatory Review*, 14 marzo 2022, <https://www.theregreview.org/2022/03/14/shapiro-marginalized-groups-multiple-languages/>, ultima consultazione Giugno 19, 2022).

[40] Comunicazione della Commissione europea, *Bussola per il digitale 2030: il modello europeo per il decennio digitale, COM (2021)118 final* e *Better Regulation Toolbox* 2021, TOOL#18, p. 145 e TOOL#28, p. 228 ss.

[41] «The idea with "rules as code" is that the government would make its single coded version available via an API [Application Programming Interface] to the public, including developers, not just those in government» (M. Waddington, *Machine-consumable legislation: A legislative drafter's perspective – human v artificial intelligence*, in *The Loophole*, giugno 2019, p. 27).

amministrativa che non implichi discrezionalità[42], come sembrerebbe possibile anche nel nostro ordinamento giuridico[43].

Non mancano però potenziali risvolti negativi. In primo luogo, la traduzione o la scrittura anche in codici comporta una estrema semplificazione (aspetto positivo)[44], ma che può tradursi in un impoverimento del dettato normativo, se non in una distorsione in ipotesi estreme[45]. In secondo luogo, il "rule as a code", quando porta all'automazione della decisione amministrativa, implica una tendenziale sovrapposizione tra *rule-making* e *adjudication* con perdita delle garanzie procedurali ad entrambi i livelli[46].

È dunque importante assicurare la trasparenza, rendendo esplicita la formulazione in codici delle regole e consentendo la partecipazione degli interessati a questo livello, così come dovrebbe essere reso pubblico il fatto che alcune previsioni non discrezionali ricevano applicazione automatizzata[47]. L'intelligenza artificiale può essere utilizzata anche per adeguare le regole al mutamento del contesto o del quadro normativo di riferimento. La revisione dello stock può non rivelarsi una operazione neutra, se espressione di politiche de-regolatorie o al contrario favorevoli a

[42] «Objective criteria in the legislation are a prerequisite for automation of case processing. […]. For example, the criterion "majority of the year" may be subject to discretion and interpretation whereas "more than 250 calendar days per year" can be assessed objectively. If fully automated case processing is introduced based on objective criteria» (così il documento dell'agenzia danese per la digitalizzazione del 2018 *Guidance on digital-ready legislation on incorporating digitisation and implementation in the preparation of legislation*, p. 11, <https://en.digst.dk/media/20206/en_guidance-regarding-digital-ready-legislation-2018.pdf>, ultima consultazione 19/06/2022).

[43] Consiglio di Stato, sez. cons. att. norm., 17 novembre 2020 n. 1322.

[44] Ad esempio, il citato document danese suggerisce che «the rules should be worded clearly and simply, unambiguously and consistently. Simple rules do not necessarily mean a brief law text. It may require more words to make it unambiguous and clear what the rules are. This does not, however, change the overall legal principle that superfluous words in the law text should be avoided» (*Guidance on digital-ready legislation*, cit., p. 8 e 9). Dello stesso tenore il *Better Regulation Toolbox* 2021 della Commissione europea, TOOL#28 Digital-ready policymaking, in particolare p. 238.

[45] «Policy is often distorted when programmers translate it into code. […] This is, in part, because the artificial languages intelligible to computers have a more limited vocabulary than human languages. Computer languages may be unable to capture the nuances of a particular policy. Code writers also interpret policy when they translate it from human language to computer code» (D.K. Citron, *Technological Due Process*, in *Washington L. Rev.*, vol. 6, 85, 2008, p. 1261).

[46] Citron, *Technological Due Process*, cit., p. 1249.

[47] «The decision is sufficiently transparent to enable the citizen to assess his/her avenues of complaint and it must be possible to verify the decision» (*Guidance on digital-ready legislation*, cit., p. 11).

una regolazione pervasiva, esiti particolarmente preoccupanti se influenzati da portatori di interessi di parte ben organizzati[48].

In generale, trasparenza e partecipazione dovrebbero dunque essere sempre assicurati, unitamente ad un costante monitoraggio e una valutazione ex post dei risultati consegui, sia per la definizione di nuove regole, che per la revisione di quelle esistenti con l'ausilio dell'intelligenza artificiale.

Intelligenza artificiale per l'attuazione amministrativa

Le numerose applicazioni di algoritmi lineari ai procedimenti per l'adozione di decisioni individuali (che vanno dalla mobilità dei docenti, al riconoscimento di incentivi ad imprese, ai contratti pubblici, alla determinazione di tariffe di servizi pubblici) hanno portato il giudice amministrativo italiano[49] a delineare un rafforzamento delle tutele tradizionali: trasparenza come reale conoscibilità[50] e spiegabilità, non esclusività della decisione algoritmica[51]. Si tratta di garanzie effettive?

Trasparenza effettiva e spiegabilità: obiettivi raggiungibili?

Le applicazioni di intelligenza artificiale utilizzate a livello di amministrazioni centrali e locali non sono facilmente individuabili, nascoste in notizie di stampa, generiche comunicazioni istituzionali[52],

[48] C.M. Sharkey, *AI for retrospective review*, in *Belmont L. Rev*., 8, 3, 2021, p. 374 ss.

[49] Consiglio di Stato, sez. VI, sentenze 13 dicembre 2019 n. 8472, 8473 e 8474, nonché Consiglio di Stato, sez. VI, sentenza 4 febbraio 2020, n. 881, cit. Tali pronunce fanno riferimento all'art. 22 GDPR, che sancisce il diritto per gli interessati a non essere sottoposti a una decisione basata unicamente sul trattamento automatizzato.

[50] «Tale conoscibilità dell'algoritmo deve essere garantita in tutti gli aspetti: dai suoi autori al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti. Ciò al fine di poter verificare che i criteri, i presupposti e gli esiti del procedimento robotizzato siano conformi alle prescrizioni e alle finalità stabilite dalla legge o dalla stessa amministrazione a monte di tale procedimento e affinché siano chiare – e conseguentemente sindacabili – le modalità e le regole in base alle quali esso è stato impostato» (Consiglio di Stato n. 8472/2019, cit.).

[51] Salvo quanto alle decisioni seriali e standardizzate (A. Masucci, *L'algoritmizzazione delle decisioni amministrative tra Regolamento europeo e leggi degli Stati membri*, in *Dir. pubblico*, 2, 2020, pp. 945-946). Sui principi elaborati dal giudice amministrativo, si rinvia a E. Carloni, *I principi della legalità algoritmica. Le decisioni automatizzate di fronte al giudice amministrativo*, in *Diritto amministrativo*, 2, 2020, p. 273 ss.

[52] Ad esempio, la direzione studi e ricerche INPS, in base al sito istituzionale, «fornisce supporto tecnico-scientifico all'elaborazione delle decisioni che l'Istituto assume nell'ambito delle proprie attività istituzionali attraverso […] l'elaborazione di statistiche, di modelli di data mining e machine learning, anche in riferimento ai big data» (<https://www.inps.it/nuovoportaleinps/default.aspx?itemdir=53263 ultimaconsultazione>, 19/06/2022).

circolari[53] o altri documenti[54] (con l'unica eccezione, al momento in cui si scrive, del regolamento di Banca d'Italia citato in nota 22). Che si tratti del controllo del flusso di traffico, dell'accesso ai parcheggi, di polizia predittiva, di verifiche del corretto pagamento di tributi o contributi, dell'individuazione di abusi di mercato, manca una chiara indicazione dei procedimenti interessati, delle tecnologie utilizzate, dei criteri in base ai quali ci si è rivolti al mercato invece di produrre internamente l'algoritmo, delle ragioni dell'utilizzo, dell'impostazione o meno di un monitoraggio del funzionamento e dei relativi esiti[55].

Questa limitata trasparenza determina ovviamente una limitata effettività delle garanzie di partecipazione: perché l'interessato dovrebbe chiedere l'accesso al codice sorgente se non sa che viene utilizzato un algoritmo? Il problema del limite alle garanzie di partecipazione che consegue alla mancanza di trasparenza si pone ancor prima di quello legato alla reale possibilità, per soggetti privi di particolari competenze tecniche, di comprendere il funzionamento di un codice sorgente. Questa stessa mancanza di trasparenza potrebbe impattare negativamente anche sull'esercizio delle garanzie processuali, dal momento che soggetti potenzialmente interessati ma non al corrente del ricorso all'intelligenza artificiale in un processo decisionale, non possono contestarne né l'uso,

[53] Ad esempio, INPS, circolare n. 23/2010, Funzione di accertamento e verifica amministrativa - Attuazione del nuovo modello organizzativo delle strutture territoriali di produzione previsto dalla circolare n. 102 del 12/08/2009 e circolare n. 23/2010, Funzione di accertamento e verifica amministrativa - Attuazione del nuovo modello organizzativo delle strutture territoriali di produzione previsto dalla circolare n. 102 del 12/08/2009.

[54] Come nel caso del cd. risparmiometro, menzionato nel Piano della performance dell'Agenzia delle entrate 2018-20, ma poi descritto solo in una decisione del Garante della Privacy del 20 luglio 2017, n. 321, Sperimentazione di una procedura basata sull'utilizzo di informazioni fornite dall'Archivio dei rapporti finanziari e degli elementi presenti nell´Anagrafe tributaria per l´individuazione di profili di evasione rilevanti.

[55] Il problema della limitata trasparenza resta attuale anche per numerose agenzie Nordamericane (D. Freeman Engstrom, D.F. Ho, *Algorithmic Accountability in the Administrative State*, 37 *Yale J. Regulation* 800, 2020) e non sembra sia stato dato seguito all'«Agency Inventory of AI use Cases» richiesto dall'Executive Order n. 13960/2000 (J.F. Weaver, *Everything Is Not Terminator. The Federal Government and Trustworthy AI*, in *Robotics, Artificial Intelligence & L.*, vol. 4, p. 291 (2021)). «In the absence of timely action of the executive or legislative branches to establish procedures to mitigate for administrative agency AI accountability gaps (like bias) and transparency gaps (like being understandable and traceable), the judiciary may dictate such procedures via remands under the administrative record provision of the Administrative Procedure Act of 194623 as it first did fifty years ago with informal adjudications» (A.A. Gavoor, *The impeding judicial regulation of artificial intelligence in administratieve State*, 97 *Notre Dame L. Rev*. 183, 2022).

né gli esiti. La trasparenza diventa dunque cruciale per l'effettività delle garanzie procedurali e processuali. Si tratta di una trasparenza che, per assicurarne l'effettività, andrebbe realizzata su due livelli: l'informazione sulle tecnologie utilizzate e le relative funzionalità andrebbe messa a disposizione del pubblico in forma semplificata e facilmente comprensibile, con possibilità di approfondimento per gli interessati[56].

Trasparenza significa anche spiegabilità, declinazione particolarmente delicata sia quando l'algoritmo viene sviluppato da unità specializzate all'interno di una pubblica amministrazione, che in caso di acquisto sul mercato. Se nella prima ipotesi un adeguato disegno organizzativo interno dovrebbe assicurare il costante e reciproco coinvolgimento di sviluppatori e funzionari che faranno uso del sistema (non solo per assicurare una progettazione funzionale alle esigenze dei secondi, ma anche per consentire a questi di comprendere e saper spiegare il funzionamento), nel secondo è cruciale assicurare trasparenza e verificabilità dell'operatività del sistema di intelligenza artificiale attraverso il bando di gara[57]. Questa è la sede sia per prevedere che il funzionamento dell'algoritmo sia oggetto di audit e "spiegabile" al soggetto pubblico (che a sua volta dovrà poterlo spiegare agli interessati), che per imporre un sistema di gestione del rischio[58].

[56] Ad esempio, il registro degli usi di intelligenza artificiale messo a disposizione del pubblico della città di Helsinki offre informazioni approfondite, pur essendo al contempi perfettamente comprensibili anche da non esperti, su: *dataset* (vale a dire le fonti di dati utilizzate nello sviluppo e nell'uso del sistema, il loro contenuto e i metodi di utilizzo), *data processing* (la logica operativa dell'elaborazione automatica dei dati e del ragionamento eseguito dal sistema e i modelli utilizzati), non discriminazione (connesso, per esempio, al numero di lingue disponibili), supervisione umana durante l'uso del servizio, *risk management* (vale a dire ai rischi legati al sistema e al suo uso e i loro metodi di gestione).

[57] «Establishing contractual pre-conditions for acquiring algorithmic systems ensures that systems that do not comply with specific conditions of transparency or fairness are not acquired or used by governments, or that, if a vendor fails to meet contractual conditions, they are subject to contractual liability. Procurement conditions also allow for interventions in the design of algorithmic systems, as well as during their use» (Ada Lovelace Institute, AI Now Institute and Open Government Partnership, *Algorithmic Accountability for the Public Sector*, 2021, pp. 33-35 e 44-45, <https://www.opengovpartnership.org/documents/%20 algorithmic-accountability-public-sector/>, ultima consultazione 19/06/2022).

[58] Linee guida per la redazione dei bandi sono state definite dal governo inglese (*Guidelines for AI procurement 2020*, in <https://www.gov.uk/government/publications/guidelines-for-ai-procurement>, ultima consultazione 19/06/2022) e dalla città di Amsterdam (*Standard Clauses for Procurement of Trustworthy Algorithmic Systems*, 2020 (<https://www.amsterdam.nl/innovatie/digitalisering-technologie/contractual-terms-for-algorithms/>, ultima consultazione 19/06/2022). Il Canada ha invece optato per una lista di fornitori pre-selezionati anche in base al rispetto di "demonstrated competence in AI ethics" (<https://buyandsell.gc.ca/procurement-data/tendernotice/PW-EE-017-34526>, ultima consultazione 19/06/2022).

Il problema della spiegabilità si pone però anche su un piano più generale e deriva dalle caratteristiche stesse delle tecnologie data driven, che si basano sull'osservazione di correlazione statistiche tra dati e sull'apprendimento automatico da questi (e non sul tradizionale metodo deduttivo). A ciò si aggiunga il frequente e già menzionato ricorso a "data from the wild" (si pensi all'uso di dati raccolti dai social media)[59], tanto che diventa arduo rendere spiegabile l'intelligenza artificiale che faccia uso del *machine learning* e praticamente impossibile quanto al *deep learning*. Un tentativo di spiegabilità dei risultati prodotti dal machine learning va comunque impostata già in sede di addestramento di un modello al fine, quantomeno, di rendere evidente graficamente l'importanza delle diverse variabili di input[60].

Non esclusività della decisione algoritmica: criticità

La non esclusività della decisione algoritmica (o sorveglianza umana nel linguaggio europeo) significa che occorre garantire l'intervento umano quando la decisione sia frutto di discrezionalità (in base alle pronunce del Consiglio di Stato), così come per i sistemi di intelligenza artificiale classificati come ad alto rischio (secondo la proposta di regolamento europeo)[61].

[59] Ad esempio, N. Cristianini, *Shortcuts to Artificial Intelligence*, cit., evidenzia che «when we communicate with other users on social networks, or we access databases of content, it would be useful to drop the pretense that this is a direct interaction, making instead explicit the presence of an intelligent agent acting as intermediator- so that we can explicitly decide which engagement actions are communicative acts aimed at the other users, and which ones are aimed at the recommending agent».

[60] «One family of approaches fundamentally attempts to describe how important different input variables are to the resulting predictions. Within this family, some methods operate on a global, or "algorithmwide," level; they do not ask how important certain input variables are to generating a prediction for a given individual on which a running model is deployed, but how important they were to the algorithm's accuracy during training across many individuals. The output of such methods is known as a variable importance plot, which displays graphically the relative importances of the different input variables. […] The other kind of importance-measuring methods attempts to operate on the individual level, explaining what the most important variables were for a given individual's predictions. But these methods are particularly novel and have yet to be thoroughly tested» (Lehr, Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, cit., pp. 708-709).

[61] Ad esempio, «i sistemi di IA destinati a essere utilizzati dalle autorità pubbliche o per conto di autorità pubbliche per valutare l'ammissibilità delle persone fisiche alle prestazioni e ai servizi di assistenza pubblica, nonché per concedere, ridurre, revocare o recuperare tali prestazioni e servizi», «i sistemi di IA destinati a essere utilizzati dalle autorità di contrasto per effettuare valutazioni individuali dei rischi delle persone fisiche al fine di determinare il rischio di reato o recidiva in relazione a una persona fisica o il

Si tratta di un presidio fondamentale, che resta però indeterminato nella sua dimensione attuativa[62]. Gli interrogativi che si pongono sono numerosi, tra questi, non solo quanto intervento umano vada assicurato[63], ma soprattutto come evitare che il decisore finale si appiattisca sulle risultanze istruttorie elaborate da sistemi di intelligenza artificiale, che si tratti di un funzionario richiesto di una licenza o di un ispettore chiamato a valutare quale impresa sottoporre a controllo.

L'intervento umano è utile se effettivamente in grado di valutare l'adeguatezza delle indicazioni, se non di realizzare una "sorveglianza umana" sul processo che ha portato a queste. Stante la difficoltà oggettiva di quanto sembra richiedersi ai funzionari, la proposta di regolamento europeo fa ricadere sugli sviluppatori l'obbligo di rendere possibile la "sorveglianza umana" per i sistemi di intelligenza artificiale ad alto rischio[64]. Ma anche in questi termini appare un obiettivo di difficile realizzazione, almeno per gli stessi motivi per cui è difficile assicurare la spiegabilità quando siano in uso tecnologie di *deep learning*.

Il presidio dell'intervento umano rischia peraltro di restare lettera morta anche per motivi che nulla hanno a che fare con la tecnologia, quanto con il funzionamento del cervello umano. Sono ipotizzabili, a fronte della limitata razionalità degli individui, due possibili opposte reazioni dell'intelligenza umana di fronte ai dati derivanti dall'intelligenza artificiale: da un lato una eccessiva deferenza[65], dall'altro diffidenza se non rifiuto. A falsare questo rapporto entrano in gioco veri e propri *bias* cognitivi[66],

rischio per vittime potenziali di reati», «i sistemi di IA destinati a essere utilizzati dalle autorità pubbliche competenti per verificare l'autenticità dei documenti di viaggio», «i sistemi di IA destinati ad assistere le autorità pubbliche competenti nell'esame delle domande di asilo, di visto e di permesso di soggiorno e dei relativi reclami per quanto riguarda l'ammissibilità delle persone fisiche che richiedono tale status» (allegato II alla proposta di regolamento europeo).

[62] B. Marchetti, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, in *BioLaw J.*, 2, 2021, p. 367 ss.

[63] Il giudice amministrativo italiano non lo spiega e la definizione che ne dà il Gruppo di esperti indipendenti ad alto livello sull'intelligenza artificiale istituito dalla Commissione europea è alquanto vaga: "la sorveglianza [umana] può avvenire mediante meccanismi di governance che consentano un approccio con intervento umano (human-in-the-loop-HITL), [vale a dire] «la possibilità di intervento umano in ogni ciclo decisionale del sistema, che in molti casi non è né possibile né auspicabile» (*Orientamenti etici per una IA affidabile*, Bruxelles, 2019, p. 18).

[64] L'art. 14, comma 3, della proposta di regolamento europeo (COM(2021)206 def.).

[65] D.S. Rubenstein, *Acquiring Ethical AI*, 73 *Florida L. Rev*. 797 (2021).

[66] N. Rangone, *Le pubbliche amministrazioni italiane alla prova dell'intelligenza artificiale: problemi e prospettive,* in *Studi parlamentari e di politica costituzionale*, 209,

come l'*automation bias*[67] (menzionato anche dalla proposta di regolamento europeo)[68] che comporta un completo affidamento nei confronti di una presunta oggettività dell'indicazione derivante dall'intelligenza artificiale. Se si dovesse verificare l'operatività di questo *bias*, la differenza tra un sistema completamente automatizzato e uno in cui ha un ruolo l'elemento umano si perderebbe completamente[69].

È anche possibile che si manifesti un atteggiamento opposto di rifiuto, che può derivare da sfiducia nella tecnica (*algoritm aversion*)[70] o da limitate competenze tecnologiche[71], oppure da un vero e proprio *bias* di avversione all'algoritmo. Potrebbe, inoltre, entrare in gioco quella che è stata definita *illusion of validity*[72] (un'ingiustificata fiducia nel ragionamento umano

2021, pp. 22-25. ivi ampi riferimenti dottrinari ai suddetti *bias*.

[67] L'*automation bias* viene descritto come "the use of automation as a heuristic replacement for vigilant information seeking and processing" (L.J. Skitka et al., *Automation Bias and Errors: Are Crews Better Than Individuals?*, in *The Int'l J. Aviation Psychology*, 10, 1, 2000, p. 85). Di particolare interesse a questo riguardo sono gli studi e gli esperimenti svolti con riferimento alle reazioni alle indicazioni delle "macchine" dei piloti di aereo (K.L. Mosier et al., *Automation Bias: Decision Making and Performance in High-Tech Cockpits*, in *Int'l J. Aviation Psychology*, 8, 1, 1997, p. 47 ss.) e in medicina (K. Goddard, A. Roudsari, J.C. Wyatt, *Automation bias: a systematic review of frequency, effect mediators, and mitigators*, in *J. American Medical Informatics Ass'n*, 19, 2021, p. 121 ss.).

[68] Così l'art. 14 comma 4, lett. a) e d) della proposta di regolamento europeo, già citata, che evidenzia anche che vanno predisposte misure affinché le persone cui è affidata al sorveglianza umana siano consapevoli «della possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio ("distorsione dell'automazione"), in particolare per i sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche» (lett. b).

[69] «We are worried that if we simply thrust the human at the output end of the running model, there is very little she can do to root out bias. The human becomes a rubber stamp for the machine, providing nothing more than a cosmetic reason to lull ourselves into feeling better about the results» (D. Lehr, P. Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, cit., p. 716). Sul punto anche Citron, (*Technological Due Process*, cit., p.1272) che evidenzia anche come «over time, human operators may lose the skills that would allow them to check a computer's recommendation».

[70] B.J. Dietvorst, J.P. Simmons, C. Massey, *Algorithm aversion: People erroneously avoid algorithms after seeing them err*, in *J. of Experimental Psychology: General*, 144, 1, 2015, p. 114 ss., p. 10-11 in Scholarly Commons.

[71] F. De Leonardis, *Big Data, decisioni amministrative e "povertà" di risorse della pubblica amministrazione*, in E. Calzolaio (a cura di), *La decisione nel prisma dell'intelligenza artificiale*, Milano, 2020, p. 159.

[72] A. Tversky, D. Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, in *Science*, 185, 4157, 1974, p. 1126.

che spesso caratterizza gli esperti)[73] o il *confirmation bias*[74], che induce a tener conto dei dati elaborati dall'intelligenza artificiale quando a supporto dell'intuizione o convinzione formatasi dal funzionario ed a rifiutarli in caso di difformità.

Seppure non sia possibile dare indicazioni definitive e generali rispetto al ruolo dei bias nel processo decisionale del funzionario, non vi è dubbio che queste diverse eventualità vadano analizzate e affrontate con strumenti adeguati. Una volta constato il potenziale ruolo di determinati *bias* (secondo alcuni nell'ambito della valutazione di impatto dell'algoritmo)[75] le diverse reazioni potrebbero essere affrontate con strumenti di empowerment cognitivo[76], per aumentare la consapevolezza della loro esistenza.

Ciò potrebbe avvenire con una formazione mirata sui *bias* che intercorrono più frequentemente nel rapporto uomo-macchina[77]. Inoltre, un sistema di accountability interna o esterna che porti il funzionario a individuare e indicare i dati generati dalla tecnologia che lo hanno indotto ad adottare una determinata decisione potrebbe aumentarne la consapevolezza[78]. Infine, un ruolo importante hanno il monitoraggio e

[73] D. Kahneman, G. Klein, *Conditions for Intuitive Expertise. A Failure to Disagree*, in *American Psychologist*, 2009, p. 517.

[74] Si tratta di un *bias* molto diffuso, che interessa gli individui in quanto destinatari delle regole e in quanto regolatori. Sul punto la letteratura è diffusa, si veda, per tutti, C. Taylor, *Biased Assimilation: Effects of Assumptions and Expectations on the Interpretation of New Evidence*, in *Social and Personality Psychology Compass*, 5, 3, 2009, p. 827 ss.; E. Zamir, D. Teichman, *Behavioural law and economics*, Oxford, 2018, p. 399; S. Stern, *Cognitive Consistency: Theory Maintenance and Administrative Rulemaking*, 63 *University of Pittsburgh L. Rev*. 589-591 (2002).

[75] Per un'analisi di impatto che cerchi di individuare specificatamente queste reazioni si veda l'art. 6 della proposta di Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration elaborate dall'European Law Institute nel 2022: «an assessment of the specific and systemic impact of the system on: […] iii. the administrative authority itself, in particular the estimated acceptance of the system and its decisions by the staff, the risks of over- or under-reliance on the system by the staff, the level of digital literacy, and specific technical skills within the authority» (consultabile in: <https://www.europeanlawinstitute.eu/news-events/upcoming-events/events-sync/news/eli-issues-guidance-on-the-use-of-algorithmic-decision-making-systems-by-public-administration/?tx_news_pi1%5Bcontroller%5D=News&tx_news_pi1%5Baction%5D=detail&cHash=f4a2a4a677e3dcf6e391d9f0a2a9bd6a>).

[76] Sulla distinzione tra *nudging* and *empowerment* sia consentito rinviare a F. Di Porto, N. Rangone, *Behaviouralm Sciences in Practice: Lessons for EU Policymakers*, in A. Alemanno, A.-L. Sibony (a cura di), *Nudge and the Law: A European Perspective?*, Oxford, 2015, p. 29.

[77] Citron, *Technological Due Process*, cit., pp. 1306-1307.

[78] M. Hallsworth, M. Egan, J. Rutter, J. Mccrae, *Behavioural Government. Using*

la valutazione ex post degli esiti dell'uso dell'intelligenza artificiale alla luce di diversi parametri, tra i quali va inclusa l'accettazione nel sistema organizzativo.

Esigenza di un quadro normativo minimo, da specificare a livello di singola amministrazione

Come già evidenziato, le amministrazioni italiane sperimentano e utilizzano l'intelligenza artificiale per esigenze conoscitive, a supporto di servizi al pubblico, nei processi decisionale per l'adozione di policies, regole, decisioni amministrative, così come nei controlli. Numerosi sono i vantaggi potenzialmente ricavabili, sia in termini di risparmi di risorse economiche ed umane, che di interventi

più mirati anche in termini prospettici. Dunque si può dire che le nuove tecnologie supportino il diritto a una buona amministrazione.

Non mancano però i rischi, in parte gestibili, a condizione che se ne riconosca l'esistenza e se affrontati nelle sedi adeguate, senza adagiarsi su facili soluzioni formali (come il riconoscimento del diritto di accesso esteso al codice sorgente dell'algoritmo, in quanto soluzione generalizzata anche per le tecnologie più avanzate). Vi è in particolare da chiedersi se le tutele tradizionali, seppure rafforzate, possano ritenersi sufficienti, o se non debba immaginarsi un nuovo diritto amministrativo dello Stato digitale[79].

Ragionare in questa direzione significa, ad avviso di chi scrive, intervenire su più livelli, uno normativo generale e uno di singola amministrazione.

Un inquadramento normativo generale minimo dell'uso dell'intelligenza artificiale nei processi decisionali pubblici[80] appare necessario sia per individuare le ipotesi in cui si possa ricorrere a decisioni automatizzate (previsione, per il vero, che discenderà e sarà riconducibile al regolamento europeo una volta approvato), sia per sancire regole quadro volte ad assicurare trasparenza, motivazione (dunque anche spiegabilità) e verificabilità dei sistemi ed esplicitare le regole minime da rispettare per rendere effettivi tali principi.

La trasparenza degli utilizzi potrebbe, in particolare, essere realizzata sia attraverso i siti delle singole amministrazioni, sia con un *repository* centrale (strumenti non necessariamente alternativi) che, con funzione di single access point, dia informazioni sui sistemi di intelligenza artificiale in uso,

behavioural science to improve how governments makedecisions, The Behavioural Insights Team, 2018.

[79] L. Torchia, *Lo Stato digitale e il diritto amministrativo*, in *Liber amicorum per Marco D'Alberti*, 2022, p. 477.

[80] S. Civitarese Matteucci, *Public Administration Algorithm Decision-Making and the Rule of Law*, in *European Public Law*, 27, 1, 2021, p. 103 ss.

sui motivi dell'utilizzo e i relativi rischi, su due livelli di approfondimento. Quanto alle modalità di declinazione della trasparenza, potrebbe essere impostato un approccio su due livelli, un primo molto semplice e intuitivo, un secondo più dettagliato e da specificare in sede di disciplina di singola amministrazione, che comprenda i dati utilizzati, la logica che ne è alla base, i rischi correlati e i sistemi di gestione del rischio.

La motivazione delle regole e delle decisioni in cui abbia avuto un ruolo l'intelligenza artificiale dovrebbe essere arricchita con l'indicazione (sintetica) delle fasi di utilizzo, le caratteristiche dei sistemi di intelligenza artificiale utilizzati, le ragioni dell'utilizzo, i dati lavorati e i meccanismi predisposti per la supervisione del relativo funzionamento.

Tale inquadramento generale dovrebbe, poi, prevedere che vengano indicati (a livello di singola amministrazione) i presidi da predisporre per la verificabilità dei sistemi (come audit, monitoraggio e valutazione *ex post*) e le clausole minime da inserire nei bandi di gara quando la tecnologia sia acquistata sul mercato.

Spetterà, inoltre, alla disciplina delle singole amministrazioni indicare, eventualmente, chi decide del ricorso a nuove tecnologie, se avviare una sperimentazione interna o acquistare sul mercato, gli indicatori per il monitoraggio e la valutazione ex post degli esiti dell'uso dell'intelligenza artificiale nei processi decisionali alla luce di parametri rilevanti per l'amministrazione di riferimento, ma che dovrebbero comunque comprendere la rilevazione degli errori riscontrati[81] e un'analisi controfattuale ove possibile[82]. In altre parole, la decisione di fare ricorso all'intelligenza artificiale o all'intelligenza umana supportata dall'intelligenza artificiale non andrebbe presa una volta per tutte, ma dovrebbe essere oggetto di valutazione alla luce dei risultati raggiunti[83].

[81] R. Cavallo Perin, I. Alberti, *Atti e procedimenti amministrativi digitali*, in R. Cavallo Perin, D.U. Galetta, *Il diritto dell'amministrazione pubblica digitale*, Giappichelli, 2020, p. 153. Freeman Engstrom, Ho, *Algorithmic Accountability in the Administrative State*, cit., propongono di impostare un "prospective benchmarking", vale a dire una comparazione tra casi decisi con il supporto dell'intelligenza artificiale e frutto di un processo decisionale tradizionale «This "human alongside the loop" approach provides critical information and a comparison set to help smoke out when an algorithm has gone astray, when encoding the past may miss new trends, when an algorithm may create disparate impact, or when "automation bias" causes excessive deference to machine outputs».

[82] «Regulators should develop tests for considering when the outcomes an algorithm creates are impermissible, based on regulatory policy goals. Regulators should begin by asking meaningful questions that can be answered by examining algorithmic outcomes, such as whether similarly situated borrowers are treated differently or whether the move from traditional pricing to algorithmic pricing has increased disparities» (Gillis, *The input fallacy*, cit.).

[83] C. Coglianese, A. Lai, *Digital Versus Human Algorithms*, 71 *Duke L. J.* 1281 (2022).

Edoardo Chiti, Barbara Marchetti, Nicoletta Rangone

The use of Artificial Intelligence by Italian Administrations Dress Rehearsal

L'impiego di sistemi di intelligenza artificiale nelle pubbliche amministrazioni italiane: prove generali

Abstract: This article aims at mapping and analyzing the ever increasing recourse by Italian public administration to AI systems. It focuses on three different types of administrations: independent authorities, smart cities and central administrations. For each group of administrations, it asks through which proceedings a decision concerning the use of AI systems is taken, which AI systems are in the process of being experimented and for which purposes, who controls the functioning of the AI. Particular attention is paid to administrative practices and legal and institutional reality.

Abstract: Questo scritto mira a fotografare e analizzare l'affacciarsi delle amministrazioni italiane sul mondo dell'IA, per individuare i processi in corso e alcune tendenze generali. Si considerano tre diversi tipi di amministrazioni: le autorità indipendenti, le smart cities e le amministrazioni centrali. Per ciascun gruppo di amministrazioni, si affrontano le seguenti domande: chi decide, all'interno dell'amministrazione, di ricorrere all'IA? Quali sono i tipi di sistemi che le amministrazioni stanno concretamente sperimentando o utilizzando? Per quali compiti e con quali obiettivi? E chi assicura la sorveglianza della macchina, correggendone gli eventuali difetti di funzionamento? Particolare attenzione è posta, nell'indagine, alle effettive pratiche amministrative e alla realtà giuridica e istituzionale.

* Questo articolo è stato originariamente pubblicato in *Biolaw Journal*, n. 2/2022, p. 489-507. Contributo sottoposto a doppio referaggio anonimo.

1. *Tre problemi*

Il processo di digitalizzazione che attraversa la società ha un impatto via via crescente sull'amministrazione e sul modo in cui essa opera. Non solo comporta la dematerializzazione dell'attività pubblica e dei rapporti tra autorità amministrative e privati, con intenti di semplificazione e modernizzazione, ma porta con sé anche l'impiego di strumenti e tecnologie digitali che modificano il potere e i modi del suo esercizio. In particolare, le amministrazioni italiane hanno cominciato a sperimentare e utilizzare applicazioni di intelligenza artificiale (IA) più o meno sofisticate per esigenze conoscitive generali, per comunicare con i cittadini (ad esempio tramite chatbot), per acquisire dati necessari a indirizzare le proprie attività e policy, per emanare le proprie decisioni.

I sistemi di IA, però, oltre che fornire alle autorità pubbliche significativi vantaggi in termini di efficienza e conoscenza, grazie a strumenti potenti di analisi dei dati e di predizione, portano con sé dei rischi, su cui da tempo si è concentrata l'attenzione della scienza giuridica, non solo italiana[1]: si

[1] C. Coglianese, *Administrative Law in the Automated State*, in *Public Law and Legal Theory Research Paper Series*, Research Paper No. 21-15, disponibile alla pagina <https://scholarship.law.upenn.edu/faculty_scholarship/2273>; C. Coglianese, D. Lehr, *Regulating by robot: Administrative Decision Making in the Machine Learning Era*, 105 Georgetown L. J. 1147 (2017); M.U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies and Strategies*, 29 *Harvard J. L. & Technol*. 353, 2016; D. Freeman, D.E. Ho, C.M. Sharkey, M.-F. Cuéllar, *Government by Algorithm: Artificial Intelligence*, in *Federal Administrative Agencies*, disponibile alla pagina <https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf.> Anche la letteratura italiana in tema di IA e diritto (pubblico) è ormai molto vasta. Si vedano, ex multis, G. Avanzini, *Decisioni amministrative e algoritmi informativi. Predeterminazione analisi predittiva e nuove forme di intellegibilità*, Napoli, 2019; A. Santosuosso, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, Milano, 2020; A. Pajno et al., *AI: profili giuridici. Intelligenza artificiale: criticità emergenti e nuove sfide per i giuristi*, in *Biolaw J*., 3, 2019, p. 205 ss.; A. Simoncini, *Profili costituzionali della amministrazione algoritmica*, in *Rivista trimestrale di diritto pubblico*, 4, 2019, p. 1149 ss.; J.-B. Auby, *Il diritto amministrativo di fronte alle sfide digitali*, in *Istituzioni del federalismo*, 2019, 3, p. 619 ss.; L. Torchia, *Lo Stato digitale e il diritto amministrativo*, in *Liber Amicorum per Marco D'Alberti*, Torino, 2022; N. Rangone, *Le pubbliche amministrazioni italiane alla prova dell'intelligenza artificiale*, in *Liber Amicorum per Marco D'Alberti*, Torino, 2022; F.F. Pagano, *Pubblica amministrazione e innovazione tecnologica, relazione al Convegno Associazione Gruppo di Pisa*, *Genova, 18-19 giugno 2021 su Il diritto costituzionale e le sfide dell'innovazione tecnologica*; M. Bassini, L. Liguori, O. Pollicino, *Sistemi di intelligenza artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018; C. Casonato, *Costituzione e intelligenza*

tratta di rischi per la privacy, per la sicurezza dei cittadini, per i diritti fondamentali, per gli stessi principi democratici.

L'IA è infatti, nelle sue applicazioni più avanzate, opaca, difficile da comprendere, non sempre affidabile e talvolta discriminatoria (a causa dei *bias* che può incorporare), sicché non è affatto certo che il suo utilizzo sia compatibile con i principi di trasparenza, motivazione, partecipazione e imparzialità ai quali l'azione amministrativa è soggetta.

Il presente contributo si inserisce in un modo specifico nella riflessione giuridica sull'uso dei sistemi di IA da parte delle amministrazioni. Il suo obiettivo è quello di fotografare e analizzare, nell'attuale momento storico, l'affacciarsi delle amministrazioni italiane sul mondo dell'IA, per individuare i processi in corso e alcune tendenze generali. Esso mira, cioè, a dare conto di come le autorità amministrative si stiano effettivamente muovendo per dotarsi di sistemi di IA utili per lo svolgimento dei loro compiti.

I problemi considerati, in particolare, sono tre. Il primo riguarda la governance del processo e il quadro normativo di riferimento: chi decide, all'interno dell'amministrazione, di ricorrere all'IA? Quali sono le pratiche utilizzate per dotarsi di un sistema di IA? Le amministrazioni, ad esempio, ricorrono al mercato oppure vi provvedono in house? E in quale quadro di regole o linee guida un'applicazione di IA viene implementata dall'amministrazione?

Il secondo problema è relativo al quomodo: quali sono i tipi di sistemi che le amministrazioni stanno concretamente sperimentando o utilizzando? E per quali compiti e con quali obiettivi? Ad esempio, le

artificiale: un'agenda per il prossimo futuro, in *Biolaw J.*, *Special Issue*, 2, 2019, p. 711 ss.; C. Casonato, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, 2019, p. 101 ss.; L. Casini, *Lo Stato nell'era di Google. Frontiere e sfide globali*, Milano, 2020; B. Marchetti, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, in *Biolaw J.*, 2021, 2, p. 367 ss.; F. Costantino, *Rischi e opportunità del ricorso delle amministrazioni alle predizioni dei Big Data*, in *Diritto pubblico*, I, 2019, p. 43 ss.; S. Civitarese Matteucci, *Umano, troppo umano. Decisioni amministrative automatizzate e principio di legalità*, in *Diritto pubblico*, 2019, p. 5 ss.; E. Picozza, *Intelligenza artificiale e diritto. Politica, diritto amministrativo and artificial intelligence*, in *Giurisprudenza italiana*, 7, 2019, p. 1657 ss.; F. Donati, *Intelligenza artificiale e giustizia*, in *Rivista AIC*, 1, 2020, p. 415 ss.; A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw J.*, 1, 2019; M.C. Cavallaro, G. Smorto, *Decisione pubblica e responsabilità dell'amministrazione nella società dell'algoritmo*, in *Federalismi.it*, 6, 2019; S. Tranquilli, *Rapporto pubblico-privato nell'adozione e nel controllo della decisione amministrativa algoritmica*, in *Diritto e società*, 2, 2020, p. 281 ss.; G. Orsoni, E. D'orlando, *Nuove prospettive dell'amministrazione digitale: Open Data e algoritmi, in Istituzioni del Federalismo*, 3, 2019, 593 ss.

applicazioni di IA sono utilizzate per finalità di comunicazione con il cittadino o nel procedimento decisorio? Si tratta di algoritmi *rule based* o anche di algoritmi *machine learning*? Quali sono i dati in possesso delle amministrazioni e come sono utilizzati per sviluppare gli algoritmi?

Il terzo e ultimo problema riguarda il controllo: chi assicura la sorveglianza della macchina, correggendone gli eventuali difetti di funzionamento, intervenendo sugli output e, nel caso, premendo lo stop botton? Queste attività sono effettivamente previste all'interno delle singole amministrazioni? Nel caso, sono conferite al programmatore del sistema o al funzionario che se ne avvale per lo svolgimento dei propri compiti? Se i compiti di sorveglianza sono attribuiti al secondo, il personale amministrativo ha le competenze tecniche necessarie?

Per raccogliere elementi utili a rispondere a queste domande, si sono interrogate le stesse pubbliche amministrazioni. Operando come gruppo di lavoro nell'ambito della più ampia ricerca Astrid sui profili giuridici della «rivoluzione dell'intelligenza artificiale»[2], si sono individuate alcune amministrazioni rappresentative di tre specifiche componenti del sistema amministrativo italiano, diverse per caratteristiche organizzative e funzionali. La prima componente è quella delle autorità indipendenti, che è stata esaminata attraverso le esperienze della Banca d'Italia, della Commissione nazionale per le società e la borsa (Consob), dell'Autorità di regolazione per energia reti e ambiente (Arera) e dell'Autorità per le Garanzie nelle Comunicazioni (Agcom). La seconda è quella delle amministrazioni centrali, esplorata a partire dai casi del Ministero della giustizia, dell'Agenzia delle entrate e dell'Istituto nazionale della previdenza sociale (INPS). La terza componente è quella delle smart cities, studiata attraverso le esperienze di Venezia, Milano, Padova e Trento. Le pratiche di queste amministrazioni sono state ricostruite per mezzo di una interlocuzione diretta, invitando le stesse amministrazioni a misurarsi con le domande sopra richiamate nel contesto di quattro seminari: i primi due, svoltosi nel maggio e nel giugno del 2021, hanno riguardato la Banca d'Italia, Consob, Arera e Agcom[3]; nel terzo, tenuto nel mese di dicembre 2021, sono state discusse le esperienze di Venezia, Milano, Padova e Trento[4];

[2] Per una descrizione dei lavori del gruppo di ricerca, si rinvia a E. Chiti, B. Marchetti E N. Rangone, *Il progetto: L'uso dell'intelligenza artificiale nel sistema amministrativo italiano*, in *BioLaw J. – Rivista di BioDiritto*, 4, 2021, p. 209 ss.

[3] I risultati sono presentati nel *Rapporto 1/2021 – L'impiego dell'IA nell'attività di CONSOB, AGCOM e ARERA e nel Rapporto 2/2021 – L'impiego dell'IA nell'attività di Banca d'Italia*, in *BioLaw J. – Rivista di BioDiritto*, 4, 2021, rispettivamente p. 211 ss. e 229 ss.

[4] Si veda il *Rapporto 3/2022 – SMART cities e intelligenza artificiale*, in *BioLaw J. – Rivista*

nel quarto, tenutosi nel gennaio 2022, sono state esaminate le pratiche del Ministero della giustizia, dell'Agenzia delle entrate e dell'INPS[5].

I limiti di questo approccio sono evidenti. Le tre componenti rappresentano solo una parte del sistema amministrativo italiano. Le amministrazioni considerate, a loro volta, non possono essere considerate esemplari dei modelli ai quali sono riconducibili, considerate le variabili e le differenziazioni interne a tali modelli. Le risposte emerse nel corso dei seminari, inoltre, potrebbero essere parziali o incomplete.

Attraverso l'analisi di questi casi e la loro discussione nella forma specifica di una interlocuzione con le stesse amministrazioni, tuttavia, è possibile almeno avviare un'indagine sulle reali pratiche di impiego dell'IA da parte delle amministrazioni italiane, verificare se amministrazioni riconducibili a modelli che presentano tratti organizzativi e funzionali molto diversi tra loro procedono nella stessa direzione o seguono linee di sviluppo differenti, individuare alcune tendenze comuni che spetterà ad ulteriori fasi della ricerca confermare e articolare.

I risultati di questa indagine saranno presentati, nelle pagine seguenti, secondo l'ordine dei problemi sopra indicati: si partirà, dunque, dai modi di acquisizione dei sistemi di IA (§ 2), per poi passare al loro utilizzo nell'esercizio delle varie attività amministrative (§3) e alla sorveglianza sul loro funzionamento (§4). Un breve paragrafo conclusivo ricapitolerà le principali conclusioni (§5).

2. *Come si acquisiscono i sistemi di intelligenza artificiale?*

Rispetto alla prima delle tre questioni, le indicazioni che sono emerse dall'interlocuzione con le amministrazioni possono essere presentate distinguendo tra le esperienze delle autorità indipendenti, delle amministrazioni centrali e delle *smart cities*.

2.1. *Come si acquisiscono i sistemi di intelligenza artificiale?*

Dalle testimonianze raccolte presso le autorità indipendenti (in particolare Banca d'Italia, Consob, Agcom) si ricavano, nonostante la

di BioDiritto, 1, 2022, p. 253 ss.

[5] *Rapporto 4/2022 - Intelligenza artificiale e amministrazioni centrali*, in *BioLaw Journal – Rivista di BioDiritto*, 4, 2021, p. 261 ss.

varietà delle scelte operate, due tendenze comuni, quella alla collaborazione inter-funzionale, sia in fase di pianificazione che in fase di valutazione dei sistemi, e quella alla auto-produzione, cioè allo sviluppo e alla produzione *in house* delle applicazioni di IA, seppure in qualche caso ricorrendo a *partnership* e collaborazioni con università e centri di ricerca.

In particolare, il dato mostra come il processo di digitalizzazione prediliga applicazioni disegnate su misura e dall'interno delle istituzioni. Gli esperti delle unità funzionali (vigilanza bancaria e finanziaria, informazione finanziaria, politica monetaria) individuano le nuove esigenze informatiche per la facilitazione e l'ottimizzazione dei loro compiti e le comunicano agli esperti informatici, i quali cercano di trovare la soluzione tecnica per soddisfarle. La cooperazione tra le diverse funzioni avviene non solo nella fase di pianificazione vera e propria, ma anche nel momento della sperimentazione del sistema.

Tali interazioni avvengono in una cornice che dovrebbe orientare e guidare le decisioni complessive: in Banca d'Italia, è stato definito un *Framework* operativo per l'intelligenza artificiale/*machine learning* (ML), che appare fondamentale per la definizione di tali scelte. Queste ultime dipendono da una valutazione volta ad accertare il valore aggiunto rappresentato dal ML rispetto alla informatica tradizionale: in tale valutazione, si deve tenere conto sia della complessità dello sviluppo e della manutenzione di sistemi di IA, sia dei rischi rispetto a *bias* e *fairness*, sia ancora dei pericoli che il sistema generi *output* errati, inutili o basati su dati incompleti o non di qualità: rischi, questi ultimi, che sarebbero assenti nell'impiego di algoritmi deterministici o di tecnologie informatiche tradizionali. Le enormi potenzialità dell'IA devono essere soppesate con la componente di rischio.

La medesima collaborazione inter-funzionale è alla base delle scelte che governano l'adozione di sistemi di IA anche in Consob, ma in questa amministrazione un ruolo importante è svolto dallo *Steering Committee Fintech*, affidato ad uno dei componenti del Collegio, cui spetta lo studio dell'impatto della digitalizzazione sulle principali aree di competenza dell'autorità e l'individuazione delle soluzioni tecniche.

La sperimentazione e la valutazione dell'applicazione concreta restano tuttavia affidate all'interazione tra gli esperti dell'unità funzionale in cui il sistema è destinato ad operare e i tecnici del dipartimento di informatica.

La necessità di applicazioni ritagliate sulle esigenze specifiche delle diverse unità funzionali spiega il ricorso, sia in Banca d'Italia, che in Consob alla auto-produzione. I sistemi di IA, dunque, non sono reperiti

sul mercato ma per lo più sviluppati in house, sfruttando le competenze informatiche interne, eventualmente in collaborazione con enti di ricerca e università. La rinuncia all'outsourcing può avere diverse spiegazioni: la prima può essere rappresentata dal peculiare regime dei dati di cui si nutre l'algoritmo: se i dati sono riservati la loro cessione a terzi per lo sviluppo e la produzione del software diventa problematica. Ad esempio, le tecniche di machine learning cui sta lavorando l'Unità di informazione finanziaria della Banca d'Italia per misurare il rischio di riciclaggio degli intermediari finanziari utilizzano dati estremamente riservati, che sarebbero incompatibili con l'affidamento a soggetti privati, al punto da aver richiesto l'esclusione della condivisione di dati in chiaro con lo stesso Dipartimento di informatica della Banca d'Italia. Un secondo vantaggio è rappresentato dall'autonomia e dalla specificità assicurate dall'auto-produzione. Le soluzioni tecniche concepite in base ad una collaborazione inter-funzionale sono «cucite su misura» per le specifiche esigenze rappresentate dalle diverse unità dell'autorità.

Inoltre, per amministrazioni con queste dotazioni, l'*in house* pare una soluzione preferibile anche dal punto di vista economico finanziario, qualora l'amministrazione sia in grado di fornire un prodotto competitivo rispetto ad applicazioni simili presenti sul mercato. Ovviamente si tratta di un'opzione non sempre percorribile, soprattutto se si considerano le amministrazioni meno attrezzate sia finanziariamente, sia in termini di competenze tecnico-informatiche. Per queste ultime, la realizzazione *in house* potrebbe risultare inadeguata rispetto ai prodotti forniti dal mercato, avere costi comunque elevati (anche quanto all'aggiornamento) e tempi di elaborazione troppo dilatati. Nel caso di Arera, per esempio, è stata praticata l'opzione della ricerca sul mercato.

2.2. *Le amministrazioni centrali: disallineamenti*

Le esperienze di Inps, Agenzia delle entrate e Ministero della giustizia mostrano stadi di sviluppo e modalità di azione parzialmente divergenti. Le prime due istituzioni sono accomunate da un processo di digitalizzazione molto avanzato e dalla sperimentazione (e messa in servizio) di chatbot e sistemi di IA consolidati, sia in funzione antifrode, che di lotta all'evasione fiscale. Il Ministero della giustizia, invece, è contraddistinto da un processo di digitalizzazione lento e disomogeneo nei diversi settori della giustizia (sostanzialmente completato per quella tributaria e amministrativa, in fase di avvio per la giustizia civile e penale), con evidenti ricadute in termini di

possibile impiego, in tale ambito, di applicazioni di IA. La digitalizzazione dei dati, la loro anonimizzazione e pseudo-anonimizzazione e il loro accesso libero costituiscono, infatti, i presupposti imprescindibili per lo sviluppo e la messa in servizio di IA, a partire dai sistemi di giustizia predittiva, sicché al momento si registra un considerevole ritardo rispetto alle amministrazioni più avanzate sul fronte della tecnologia digitale, come quelle francesi[6].

Tali divergenze, peraltro, sono comprensibili se si considera una differenza importante tra le prime due amministrazioni e il Ministero della giustizia: Inps e Agenzia delle entrate vantano un approccio unitario e centralizzato sulle scelte in materia di tecnologie dell'informazione e della comunicazione (ICT) e IA, mentre il Ministero fatica a svolgere un ruolo di direzione unitaria, in ragione della struttura estremamente

[6] In Francia, una riforma voluta nel 2016 (*loi* n. 2016/1321) su *La République Numérique* (Stato digitale) ha previsto la messa a disposizione gratuita del pubblico di tutti i dati in possesso della p.a. e di tutte le decisioni dei giudici. Ovviamente, questo ha comportato il libero accesso a moltissimi dati giudiziari e dunque la possibilità di algoritmi di giustizia predittivi per predire il possibile esito di una controversia. Preoccupato dei possibili effetti di tale apertura, il legislatore francese ha stabilito sia un divieto di profilazione dei giudici (intervenendo, dunque, sul lato degli utenti della giustizia), sanzionato con la pena della reclusione fino a 5 anni (*loi* n. 2019/222), sia il divieto di decisioni giudiziarie che si fondino su valutazioni del comportamento di una persona basate su un trattamento automatizzato di dati personali riguardanti aspetti della personalità di tale persona (*loi* n.78/78 come modificata dalla legge 2018/493). Simili divieti, al contrario, non sono stabiliti, ad esempio, né negli Stati Uniti, né nel Regno Unito, dove sono stati realizzati programmi che, servendosi degli open data giudiziari, profilano i magistrati per ricostruire il loro pensiero giuridico e giudiziario, al fine di impostare su base predittiva la difesa (*Ravel Law*) ovvero che forniscono al giudice, come abbiamo visto nel caso di COMPAS, predizioni sulla probabilità di recidiva di un imputato. Anche in Cina, algoritmi di giustizia predittiva trovano impiego al di fuori di una qualsiasi cornice normativa. In argomento cfr. R.W. Campbell, *Artificial Intelligence in the Courtroom: The Delivery of Justice in the Age of Machine Learning*, 18.2 *Colorado Technology L. J.* 323, 2020; E. Volokh, *Chief Justice Robots*, 68 *Duke L. J.*, 2019; R.E. Stern, B.L. Liebman, M. Roberts, A.Z. Wang, *Automating Fairness? Artificial Intelligence in the Chinese Court*, 59 *Columbia J. Transnational L.* 515, 2021; J. Deng, *Should the Common Law System Welcome Artificial Intelligence: A Case Study of China's Some-type Case Reference System*, 3 *Georgetown L. Technol. Rev.* 223, 2019; E. Niiler, *Can AI Be a Fair Judge in Court? Estonia Thinks So*, Wired, Mar. 25, 2019, scaricabile al sito <https://www.wired.com/story/can-AI-be-fair-judge-court-estonia-thinks-so>; T. Kerikmäe E E. Pärn-Lee, *Legal Dilemmas of Estonian artificial intelligence strategy: in between of e-society and global race*, in *AI & Society*, 36, 2021, 561 ss.; F. G'sell, *Les progress à petits pas de la "justice predictive" en France*, ERA Forum, 21, 2020, 299 ss. In generale, in prospettiva europea, si veda il noto lavoro di A. Garapon E J. Lassègue, *La giustizia digitale. Determinismo tecnologico e libertà*, Bologna, 2021.

articolata, differenziata e decentrata dell'amministrazione della giustizia che ha favorito, tra l'altro, l'avvio di sperimentazioni autonome di IA da parte di alcuni uffici giudiziari. La spinta che si produrrà per effetto dei finanziamenti del Piano Nazionale di Ripresa e Resilienza (PNRR) accelererà il processo di digitalizzazione e favorirà presumibilmente la regia ministeriale di raccolta e analisi dei dati. Al momento, però, non sono maturi i tempi per la progettazione e la messa in servizio di sistemi di giustizia predittiva.

Dal punto di vista del reperimento dei sistemi di IA, l'Agenzia delle entrate e l'Inps seguono percorsi parzialmente differenti. Nella prima è la Direzione tecnologia e innovazione che, su sollecitazione del vertice, dota l'Istituzione degli strumenti digitali più adatti allo svolgimento dei compiti dell'agenzia. A tal fine, la Direzione si avvale della consulenza e della collaborazione della SOGEI, ovvero della Società generale di informatica S.p.A. (il cui capitale è detenuto al 100% dal MEF), la quale, a sua volta, può sia rivolgersi al mercato, indicendo specifiche gare, che utilizzare sistemi open. Il reperimento del sistema di IA vede quindi, in questo caso, l'interposizione tra l'Agenzia e il mercato di un soggetto che, in ragione della sua missione, risulta particolarmente attrezzato per governare la scelta tecnologica.

Quanto all'Inps, la cui sperimentazione di sistemi di IA nasce da un confronto «dal basso» sulle specifiche esigenze dell'amministrazione, il canale principale di fornitura delle applicazioni tecnologiche è stato rappresentato, fino ad ora, dalla Consip, anche se nella programmazione dell'istituto decine di progetti, caratterizzati da esigenze di manutentabilità (e di innovatività) significative, preludono al possibile reperimento dell'IA sul mercato, anche grazie ai finanziamenti PNRR destinati a coprirne i costi.

2.3. *Le* smart cities*: gara pubblica o auto-produzione*

Le esperienze delle città intelligenti indagate nella ricerca sono riconducibili a due modelli alternativi di acquisizione di IA: la gara pubblica o l'auto-produzione.

Un esempio significativo del primo tipo è la gara internazionale bandita dalla città di Venezia per la realizzazione di una piattaforma di controllo della città lagunare (control room) al fine di raccogliere tutti i dati, provenienti dalla municipalità, ma anche da soggetti privati, relativi al traffico presente in laguna, sia pedonale, sia acqueo, sia stradale. Il dato interessante riguarda il rapporto tra le scelte operate dall'amministrazione

(che ha bandito la procedura) e le opzioni tecniche concretamente attuate dalla società che si è aggiudicata la gara (*Mindicity*). La procedura indetta dall'amministrazione è stata regolata dalla disciplina del parternariato per l'innovazione (art. 65 del d.lgs. 50 del 2016, Codice degli appalti), prevista dal Codice per il caso in cui la pubblica amministrazione debba procurarsi prodotti, servizi e lavori innovativi che non sono reperibili sul mercato. L'obiettivo che l'amministrazione intendeva conseguire era la realizzazione di una piattaforma di *Urban Intelligence*, capace di raccogliere ed elaborare una enorme quantità di dati utile ad affrontare gli annosi problemi di mobilità della città.

La società aggiudicataria ha dimostrato di possedere una tecnologia sofisticata, innovativa e potente, capace di rilevare in tempo reale il traffico pedonale, tramite alcuni sensori acustici, quello dei canali, tramite l'installazione di 80 telecamere, e quello stradale di accesso alla laguna. Rispetto alle modalità di raccolta dei dati e sul loro uso, tuttavia, si registra una discrasia informativa tra quanto è stato presentato e valutato in sede di gara, da un lato, e i dettagli operativi e pratici relativi alla fase di esecuzione della prestazione, dall'altro. Nel caso specifico, ad esempio, l'amministrazione non si è confrontata con la società circa le modalità di rilevazione dei flussi, né sono stati discussi e affrontati i possibili profili di rischio, legati alla violazione della privacy prodotta dalla installazione di videocamere o alle possibili implicazioni (anche sul fronte sanzionatorio) correlate alla registrazione dei flussi nei canali per eccesso di velocità. Tali aspetti sono emersi e hanno formato oggetto di discussione solo in un secondo momento, nella fase esecutiva del contratto, a dimostrazione delle difficoltà dell'amministrazione, in sede di gara, di prevedere in modo esaustivo i caratteri che il sistema deve avere, anche sotto il profilo della accuratezza, robustezza, sicurezza.

La soluzione di ricorrere al mercato, in altre parole, può significare una prevalenza della «visione» privata su quella pubblica, qualora la pubblica amministrazione non si mostri capace o non abbia le competenze per comprendere a fondo le modalità e i contenuti dell'applicazione di IA, per controllarne gli aspetti di rischio, per individuarne le criticità anche con riguardo alle possibili violazioni di diritti. Di qui l'importanza di una elaborazione accorta del bando e dei capitolati di gara e della fase di aggiudicazione.

Il secondo modello, dell'autoproduzione, è stato invece prescelto da Trento smart city, che lo ha preferito alla ricerca nel mercato: qui le diverse applicazioni di IA sono state sviluppate in house con la collaborazione della Fondazione Bruno Kessler (FBK), un ente di ricerca della costellazione

pubblica, e nell'ambito di progetti europei volti a promuovere mobilità sostenibile, sistemi di illuminazione intelligente, algoritmi per la sicurezza degli attraversamenti pedonali (ad esempio *Stardust* e *See Roads* 2).

In questa ipotesi, vale quanto rilevato sopra in termini di vantaggi legati ad autonomia, controllo e specificità dei sistemi. Anche in ragione della dimensione dell'ente e delle limitate competenze interne, però, la collaborazione con istituti di ricerca ad alta competenza tecnica come FBK risulta un presupposto fondamentale per l'auto-produzione, considerata anche la necessità di individuare soluzioni tecniche competitive.

2.4. *Il problema delle competenze*

Il principale problema che emerge dal quadro che si è tratteggiato è quello delle competenze. In tutti i casi considerati, è indubbio che l'amministrazione procedente debba poter fare affidamento su adeguate competenze interne sia quando decide di ricorrere al mercato, perché solo così può governare la scelta tecnologica e controllarne le implicazioni in sede di gara, sia quando procede all'auto-produzione, dovendo sviluppare software competitivi, anche sotto il profilo della manutenzione e dell'aggiornamento. Il dato della expertise tecnica interna all'amministrazione, dunque, è imprescindibile, quale che sia il metodo di acquisizione prescelto per dotarsi di IA.

In questa direzione spinge, del resto, anche il processo di graduale costruzione del quadro giuridico. Il mondo dell'IA manca ancora di una disciplina generale, così come mancano regole che governano l'impiego di software di IA da parte delle autorità pubbliche. Tuttavia, è attualmente in corso di discussione la proposta di regolamento presentata dalla Commissione, che prefigura una disciplina complessiva della materia destinata a riguardare anche le amministrazioni pubbliche[7]. Se la proposta sarà approvata, il regolamento stabilirà il doveroso rispetto da parte di chi fornisce un sistema di IA di requisiti rigorosi in tema di data governance, di sicurezza e accuratezza, di informazione, di sorveglianza umana oltre che l'assoggettamento a una procedura di conformità quale condizione per l'immissione nel mercato o la messa in servizio del sistema. Il suo impatto sulle amministrazioni sarà significativo, considerato che genererà obblighi di conformità alla disciplina europea soprattutto per le amministrazioni che decidano di sviluppare in house le applicazioni di IA di cui necessitino. Gli oneri derivanti dall'applicazione del regolamento costituiranno, quindi, un

[7] COM (2021) 206.

elemento da soppesare in vista dell'alternativa mercato/auto-produzione.

Inoltre, il Codice dell'amministrazione digitale stabilisce già, all'art. 68, i criteri che l'amministrazione è tenuta a seguire per decidere come procedere all'acquisizione di programmi informatici (e IA)[8]. Si tratta, in particolare, della valutazione dei costi complessivi (di produzione e manutenzione) del sistema, dei vantaggi che esso comporta in termini di interoperabilità e cooperazione applicativa e delle garanzie che il fornitore è in grado di dare in materia di livelli di sicurezza, conformità alla normativa, protezione dei dati personali, servizio (tenuto conto della tipologia di software acquisito). La preferenza del legislatore è però per i sistemi aperti o in auto-produzione, dal momento che l'accesso a software di tipo proprietario (con licenza d'uso) è ritenuto ammissibile solo se l'autorità dimostra motivatamente l'impossibilità di accedere a soluzioni già disponibili all'interno dell'amministrazione o a software liberi[9].

La decisione sui metodi di acquisizione dei sistemi di IA, dunque, è complessa e tutt'altro che neutra: essa può avere effetti sul piano organizzativo interno (effetti crescenti, alla luce del futuro regolamento europeo) e sul piano dei rapporti con i cittadini; deve tenere conto dei costi complessivi, ma al contempo favorire l'apertura e l'interoperabilità; è chiamata a reperire tecnologie innovative (e facilmente manutenibili) e, allo stesso tempo, ben cucite sulle esigenze delle specifiche funzioni; deve assicurare il rispetto della privacy e al tempo stesso sfruttare i vantaggi offerti dalla grande disponibilità di dati pubblici. Si tratta di un quadro articolato, che conferma l'esigenza che le amministrazioni si dotino al più presto delle competenze necessarie a governare efficacemente l'acquisizione degli strumenti di IA.

3. *Quali impieghi e per quali scopi?*

Anche rispetto alla seconda delle tre questioni poste al centro dell'indagine, è possibile presentare le indicazioni emerse distinguendo tra le esperienze delle autorità indipendenti, delle amministrazioni centrali e delle smart cities.

[8] L'art. 68, co. 1, del Codice prevede sei modalità di acquisizione: l'autoproduzione, riutilizzo di software elaborati in autoproduzione; accesso a *software* liberi o codici sorgente aperti; acquisto di software di tipo proprietario (mediante ricorso a licenza d'uso); utilizzo di software che sono la combinazione delle precedenti soluzioni.

[9] Art. 68, co. 1-ter.

3.1. *Le autorità indipendenti: diverse velocità*

Quanto alle autorità indipendenti prese in considerazione, i regolatori dei mercati finanziari vantano una significativa esperienza di uso di nuove tecnologie nelle attività di vigilanza, mentre sembra più limitata quella dei regolatori dei servizi pubblici.

Nel caso della Banca d'Italia, i cui sistemi di IA sono realizzati in house sulla base di una ben definita struttura organizzativa, le tecnologie utilizzate vanno dal *natural language processing* per analizzare messaggi postati su social media (al fine costruire indicatori di sentiment, ad esempio, per misurare aspettative di inflazione o costruire *early woarning indicator* sulle banche italiane) o per processare esposti della clientela a fini di vigilanza. Viene fatto ricorso al machine learning per prevedere, ad esempio, probabilità di default di imprese italiane, per la classificazione automatica delle operazioni sospette, per la rilevazione di indicatori di rischi di riciclaggio o di infiltrazioni mafiose. *Neural networks* possono essere utilizzati per fare previsioni di variabili, come la produzione industriale. Il *web scraping*, consente, ad esempio, di fare ricerche su *Google trend* per prevedere fenomeni come la disoccupazione.

La Consob, che realizza anch'essa sistemi in house e ha optato per una governance della digitalizzazione finanziaria diffusa e coordinata da uno *Steering Commettee Fintech*, vanta numerose e diversificate esperienze. Quanto alla vigilanza sui prodotti finanziari, una prima sperimentazione è stata volta alla selezione dei documenti sintetici (cartacei in formato pdf) che illustrano le caratteristiche dei prodotti finanziari PRIIPs rivolti agli investitori al dettaglio. Il prototipo, sviluppato nel 2019 in collaborazione con una università appositamente selezionata, si basa sull'applicazione di tecniche di *natural language processing* per estrarre dai documenti parole e concetti chiave. È attualmente allo studio l'uso del machine learning per effettuare uno screening automatico e selezionare i documenti su cui potrà lavorare l'analista.

Per contrastare l'offerta abusiva *on line* di attività finanziarie riservate, è in fase di sviluppo un progetto di *intelligent crawling*: un motore di ricerca basato sul *machine learning* analizza il web per individuare le piattaforme *on line* interessate da abusi di mercato e attraverso l'elaborazione del linguaggio naturale viene effettuato il *text mining* degli esposti. In prospettiva, il *machine learning* dovrebbe portare a una individuazione autonoma (non più sulla base degli esposti) dei siti web attraverso i quali sono svolte attività abusive.

La vigilanza dell'ordinato svolgimento delle negoziazioni sui mercati per prevenire, individuare e sanzionare abusi si svolge da tempo attraverso analisi statistiche e algoritmi che lanciano *alert*. L'obiettivo è definire sistemi intelligenti che ottimizzino le analisi condotte per dare seguito ai segnali di anomalia. Questo *supervised machine learning* in grado di replicare i ragionamenti degli analisti è in corso di definizione in collaborazione con una università italiana.

L'Agcom, invece, ha sviluppato un sistema di prevenzione e *detection* algoritmica del linguaggio d'odio sui servizi media e piattaforme *on line* attraverso l'IA. In particolare, l'Autorità svolge analisi dell'informazione on line attraverso tecniche di *natural language processing* per rilevare l'uso di espressioni di odio (*hate speech*) e i reati connessi. I risultati di questa attività sono messi a disposizione del pubblico attraverso l'Osservatorio sulla disinformazione online (il cui terzo e ultimo numero risale però al 2020)[10].

In prospettiva, l'IA potrebbe essere utilizzata dall'Agcom e dall'Arera nei procedimenti di regolazione (ad esempio per processare le risposte ricevute in sede di consultazione o le denunce) e per svolgere controlli sulle imprese (ad esempio, per analizzare i dati stoccati dagli operatori di servizio al fine di verificare il rispetto dei parametri di qualità).

3.2. *Le amministrazioni centrali: una pluralità di tecnologie e di funzionalità*

Anche rispetto alle amministrazioni centrali si registrano diversi gradi di avanzamento nell'impiego di sistemi di IA. Se Inps e Agenzia delle entrate segnalano vari usi di *machine learning*, *deep learning* e *natural language processing*, l'amministrazione della giustizia sembra trovarsi a uno stadio precedente volto alla digitalizzazione dei dati[11].

Le esperienze dell'Agenzie delle entrate e dell'Inps mostrano diversi usi dell'intelligenza volti, rispettivamente, a migliorare i servizi all'utenza, a razionalizzare e rendere più efficienti le procedure interne, a supportare i

[10] Si veda <https://www.agcom.it/documents/10179/19226924/Documento+generico+29-06-2020/3b8d1a2d-61fc-4865-b5b0-bb6343933465?version=1.0>.

[11] Le sperimentazioni sull'uso di nuove tecnologie per l'amministrazione della giustizia si svolgono prevalentemente a livello decentrato e sono basato sulla collaborazione con università; questa frammentazione ha il limite, tra gli altri, di operare su campioni limitati di dati, laddove l'intelligenza artificiale opera utilmente con big data. Sarebbe dunque importante una centralizzazione presso il Ministero della giustizia.

processi decisionali e di controllo.

Con riferimento all'IA al servizio del pubblico, sia l'Agenzia delle entrate che l'Inps utilizzano chatbot intelligenti per l'assistenza agli utenti. Ad esempio, Inps mette a disposizione assistenti virtuali per guidare utenti finali (cittadini o professionisti)[12], basati su motori di ricerca cognitivi capaci di comprendere il linguaggio – spesso eterogeneo – dell'utente, nel rispetto delle indicazioni del Garante della privacy (per evitare profilazioni è operato un *matching* per segmenti di utenza). Un esempio di *chatbot* basata sull'IA e volta a orientare il cittadino nella prestazione di un servizio è quella relativa alla Nuova assicurazione sociale per l'impiego-NASPI (in produzione). Per la migliore fruibilità del sito istituzionale, l'Inps sta sperimentando motori di ricerca «cognitivi», basati su meccanismi di deep learning e *natural language processing* e oggetto di sperimentazione (ad esempio, il «portale della genitorialità» per l'assistenza all'apertura di pratiche che riguardano i bonus bebè e la genitorialità).

Quanto all'IA per ottimizzare le procedure interne, può essere menzionato il sistema di smistamento a livello nazionale delle pratiche INPS (che consente un processo decisionale completamente automatizzato di *resource planning*), la classificazione automatica delle PEC-Inps, l'analisi predittiva attraverso *machine learning* sulla mediabilità delle controversie Inps. L'Agenzia delle entrate riporta l'uso del *text mining* per la categorizzazione automatica o semiautomatica dei documenti basata sul riconoscimento di pattern all'interno di testi non strutturati così da facilitare l'estrazione di dati di interesse.

La funzione più delicata da un punto di vista dei diritti che possono essere incisi è quella dell'IA a supporto delle decisioni e dei controlli. Agenzia delle entrate e Inps segnalano varie applicazioni, che non sostituiscono mai l'intervento umano. Ad esempio, l'Agenzia entrate utilizza un applicativo informatico per applicare il redditometro, così come il cosiddetto «evasometro anonimizzato». Dal 2005, Inps utilizza il *data mining* e *machine learning* per rilevare i fenomeni fraudolenti nella lotta all'evasione contributiva.

[12] «Supporto all'utente (es. su richieste sullo stato delle pratiche, sull'invio di documentazione). Supporto all'operatore (es. scadenze previste dall'iter procedurale, risoluzione di problematiche procedurali/instradamento ticket mediante «chatbot»)», slides INPS quarto seminario.

3.3. *Le* smart cities*: la collaborazione con i privati*

L'esperienza delle smart cities conferma che la varietà delle tecnologie e delle funzionalità utilizzate da alcune amministrazioni italiane.

La piattaforma open source *MindIcity*[13], ad esempio, offre alla città di Venezia – e, in prospettiva, ad altre città – una *Smart Control Room* che, attraverso *machine learning* e *deep learning*, lavora dati raccolti attraverso l'*internet of things* e di provenienza diversa. Alcuni sono «prodotti dalla città» attraverso videocamere intelligenti, sensori collocati in aree strategiche (pedonali, stradali e autostradali) come già evidenziato; altri forniti da soggetti terzi di natura privata, come le società TIM e Abertis. Al progetto collaborano l'Università di Bologna e società specializzate, come Engine[14].

Anche a Padova una importante spinta all'uso dell'intelligenza per la realizzazione di una smart city viene dalla collaborazione con privati e università. La prospettiva è di utilizzare dati raccolti con l'*internet of things* (da telecamere collocate da soggetti pubblici e privati) o con rilevazioni tradizionali (come le informazioni sulla sicurezza stradale contenute nelle comunicazioni dei taxisti alle cooperative di riferimento), poi lavorati con l'IA per progetti di interesse locale o regionale, come la mobilità sostenibile.

A Trento, le nuove tecnologie per la città, sviluppate in collaborazione con università e centri di ricerca, lavorano dati pubblici e dati raccolti con l'*internet of things*, come sensori nell'asfalto per dare informazioni in tempo reale sulla disponibilità di parcheggi, telecamere dislocate sul territorio comunale per individuare pericoli potenziali (e consentire, quando poste su piste ciclabili, l'illuminazione al passaggio di utenti in zone poco frequentate), gestione intelligente dei semafori, sicurezza degli attraversamenti pedonali.

I dati raccolti dalle smart cities consentono l'offerta di nuovi servizi ai cittadini (che potrebbero essere offerti tramite app o siti istituzionali), come l'indicazione della disponibilità di posti nei parcheggi, la concentrazione di traffico di veicoli o pedonale in una determinata zona, la qualità dell'acqua per la balneazione. In prospettiva, la resa di informazioni potrebbe essere utilizzata per la modifica di comportamenti o stili di vita in linea con gli obiettivi dell'agenda 2030 delle Nazioni Unite.

I dati possono essere (e in alcuni casi vengono) utilizzati anche a supporto dell'enforcement, come negli esempi di applicazioni di polizia

[13] Si veda <https://dilservice.it/it/soluzioni/smart-city>.

[14] Si veda <https://www.fabbricadigitale.com/smart-control-room-venezia-mindicity/>.

predittiva, per la rilevazione della violazione di limiti velocità o del mancato uso delle cinture di sicurezza.

I dati raccolti dalle smart cities, infine, possono supportare, in prospettiva, l'adozione di politiche e la verifica di politiche pubbliche (ad esempio per la sostenibilità) e la definizione di pianificazioni (come quelle per trasporti intermodali e a limitato impatto sull'ambiente).

3.4. *La rilevanza delle condizioni e le questioni aperte*

Le indicazioni raccolte suggeriscono tre considerazioni generali, relative alle condizioni che facilitano l'uso di strumenti di IA e alle questioni aperte.

La prima considerazione riguarda gli incentivi (economici e non) all'impiego dell'IA nelle pubbliche amministrazioni.

Il ruolo dei finanziamenti o co-finanziamenti europei è rilevante nel supporto all'introduzione di sistemi di IA sia a livello locale che nelle amministrazioni centrali e indipendenti. Tanto Padova quanto Trento sottolineano l'impulso dato dai progetti europei, che hanno favorito, tra l'altro, partenariati pubblico privati (ad esempio, un algoritmo in uso a Trento è stato messo a disposizione da un'impresa coreana).

Nelle amministrazioni centrali, la Commissione europea supporta dal 2021 l'Agenzia delle entrate in un progetto per l'adozione di tecniche di IA nel contrasto all'evasione; il programma operativo nazionale-PON Legalità 2014 ha consentito di censire l'uso dell'IA e già menzionava l'uso del machine learning per l'antifrode INPS. La *detection* algoritmica del linguaggio d'odio da parte di Agcom ha ricevuto impulso anche da un progetto finanziato dalla Commissione europea (*Innovative Monitoring Systems and Prevention Policies of Online Hate Speech-IMSyP*) che ha portato l'autorità a lavorare nell'ambito di un consorzio di ricerca internazionale. Importanti sono poi, per tutte le amministrazioni, il RRF e il PNRR.

Gli incentivi di natura non economica, invece, derivano dal confronto in sede internazionale. Si pensi al premio UNESCO che ha riconosciuto il sistema INPS di smistamento delle PEC come tra i migliori dieci progetti sviluppati per l'amministrazione. Per Banca d'Italia, gli scrutini periodici di *mutual evaluation* realizzati a livello OCSE nel Gruppo intergovernativo d'Azione Finanziaria Internazionale-GAFI possono essere occasione per lo scambio di buone pratiche anche con riferimento all'uso dell'IA nella vigilanza (ad esempio, in questa sede è stato valutato positivamente il sistema di indicatori del rischio di riciclaggio). Di estrema rilevanza è inoltre la partecipazione a sistemi europei o reti di regolatori. Ad esempio,

nell'ambito del sistema europea di banche centrali si stanno sviluppando sistemi di digitalizzazione dei procedimenti di vigilanza (*suptech*), a livello sia centralizzato che decentralizzato presso le autorità nazionali, così da avere a disposizione in prospettiva strumenti omogenei. I progetti Consob sono oggetto di confronto in sede europea nell'ambito dell'Autorità europea degli strumenti finanziari e dei mercati (ESMA).

In secondo luogo, si conferma, nel caso dell'IA acquisita sul mercato, il ruolo cruciale dei bandi di gara, già evidenziato in precedenza. Con riferimento all'esigenza di accountability degli sviluppatori, dalle esperienze delle smart cities emerge che una stretta collaborazione di questi con l'ente locale si attiva talvolta a monte gara (come nel caso del bando monopattini, in cui il comune di Trento ha chiesto che i dati raccolti venissero messi a disposizione in formato aperto e standard e tramite interfacce di programmazione delle applicazioni-API), ma più spesso a valle della selezione (come nel caso di Venezia, anche se ciò può essere riconducibile alla tipologia di procedura di selezione utilizzata). A fronte del coinvolgimento di soggetti terzi, è cruciale assicurare la trasparenza e l'*accountability* dell'operatività del sistema di IA attraverso chiare indicazioni nei bandi di gara[15]. A questo riguardo, linee guida per la redazione dei bandi sono state definite dal governo inglese[16] e dalla città di Amsterdam[17], nell'ottica di assicurare la verifica della qualità dei dati, la trasparenza del funzionamento dell'algoritmo e tale da poter essere oggetto di audit e spiegabilità al soggetto pubblico, così come l'impostazione di sistemi di gestione del rischio. Il Canada ha invece optato per una lista di fornitori pre-selezionati anche in base al rispetto di «demonstrated competence in AI ethics»[18].

[15] Si veda, in particolare, il documento di Ada Lovelace Institute, AI Now Institute and Open Government Partnership, *Algorithmic Accountability for the Public Sector*, 2021, disponibile alla pagina https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/, dove si legge: «Establishing contractual preconditions for acquiring algorithmic systems ensures that systems that do not comply with specific conditions of transparency or fairness are not acquired or used by governments, or that, if a vendor fails to meet contractual conditions, they are subject to contractual liability. Procurement conditions also allow for interventions in the design of algorithmic systems, as well as during their use» (33-35 e 44-45).

[16] *Guidelines for AI procurement 2020*, in <https://www.gov.uk/government/publications/guidelines-for-ai-procurement>.

[17] City of Amsterdam, *Standard Clauses for Municipalities for Fair Use of Algorithmic Systems*, 2020 (<https://www.amsterdam.nl/innovatie/digitalisering-technologie/contractual-terms-for-algorithms/>).

[18] Cfr. <https://buyandsell.gc.ca/procurement-data/tender-notice/PW-EE-017-34526>.

La terza considerazione generale ha ad oggetto la trasparenza e l'accountability dell'amministrazione nei confronti di cittadini e imprese. Le informazioni sugli usi di sistemi di IA da parte delle pubbliche amministrazioni non sono facilmente reperibili da parte di cittadini e imprese, che si tratti del controllo dei parcheggi, di polizia predittiva, di verifiche del corretto pagamento di tributi o contributi, dell'individuazione di abusi di mercato. Le informazioni a disposizione del pubblico, infatti, derivano in ampia misura da notizie di stampa (come nei casi polizia predittiva), da generiche comunicazioni istituzionali[19], circolari[20] o altri documenti[21]. L'informazione sulle tecnologie utilizzate e le relative funzionalità andrebbe messa disposizione del pubblico in forma semplificata e facilmente comprensibile, con possibilità di approfondimento per gli interessati[22]. Interessante al riguardo il registro delle applicazioni di IA della città di Helsinki[23], che spicca per chiarezza e facile consultabilità da parte dei cittadini.

[19] Ad esempio, la direzione studi e ricerche INPS, in base al sito istituzionale «fornisce supporto tecnico-scientifico all'elaborazione delle decisioni che l'Istituto assume nell'ambito delle proprie attività istituzionali attraverso […] l'elaborazione di statistiche, di modelli di data mining e machine learning, anche in riferimento ai big data»: <https://www.inps.it/nuovoportaleinps/default.aspx?itemdir=53263>.

[20] Si veda, ad esempio, INPS, circolare n. 23/2010, Funzione di accertamento e verifica amministrativa – Attuazione del nuovo modello organizzativo delle strutture territoriali di produzione previsto dalla circolare n. 102 del 12/08/2009 e circolare n. 23/2010, Funzione di accertamento e verifica amministrativa - Attuazione del nuovo modello organizzativo delle strutture territoriali di produzione previsto dalla circolare n. 102 del 12/08/2009.

[21] Come nel caso del cd. risparmiometro, menzionato nel Piano della performance dell'Agenzia delle entrate 2018-20, ma poi descritto solo in una decisione del Garante della Privacy del 20 luglio 2017, n. 321, Sperimentazione di una procedura basata sull'utilizzo di informazioni fornite dall'Archivio dei rapporti finanziari e degli elementi presenti nell'Anagrafe tributaria per l'individuazione di profili di evasione rilevanti.

[22] Interessanti le soluzioni anche grafiche prospettate, ad esempio, in materia di polizia predittiva dal rapporto Britainthinks, *Complete transparency, complete simplicity. How can the public sector be meaningfully transparent about algorithmic decision making?*, 17 giugno 2021, 9.

[23] Si vedano le informazioni disponibili alla pagina <https://ai.hel.fi/en/ai-register/>, dove si chiarisce che «AI Register is a window into the artificial intelligence systems used by the City of Helsinki. Through the register, you can get acquainted with the quick overviews of the city's artificial intelligence systems or examine their more detailed information based on your own interests. You can also give feedback and thus participate in building humancentered AI in Helsinki». Si veda anche D. Marchetti, *L'intelligence finanziaria: tre esempi di applicazione, in Rapporto 2/2021*, cit., p. 242 ss., che richiama il progetto della Unità di Informazione Finanziaria per l'Italia volto alla costruzione di un *blind learning environment*, nel quale è possibile «testare le tecniche di classificazione senza "vedere" i dati sottostanti» (242).

Infine, si registra una evoluzione della tutela della riservatezza: da ostacolo allo sviluppo a esempio di applicazione dell'IA. In alcuni contesti, il timore di violare la normativa sulla privacy ostacola la condivisione, come nel caso riportato dell'operatore di trasporto urbano della città di Padova che potrebbe meglio organizzare il servizio in base alle esigenze delle scuole se avesse a disposizione i dati relativi ai flussi degli studenti. In realtà, una risposta viene data proprio attraverso l'uso dell'IA: sempre a Padova, i dati raccolti con telecamere vengono anonimizzati con tecnologie di machine learning, similmente a quanto avviene a Venezia, dove viene memorizzato solo il metadato generato e non tutto il flusso video. Le esigenze del rispetto della privacy vengono affrontate anche dalle amministrazioni che producono internamente i sistemi di IA, perché i dati lavorati non possono necessariamente essere visibili da dipartimenti diversi di una stessa istituzione. Per superare questo problema, la Banca d'Italia ha creato un *blind learning environment*, che consente di evitare la condivisione di dati «in chiaro» all'interno della Banca stessa. Anche l'Inps sfrutta l'IA per simulare (replicare) i dati per poi anonimizzarli ma in modo da consentire da decriptazione, ove necessaria.

4. *Chi controlla la macchina?*

Il quadro tratteggiato nelle pagine precedenti va completato considerando un ultimo aspetto dell'impiego dei sistemi di IA nelle amministrazioni italiane, quello relativo alla sorveglianza sul loro effettivo funzionamento: a controllare la macchina è il programmatore del sistema o un funzionario amministrativo?

In questo secondo caso, come è individuato e quali sono le competenze richieste per lo svolgimento dei suoi compiti? E in cosa consistono esattamente le attribuzioni attraverso le quali si realizza la sorveglianza sulla macchina?

4.1. *Una tendenza unitaria:* Human Out of the Loop

La risposta che emerge dall'indagine svolta è relativamente chiara. Le amministrazioni considerate in questa ricerca sperimentano strumenti di IA al di fuori di un quadro normativo che definisca in modo chiaro l'uso di tali strumenti all'interno del procedimento e il ruolo che deve rivestire

il decisore umano. Si tratta di una tendenza comune, che prescinde dalle specificità organizzative e funzionali delle amministrazioni prese in esame e dai modelli ai quali sono riconducibili.

Le autorità indipendenti, ad esempio, riconoscono l'esigenza che gli strumenti di IA siano soggetti a un processo di «validazione», ma tendono a risolvere tale processo nei termini puramente funzionali di una verifica della efficacia degli strumenti rispetto ai risultati attesi, individuati nel miglioramento delle attività istituzionali: si pensi, tra gli altri, all'approccio S.M.A.R.T. della Consob, che richiede, appunto, di «testare» ogni tipo di strumento di IA per provarne l'efficacia e la validità[24]; e *al Framework operativo per l'intelligenza artificiale/machine learning della Banca d'Italia*, l'architettura IT già menzionata sviluppata dal Dipartimento informatica della stessa Banca d'Italia e volta, tra le altre cose, a definire le funzionalità che la piattaforma deve offrire, a individuare il valore aggiunto che l'IA può apportare a nuovi business case, misurato nei termini di uno score chiamato «AI-ness», e a facilitare il *testing* di modelli di *machine learning*[25]. L'esigenza che sia assicurato un controllo umano sui vari passaggi del funzionamento della macchina è talora riconosciuta, ma si pone l'accento sulle difficoltà che tale controllo incontrerebbe: con riferimento all'Autorità per le garanzie nelle comunicazioni, in particolare, si osserva come l'operatore verrebbe a conoscenza di dati che «possono essere coperti da privacy o segreto commerciale, così come possono esserlo gli algoritmi»; e si rileva che «l'apprendimento automatico (*machine learning*, *deep learning*) si basa su tecniche "non esplicite" di programmazione dei computer», che complicano in misura significativa l'interpretazione dei risultati[26]. In ogni caso, l'esigenza di un controllo umano sul funzionamento della macchina non si traduce mai nella definizione di specifici requisiti, tanto meno di requisiti definiti da norme giuridiche. Vi sono d'altra parte, pratiche operative che paiono suscettibili di ulteriori sviluppi e potrebbero rappresentare degli esempi da seguire: è il caso del sistema di monitoraggio dei mercati energetici all'ingrosso, sviluppato dall'Agenzia europea di collaborazione dei regolatori dell'energia – ACER e utilizzato come strumento di «intelligenza condivisa» da ACER e Arera,

[24] Togna, *L'approccio SMART di Consob alla digitalizzazione finanziaria, in Rapporto 1/2021 – L'impiego dell'intelligenza artificiale nell'attività di CONSOB, AGCOM e ARERA*, cit., p. 213 ss.

[25] Federico, *AI e prospettive di evoluzione dell'architettura informatica, in Rapporto 2/2021 – L'impiego dell'intelligenza artificiale nell'attività di Banca d'Italia*, cit., pp. 229-230.

[26] Ragucci, *AGCOM e l'intelligenza artificiale per fronteggiare l'hate speech, in Rapporto 1/2021*, cit., p. 222 ss.

insieme ai regolatori indipendenti degli altri Stati membri dell'Unione, il cui utilizzo prevede una procedura di verifica manuale, svolta in più fasi e volta a individuare errori del modello, validare i casi sospetti utilizzando informazioni aggiuntive a disposizione dell'autorità nazionale e ad aggiornare il modello sulla base dei difetti riscontrati[27].

Un discorso non diverso vale per le amministrazioni centrali e per le smart cities considerate nell'indagine. Anche in queste amministrazioni, infatti, il processo di «validazione» degli strumenti di IA è condotto secondo criteri strettamente funzionali e mancano meccanismi formalizzati in disposizioni giuridiche o consolidati sul piano operativo e volti a garantire la sorveglianza umana sul funzionamento della macchina. Rispetto all'esperienza delle autorità indipendenti, tuttavia, in alcune amministrazioni centrali pare registrarsi una maggiore attenzione per l'esigenza dell'intervento umano e di una formalizzazione in un quadro di regole giuridiche: è il caso, in particolare, dell'Agenzia delle entrate, per la quale si è riconosciuta tanto la necessità di una disciplina degli algoritmi di IA collegati all'adozione di provvedimenti amministrativi, quanto l'opportunità di stabilire il principio che l'esperto di dominio sia il funzionario amministrativo responsabile dell'adozione delle misure amministrative capaci di produrre effetti sugli amministrati, anche in funzione dell'*accountability* dell'amministrazione[28]. Nel caso delle smart cities, al contrario, il distacco dalle esigenze di controllo umano è ancora maggiore di quello che caratterizza l'esperienza delle autorità indipendenti: le piattaforme di *urban intelligence* sperimentate in alcune città, esemplificate dal caso di Venezia, sembrano prescindere dalla sorveglianza umana sui loro modi di funzionamento; e in ogni caso tale sorveglianza non sarebbe svolta dall'amministrazione, ma dai privati che sviluppano e gestiscono le piattaforme[29].

4.2. *Uno sviluppo problematico*

Una simile tendenza, comune a tutte le amministrazioni considerate nella ricerca, si spiega a partire dalla logica essenzialmente funzionale che

[27] Lo Schiavo e Lazza, *ARERA e l'«intelligenza assistita», l'«intelligenza condivisa» e l'«intelligenza della filiera»,* in *Rapporto 1/202*, cit., p. 224 ss., p. 225.

[28] G. Buono, *L'esperienza dell'Agenzia delle Entrate, in Rapporto 3/2021 – L'impiego dell'intelligenza artificiale nell'attività di Agenzia delle entrate*, INPS e Ministero della giustizia, cit.

[29] È quanto emerge dalle esperienze presentate nel Rapporto 4/2021 – *L'impiego dell'intelligenza artificiale nell'attività delle Smart Cities*, cit.

sta alla base del ricorso delle amministrazioni ai sistemi di IA. Quelle che emergono dall'indagine, infatti, sono amministrazioni che vogliono sperimentare tecnologie innovative per migliorare la qualità e l'efficacia delle proprie attività. Le amministrazioni, in altri termini, si orientano verso i sistemi di IA perché questi ultimi permettono di innovare radicalmente i modi di svolgimento delle attività di cui sono responsabili e i risultati attesi, indipendentemente dalle specificità funzionali di ciascuna amministrazione. Sia che procedano direttamente, attraverso i propri servizi informatici, come nel caso della Banca d'Italia, sia che si rivolgano ai privati, come avviene spesso per le *smart cities*, le amministrazioni accettano sin dall'inizio che la dinamica interna del processo di sviluppo di sistemi di intelligenza risponda alla logica della innovazione. La prospettiva complessiva non è quella, essenzialmente difensiva, della gestione e attenuazione dell'impatto sull'amministrazione e sugli amministrati di una tecnologia ritenuta pericolosa, ma piuttosto quella della sperimentazione di una tecnologia fortemente abilitante, capace, cioè, di esercitare una forza trasformatrice e positiva sulle attività di ciascuna amministrazione. In questo quadro, il parametro per la validazione del sistema di IA è sempre inevitabilmente funzionale: è la capacità del sistema di raggiungere risultati corretti che deve essere verificata e accertata; e la sorveglianza sul funzionamento della macchina non è una garanzia da assicurare in quanto tale, ma solo uno dei vari strumenti che possono essere utilizzati, in alcune specifiche occasioni, per controllare la correttezza degli output del sistema. Nella medesima prospettiva, poi, il diritto rileva come una forza limitante. Vi è la consapevolezza che i principi e le regole del diritto amministrativo tradizionale non siano facilmente applicabili a una realtà tecnica molto diversa da quella a partire dalla quale tali principi e regole sono stati costruiti. Ma questa consapevolezza non spinge a definire un nuovo quadro regolatorio capace di inquadrare e governare l'uso dell'IA. Piuttosto, il diritto rappresenta una forza che contrasta quella dell'innovazione. Se la costruzione di un insieme di regole adatto alla nuova realtà è un processo ritenuto inevitabile, esso rappresenta comunque un secondo momento di un processo che deve ora lasciare spazio alla innovazione tecnica[30].

Allo stesso tempo, lo sviluppo di cui si è dato conto è ovviamente problematico. Il principale punto critico è il rapporto con i requisiti della sorveglianza umana stabiliti dalla normativa europea e dalla giurisprudenza amministrativa italiana.

[30] Per un'affermazione particolarmente netta di questa prospettiva, comune a tutti gli interventi raccolti, si veda F. Meneghetti, *La piattaforma di Urban Intelligence della città di Venezia*, in *Rapporto 4/2021*, cit.

Quanto alla normativa dell'Unione, la già richiamata proposta di regolamento della Commissione europea in materia di IA, in linea con l'approccio del regolamento generale sulla protezione dei dati (General Data Protection Regulation – GDPR)[31], che riprende e sviluppa notevolmente, ha individuato nella «sorveglianza umana» uno dei requisiti obbligatori applicabili ai sistemi di IA ad alto rischio. Ad avviso della Commissione, in particolare, la sorveglianza umana è necessaria per prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali. L'esigenza di garantire il controllo umano, inoltre, dovrebbe essere tenuta in considerazione sin dal momento della progettazione dei sistemi di IA ad alto rischio[32]. Ancora, può dirsi soddisfatta quando sono garantire alcune specifiche condizioni, ovvero là dove la persona alla quale sono affidati i compiti di sorveglianza comprende appieno le capacità e i limiti del sistema di IA, è in grado di monitorarne debitamente il funzionamento e di individuarne tempestivamente le disfunzioni, è consapevole della possibilità di una «distorsione dell'automazione», è in grado di interpretare correttamente l'output del sistema di IA, di decidere di non usarlo o di ignorarne, annullarne o ribaltarne l'output, così come di intervenire sul funzionamento del sistema anche interrompendolo mediante un pulsante di arresto[33].

Nella stessa direzione vanno le pronunce nelle quali il Consiglio di Stato ha affrontato il problema della compatibilità costituzionale dell'uso di IA da parte dell'amministrazione[34]. In quella giurisprudenza, infatti, il giudice amministrativo ha affermato che una delle condizioni perché si possa ricorrere alle macchine per lo svolgimento di compiti tradizionalmente svolti da persone è il mantenimento della sorveglianza umana. L'impiego di sistemi di IA, più precisamente, richiede che il funzionario controlli la macchina, fino al punto di correggerne l'eventuale funzionamento scorretto, intervenire sugli output e premere lo stop botton, se questa è l'unica misura in grado di evitare gli effetti pregiudizievoli conseguenti al ricorso all'IA. In assenza di questo tipo di sorveglianza, una decisione

[31] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, in GU L 119 2016. Si veda, in particolare, l'art. 22, che stabilisce il cosiddetto principio di non esclusività, in base al quale l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato. Il principio, peraltro, non si applica ai casi indicati nello stesso art. 22, co. 2.

[32] *Ibidem*, considerando 48 e art. 14, co. 1.

[33] *Ibidem*, art. 14, co. 4.

[34] Si vedano, ad esempio, Consiglio di Stato, Sez. VI, sentenza 8 aprile 2019, n. 2270; Consiglio di Stato, sez. VI, n.
881/2020; Consiglio di Stato, sez. VI, n. 2270/2019; Consiglio di Stato, sez. VI, n. 8472/2019.

amministrativa automatizzata che incida sulla sfera giuridica soggettiva dei destinatari è da considerarsi incostituzionale[35].

Confrontato con questa cornice giuridica, il quadro che emerge dall'indagine svolta è preoccupante. Se in alcuni casi è stata riconosciuta l'esigenza di una sorveglianza umana e anche di una sua formalizzazione in disposizioni giuridiche, le amministrazioni paiono ben lontane, al momento, dal soddisfare le condizioni minime richieste dal diritto europeo e dalla giurisprudenza nazionale almeno per i sistemi di IA più capaci di produrre conseguenze per i cittadini e gli operatori economici che si confrontano con l'amministrazione. Non è chiaro si vi sia un funzionario al quale sia imputabile l'algoritmo, quali competenze tecniche eventualmente abbia, come sia costruita giuridicamente la sua responsabilità. Sembra mancare, inoltre, una politica interna volta ad attrezzare l'amministrazione alla sorveglianza umana, a partire dalla formazione del personale, che deve essere in grado di comprendere la logica e i modi di funzionamento dei sistemi di IA. A ciò si accompagna, in alcuni casi, il rafforzamento crescente e rapidissimo della capacità di svolgere una sorveglianza umana da parte dei privati che siano stati eventualmente coinvolti nello sviluppo dei sistemi di IA: l'esempio principale è quello delle imprese che sviluppano e gestiscono le piattaforme per le smart cities, le cui capacità di gestione dei sistemi permettono loro di sostituirsi integralmente al controllo delle amministrazioni.

Naturalmente, si deve considerare che si tratta di un processo in corso e che la *juridification* dei modi di impiego dei sistemi di IA da parte delle amministrazioni è una questione inevitabilmente destinata ad acquisire una importanza crescente nel prossimo futuro. La stessa realtà amministrativa delle autorità considerate, del resto, potrebbe offrire un punto di partenza meno rudimentale di quanto non sia emerso in questa ricerca. E il giudizio potrebbe rivelarsi meno severo se si distinguesse precisamente tra i diversi tipi di IA utilizzati, dato che il funzionamento degli *algoritmi rule based* pone problemi del tutto diversi da quello degli algoritmi machine learning. Resta il problema, in ogni caso, di un'amministrazione che spinge verso l'innovazione senza farsi carico del problema della sorveglianza umana, diversamente da quanto avviene in altri Stati membri dell'Unione europea, a partire dall'esperienza già richiamata dell'ordinamento francese.

[35] Per una discussione di questa giurisprudenza, si vedano Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, cit.; e Marchetti, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, cit.

5. *Conclusioni*

In questo scritto, si è tentato di fotografare i modi nei quali alcune amministrazioni, riconducibili a diversi modelli del complessivo sistema amministrativo italiano, stanno sperimentando, nell'esercizio delle proprie attività, sistemi di IA. L'attenzione è stata portata alle pratiche reali di queste amministrazioni, in modo da comprenderne le esperienze concrete.

Le conclusioni principali sono tre. Anzitutto, le pratiche utilizzate per dotarsi di un sistema di IA variano da caso a caso, ma sono ordinabili in base a un'alternativa chiara, quella tra la produzione in house delle applicazioni di IA e il ricorso al mercato attraverso la gara pubblica. Le due opzioni rispondono a logiche diverse e riflettono le differenti capacità tecniche delle amministrazioni. Quale che sia il metodo prescelto, in ogni caso, si pone il problema delle competenze tecniche interne all'amministrazione, senza le quali quest'ultima è incapace di governare la scelta tecnologica e gestirne le implicazioni. Se tale *expertise* sia effettivamente presente in tutte le amministrazioni considerate, è una questione alla quale non è possibile dare una risposta univoca.

In secondo luogo, si registra una notevole varietà anche nelle tecnologie di IA effettivamente utilizzate dalle amministrazioni e nelle finalità del loro impiego, con diversi gradi di avanzamento tra un'autorità e l'altra. Il quadro che emerge, in ogni caso, è certamente ricco e segnala l'intenzione di quasi tutte le amministrazioni esaminate di sfruttare i sistemi di IA come strumento di innovazione delle proprie attività istituzionali. Allo stesso tempo, l'analisi mostra l'importanza delle condizioni che facilitano l'uso di strumenti di IA, dagli incentivi, non solo economici, alla chiarezza dei bandi di gara, e permette di mettere a fuoco le questioni più urgenti e al momento irrisolte, a partire da quelle relative alla trasparenza e all'*accountability* dell'amministrazione nei confronti di cittadini e imprese.

Da ultimo, se le pratiche delle amministrazioni sono relativamente varie e articolate rispetto ai modi di acquisizione e di utilizzo dei sistemi di IA, risultano decisamente più omogenee quando si considerano i meccanismi di controllo. Le amministrazioni considerate, infatti, sperimentano strumenti di IA al di fuori di un quadro normativo che definisca in modo chiaro il loro impiego all'interno del procedimento e il ruolo del decisore umano. Si tratta di una tendenza che si spiega a partire dalla logica essenzialmente funzionale che muove le amministrazioni nella ricerca di tecnologie di IA, ma che risulta assai problematica rispetto al quadro regolatorio in corso di sviluppo, anche per effetto del diritto dell'Unione europea.

Queste conclusioni, naturalmente, rappresentano a loro volta dei punti di avvio di un'ulteriore fase della ricerca. Restituiscono l'immagine di un quadro in veloce divenire, caratterizzato, per un verso, da una forza attrattiva verso il digitale e da una varietà di soluzioni, per altro verso, dalla difficoltà di affrontare direttamente i problemi che l'uso di sistemi di IA inevitabilmente porta con sé. Se le dinamiche qui ricostruite possano essere generalizzate ed estese ad altre componenti del sistema amministrativo italiano, e se quest'ultimo sia in grado di farsi carico delle questioni aperte, sono domande alle quali spetta alle prossime fasi della ricerca tentare di rispondere.

Nicoletta Rangone

Artificial intelligence and human intelligence in support of good administration

Intelligenza artificiale e intelligenza umana a supporto di una buona amministrazione

Sommario: 1. Il difficile rapporto tra intelligenza artificiale e intelligenza umana - 2. Fiducia nelle istituzioni e fiducia nell'intelligenza artificiale - 3. Considerazioni conclusive

Abstract: The contribution starts with analysing the use of algorithms by public administrations as support for the goal of good administration. However, in dealing with risks, easy formal solutions that undermine an already shaky trust in the administration and AI should be avoided. In addition, the public official's trust in the administration using AI is also relevant to nurturing trust in AI. While procedural guarantees that must also be ensured in an administrative procedure, the author suggests a comparison of the advantages and disadvantages in deciding whether to rely on AI or AI-supported human intelligence associated with human decision-making alone. In addition, such decisions will then have to be reviewed in the light of monitoring, reports and appeals lodged against the outcomes of partially or fully automated decisions.

Abstract: Il contributo parte dall'analisi dell'utilizzo degli algoritmi da parte delle pubbliche amministrazioni come supporto all'obiettivo di una buona amministrazione e ne evidenzia i rischi. Per fronteggiare il secondo occorre non limitarsi a soluzioni formali che finirebbero per minare la fiducia dei cittadini e delle imprese nelle amministrazioni pubbliche che usano IA, facendo anche attenzione ad alimentare la fiducia del funzionario pubblico nell'IA. Se vanno assicurate garanzie procedurali nel processo decisionale, a monte di questo la scelta di fare ricordo all'IA dovrebbe basarsi su una comparazione tra i relativi vantaggi e svantaggi rispetto all'intelligenza umana e, a valle, le decisioni parzialmente o completamente automatizzate adottate dovrebbero essere oggetto di un attento monitoraggio e se possibile confronto rispetto a interventi tradizionali.

* Questo articolo è stato originariamente pubblicato in *Munera*, 2, 2022, pp. 101-111. Le considerazioni svolte sono frutto della ricerca avviata nell'ambito del progetto PRIN 2017 "governance of/through Big Data: challenges for european law". L'autore è titolare della cattedra Jean Monnet EU Approach to Better Regulation; il sostegno della Commissione europea alla produzione di questa pubblicazione non costituisce un'approvazione del contenuto, che riflette esclusivamente il punto di vista dell'autore, e la Commissione non può essere ritenuta responsabile per l'uso che può essere fatto delle informazioni ivi contenute.

1. *Il difficile rapporto tra intelligenza artificiale e intelligenza umana*

Gli algoritmi supportano da tempo varie pubbliche amministrazioni italiane, si pensi all'attribuzione di sedi lavorative agli insegnati[1], al calcolo delle tariffe di energia e rifiuti dal parte dell'autorità indipendente di settore[2], alle sovvenzioni allo spettacolo[3], all'e-procurement con aggiudicazione al prezzo più basso[4]. Oltre a questi algoritmi estremamente semplici, basati su una logica lineare, emergono applicazioni dell'intelligenza artificiale, a supporto all'organizzazione, della prestazione di servizi e del processo decisionale di varie amministrazioni, centrali e locali[5].

Queste applicazioni dell'intelligenza artificiale nel settore pubblico (oltre che privato) hanno messo in luce numerosi vantaggi. Si pensi alla possibilità di processare una enorme mole di dati e dunque di arrivare a decisioni particolarmente informate e precise, così come ad una programmazione dei controlli mirata alle attività o imprese a più alto rischio (che si tratti di controlli sulla manipolazione dei mercati svolti dalle autorità di vigilanza dei mercati finanziari o della rilevazione di anomalie nell'aggiudicazione di appalti pubblici). Questo si risolve non solo in un risparmio di tempo e risorse finanziarie, ma anche nel raggiungimento di risultati irrealizzabili dalle pubbliche amministrazioni che si affidino alla sola intelligenza umana.

[1] Queste decisioni sulla mobilità hanno dato origine ad un importante contenzioso (Tar Lazio sez. III*bis*, n. 3742/2017 e n. 3769/2017 e Consiglio di Stato VI n. 8472/2019, n. 8473 e n. 8474 dello stesso anno, n. 881/2020). Anche con l'apertura dell'anno scolastico 2021/2022, numerosi errori dell'algoritmo nell'attribuzione delle cattedre vengono riportati dai quotidiani (si veda ad esempio, La Stampa, 10 settembre 2021).

[2] G. Avanzini, *Decisioni amministrative e algoritmi informatici. Predeterminazione, analisi predittiva e nuove forme di intellegibilità*, Napoli, Editoriale Scientifica, 2019, p. 53-57.

[3] D.M. 27 luglio 2017, *Criteri e modalità per l'erogazione, l'anticipazione e la liquidazione dei contributi allo spettacolo dal vivo, a valere sul Fondo unico per lo spettacolo di cui alla legge 30 aprile 1985, n. 163*.

[4] Consiglio di Stato sez. consultiva per gli atti normativi, adunanza 17 novembre 2020, n. 1322/2020.

[5] Per una analisi si rinvia ai rapporti curati da E. Chiti, B. Marchetti, N. Rangone, *Il progetto: L'uso dell'intelligenza artificiale nel sistema amministrativo italiano*, in *Biolaw Journal*, 4, 2021, pp. 209-210; Rapporto 1/2021 – *L'impiego dell'intelligenza artificiale nell'attività di CONSOB, AGCOM e ARERA,* p. 211-227, Rapporto 2/2021 – *L'impiego dell'intelligenza artificiale nell'attività di Banca d'Italia*, pp. 229-244. Inoltre, E. Chiti, B. Marchetti, N. Rangone, *Smart cities e Amministrazioni centrali di fronte all'intelligenza artificiale: esperienze a confronto*, in *Biolaw Journal*, 1/2022, pp. 251-252; Rapporto 3/2022 - *Smart cities e Amministrazioni centrali di fronte all'intelligenza artificiale: esperienze a confronto*, pp. 253-259, Rapporto 4/2022, *Intelligenza artificiale e amministrazioni centrali*, pp. 261-274.

A ciò si aggiungano i temi centrali dell'eliminazione dell'errore umano e della riduzione delle occasioni di corruzione. Dunque si può dire che le nuove tecnologie supportino il diritto a una buona amministrazione.

Una volta uscito dall'oscurità, l'uso dell'intelligenza nel settore pubblico ha dato anche origine a numerose critiche, legate agli errori e dunque ai rischi per gli interessi (ad esempio, a ricevere un sussidio se ne ricorrono i presupposti) e diritti fondamentali legati alla dignità umana, alla libertà, all'uguaglianza, alla riservatezza. Ed invero, i menzionati esiti positivi non sono scontati, perché come ormai noto, l'errore umano e le discriminazioni possono essere perpetrate (se non amplificate) dall'intelligenza artificiale proprio in quanto progettata da esseri umani (quello che viene definito *bias in bias out*, o *garbage in garbage out*)[6].

Al contempo, i sistemi di intelligenza artificiale si alimenta di dati che possono essere il frutto di ambiguità e incoerenze connesse all'intervento umano. Si pensi ad un algoritmo che sia stato addestrato con i dati risultanti da ispezioni in cui funzionari diversi siano giunti a conclusioni divergenti su una stessa attività[7]. È quello che D. Kahaneman, O. Sibony e C.R. Sunstein chiamano "Noise. A flaw in human judgment"[8]. L'imprecisione dei dati è anche connessa al diffuso ricorso a dati già esistenti, generati e condivisi continuamente *online* a livello globale (l'uso di "*data from the wild*")[9], che implica un'immediata disponibilità di numerosissimi dati gratuiti, ma al contempo non verificati né dai proprietari, né da soggetti terzi.

Cosicchè, una progettazione non adeguata e dati non verificati minano l'imparzialità che si è portati a ritenere insita nell'intelligenza artificiale. E questo risulta quasi imperdonabile agli occhi delle persone, che perdono fiducia nella "macchina" molto più rapidamente di quando sia invece un

[6] S. Mayson, *Bias in, Bias Out*, in *Yale Law Journal*, vol. 128, 2019, pp. 2219-2300: p. 2224 nota 23; A. Caliskan, J.J. Bryson, A. Narayanan, *Semantics derived automatically from language corpora contain human-like biases*, in *Science*, n. 356, 2017, pp. 183-186; C. O'Neil, *Weapons of math destruction. How Big Data increses inequality and threatens democracy*, London, Penguin, 2016, p. 21; C.R. Sunstein, *Algorithms, correcting biases*, in *Social Research: An Int Quart*, vol. 86, n. 2, 2019, p. 499-511.

[7] L. Lorenz, J. van Erp, A. Meijer, *Machine-learning algorithms in regulatory practice: Nine organisational challenges for regulatory agencies*, in *Technology and Regulation*, 2022, pp. 1-11: p. 7.

[8] Little Brown Spark, 2021. La soluzione proposta dagli autori è proprio l'uso dell'intelligenza artificiale.

[9] N. Cristianini, *Shortcuts to Artificial Intelligence,* in M. Pelillo, T. Scantamburlo, a cura di, *Machines We Trust: Perspectives on Dependable AI*, The MIT Press, Cambridge, Massachusetts, London, England, pp. 15 e 16.

essere umano a sbagliare[10]. Ciò appare a ben vedere paradossale, perché se gli errori e le discriminazioni frutto dell'uso dell'intelligenza artificiale possono in ampia misura essere limitati, così non può dirsi di quelli delle persone fisiche che, proprio in quanto umane, possono operare valutazioni contraddittorie e sono influenzate da *bias* che portare con sé errori di giudizio[11]. A quest'ultimo riguardo, quanto all'interazione uomo-macchina, possono entrare in gioco distorsioni cognitive in grado di minare la garanzie dell'intervento umano, richiesto sia dal regolamento sulla privacy[12], sia dalla proposta di regolamento sull'intelligenza artificiale per i sistemi ad alto rischio[13]. Si pensi al completo affidamento nei confronti della presunta oggettività delle indicazioni che vengono dall'intelligenza artificiale (distorsione dell'automazione)[14] o, al contrario, l'eccessiva diffidenza (avversione all'algoritmo)[15], o ancora l'eccessiva fiducia nel ragionamento umano (*illusion of validity*)[16] che spesso caratterizza gli esperti[17].

Tuttavia, "demonstrating that machine learning can outperform humans in the completion of some tasks does not mean they will outperform humans in every task" [18]. L'uso della discrezionalità è in molte ipotesi fondamentale per risolvere casi concreti alla luce di regole generali e più vi è bisogno di valutazione umana, così che meno spazio andrebbe lasciato all'automazione. Al contempo vi è da dire che l'intelligenza artificiale "è sopravvalutata nella

[10] J. Jaurnig et al., *People Prefer Moral Discretion to Procedurally Fair Algorithms: Algorithm Aversion Beyond Intransparency,* in *Philosophy & Technology*, vol. 35, n. 1, 2021, pp. 1-25; B.J. Dietvorst, J.P. Simmons, C. Massey, *Algorithm aversion: People erroneously avoid algorithms after seeing them err*, in *Journal of Experimental Psychology: General*, vol. 144, n. 1, 2015, pp. 114-126.

[11] C. Coglianese, *Algorithm vs. Alvorithm*, 72 *Duke L. J*. 1288, 2022.

[12] Art. 22, regolamento europeo n. 679 del 27 aprile 2016.

[13] Art. 14, comma 3

[14] L.J. Skitka et al., *Automation Bias and Errors: Are Crews Better Than Individuals?*, in *The International Journal of Aviation Psychology*, 2000, vol. 10, n. 1, pp. 85-97; K.L. Mosier et al., *Automation Bias: Decision Making and Performance in High-Tech Cockpits*, in *International J.Aviation Psychology*, 1997, n. 8, vol. 1, p. 47-63; K. Goddard, A. Roudsari, J.C. Wyatt, *Automation bias: a systematic review of frequency, effect mediators, and mitigators*, in *J. Am. Med. Inform. Assoc.*, 2012, vol. 19, pp. 121-127.

[15] B.J. Dietvorst, J.P. Simmons, C. Massey, *Algorithm aversion: People erroneously avoid algorithms after seeing them err*, cit.

[16] A. Tversky, D. Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, in *Science* , vol. 185, 1124, 1974, p. 1126; D. Kahneman, G. Klein, *Conditions for Intuitive Expertise. A Failure to Disagree*, in *American Psychologist*, 2009, pp. 515-526: p. 517.

[17] L. Lorenz, J. van Erp, A. Meijer, *Machine-learning algorithms in regulatory practice. Nine organisational challenges for regulatory agencies*, cit., p. 6.

[18] C. Coglianese, *Algorithm vs. Alvorithm*, cit., p. 1312.

sua capacità di prevedere fenomeni instabili": in condizioni di incertezza, se il futuro non è come il passato, i big data possono risultare ingannevoli[19].

Il quesito che si pone allo studioso non è, dunque, tanto l'alternativa tra intelligenza umana e intelligenza artificiale, quanto in quali ipotesi è preferibile ricorrere all'intelligenza artificiale in luogo dell'intelligenza umana e in quali ipotesi si dovrebbe supportare l'intelligenza umana attraverso l'intelligenza artificiale.

A ben vedere questa è la direzione intrapresa dalle istituzioni europee che hanno proposto un regolamento sull'uso dell'intelligenza artificiale nel settore pubblico e privato[20] che costituirà (una volta approvato definitivamente) un'importante guida per decidere se e come fare uso dell'intelligenza artificiale. Il regolamento, oltre a prevedere attività che non possono essere affidate all'intelligenza artificiale (come la classificazione dell'affidabilità delle persone sulla base del loro comportamento o di caratteristiche personali e l'identificazione biometrica remota in tempo reale salvo quando necessaria alla prevenzione o il contrasto di crimini)[21], prevede numerosi presidi quanto alle altre. In particolare, le attività qualificabili come ad alto rischio (come la polizia predittiva "person based", la valutazione di candidati ad un posto di lavoro o studenti per l'ammissione a istituzioni di formazione, l'accesso a servizi di assistenza pubblica)[22] sono sottoposte a una serie di cautele, come l'istituzione, la documentazione e la manutenzione di un sistema di gestione dei rischi[23], la predisposizione di pratiche di *governance* e gestione dei dati utilizzati per l'addestramento dei modelli[24], la conservazione della documentazione tecnica[25], la conservazione

[19] *Intelligenza artificiale poco intelligente senza psicologia*, conversazione di R. Viale con G. Gingerezer, in *Corriere della Sera*, 19 dicembre 2021, pp. 26-27. G. Gingerezer (*How to stay Smart in a Smart World. Why Human Intelligence Still Beats Algorithms*, Penguin, 2021) elabora in particolare una proposta di una intelligenza artificiale "psicologica" che analizza le intuizioni degli esperti e trasforma le loro euristiche in algoritmi.

[20] Proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021 che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione europea, COM(2021) 206 final.

[21] Art. 5, proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021, cit.

[22] Allegato III, proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021, cit.

[23] Art. 9, proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021, cit.

[24] Art. 10, proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021, cit.

[25] Art. 11, proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021, cit.

delle registrazioni del funzionamento tale da garantire la tracciabilità del sistema durante tutto il suo ciclo di vita[26], la trasparenza per gli utenti[27], la sorveglianza umana volta prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali[28], ed ancora una progettazione e sviluppo tali da consentire accuratezza, robustezza e cibersicurezza durante il ciclo di vita[29]. I rischi (non considerati alti) connessi ad altri usi dell'intelligenza artificiale (come i controlli sull'igiene degli alimenti), vanno comunque minimizzati attraverso previsioni che ne rendono trasparente l'uso e il funzionamento: "i fornitori garantiscono che i sistemi di IA destinati a interagire con le persone fisiche siano progettati e sviluppati in modo tale che le persone fisiche siano informate del fatto di stare interagendo con un sistema di IA, a meno che ciò non risulti evidente dalle circostanze e dal contesto di utilizzo" [30].

Si tratta di un inquadramento normativo assente in ordinamenti che pure fa uso da tempo dell'intelligenza artificiale come quello nordamericano[31], teso a definire regole che consentano lo sviluppo di tutte quelle tecnologie di supporto ad una buon amministrazione (quanto al settore pubblico), ma che siano anche affidabili[32] e meritevoli della fiducia di cittadini e imprese[33].

[26] Art. 12, proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021, cit.
[27] Art. 13, proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021, cit.
[28] Art. 14, proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021, cit.
[29] Art. 15, proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021, cit.
[30] Art. 52, proposta di regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021, cit.
[31] Per una mappatura degli usi da parte delle agenzie federali, si veda D.F. Engstrom, D.E. Ho, C.M. Sharkey, M.-F. Cuellar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, Report submitted to the Administrative Conference of the United States. Una normativa sugli usi è auspicata in dottrina (sul punto si veda A.A. Gavoor, *The Impending Judicial Regulation Of Artificial Intelligence In The Administrative State*, 97 *Notre Dame L. Rev*. 180-188 (2022). Per un confronto tra i due approcci, B. Marchetti, L. Parona, *La regolazione dell'intelligenza artificiale: Stati Uniti e Unione europea alla ricerca di un possibile equilibrio*, in *DPCE online*, vol. 51, n. 1, 2022, pp. 237-252.
[32] Gruppo di esperti ad alto livello sull'intelligenza artificiale, *Orientamenti etici per un'IA affidabile*, 2019.
[33] Commissione europea, *Libro sull'intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia*, COM(2020) 65 final. Commissione europea, *Creare fiducia nell'intelligenza artificiale antropocentrica*, COM(2019) 168 final.

2. *Fiducia nelle istituzioni e fiducia nell'intelligenza artificiale*

La fiducia di cittadini, imprese e funzionari pubblici nel diritto e nelle istituzioni pubbliche è un fattore intangibile, ma cruciale per l'effettività del diritto, vale a dire perché questo produca effetti e produca gli effetti voluti[34]. Per rinsaldare la fiducia nelle istituzioni entrano in gioco diversi fattori, tra i quali la chiarezza e l'accessibilità del quadro normativo. Di converso, un quadro giuridico sovrabbondante, poco chiare, troppo restrittivo e che impone oneri burocratici non giustificati e norme ingiustificatamente severe, alimenta l'inosservanza, la creative *compliance*, la corruzione e, in ultima analisi la sfiducia[35]. Al contempo risulta fondamentale l'esperienza del rapporto di cittadini e imprese con le pubbliche amministrazioni, che dovrebbero supportare l'adempimento (ad esempio attraverso la precompilazione dei moduli con dati già a disposizione, la messa a disposizione di liste di adempimenti, la disponibilità a spiegazioni e chiarimenti), prevenire gli errori e aiutare a superarli, essere cooperative e non aggressive nei controlli[36].

Quale ruolo può svolgere l'intelligenza artificiale per rinsaldare la fiducia? Le nuove tecnologie possono supportare l'approccio collaborativo (migliorando i servizi all'utenza e le prestazioni amministrative, favorendo controlli più mirati e dunque diminuendo i costi di controllo non necessari), ma possono anche minare la fiducia nelle pubbliche amministrazioni. Ad esempio, una chatbot "intelligente" che risponde alle parole di uso comune e guida anche l'utente inesperto verso il servizio voluto può essere di aiuto al cittadino, ma andrebbe sempre assicurata la possibilità di parlare con una persona fisica (che sia anche disponibile e competente) al cittadino che necessiti di una risposta personalizzata o a colui che per i più diversi motivi

[34] M. D'Alberti, introduzione a G. Corso, M. de Benedetto, N. Rangone, *Diritto amministrativo effettivo. Una introduzione*, Il Mulino, 2022, pp. 15-17.

[35] Una serie di studi ha individuato una delle determinanti della corruzione amministrativa proprio nel quadro normativo, quando questo risulti non chiaro, sovrabbondante, all'origine di rilevanti costi di adempimento (per tutti, M. De Benedetto, *Corruption from a regulatory perspective*, Hart, 2021). Si veda anche C.A. Dunlop, C.M., Radaelli, *Regulation and Corruption: Claims, Evidence and Explanations*, in A. Massey (a cura di), *A Research Agenda for Public Administration*, E. Elgar publishing, 2019. In un circolo vizioso la corruzione (anche solo percepita) non fa che minare ulteriormente la fiducia nelle istituzioni (S.D. Morris, J.L. Klesner, *Corruption and Trust: Theoretical Considerations and Evidence From Mexico*, in *Comparative Political Studies*, vol. 43, n. 10, 2000, pp. 1258-1285).

[36] Per una analisi di quanto si intende per amministrazione collaborativa e di supporto all'adempimento sia consentito rinviare a Corso, de Benedetto, Rangone, *Diritto amministrativo effettivo. Una introduzione*, cit., p. 204 ss.

non si senta a suo agio con le tecnologie. In altre parole, per non minare la fiducia, "an automated State" dovrebbe essere anche "an empathic one"[37].

D'altro canto, la stessa intelligenza artificiale deve essere affidabile, ingenerare fiducia e sicurezza nei cittadini e nelle imprese, direzione in cui - come evidenziato nel paragrafo precedente - muove l'approccio europeo alla disciplina degli usi. Della rapida erosione della fiducia nei confronti dell'intelligenza artificiale che "sbaglia", perché ad esempio chiede erroneamente la restituzione di sussidi già erogati o non eroga sussidi agli aventi diritto, si è già detto. Come alimentare la fiducia nell'intelligenza artificiale o meglio nelle amministrazioni che usano l'intelligenza artificiale?

Non vi è dubbio che la trasparenza degli usi sia fondamentale, indipendentemente dal fatto che questo si realizzi o meno nell'ambito di un procedimento amministrativo, come nel caso della trattazione attraverso l'intelligenza artificiale di esposti indirizzati a Banca d'Italia. In base al regolamento pubblicato da questa istituzione nel marzo del 2022, seppure la segnalazione non porti all'avvio un procedimento amministrativo, colui che presenta l'esposto e gli intermediari potenzialmente interessati devono sapere che le informazioni acquisite vengono trattate attraverso *machine learning*[38]. Devono poi essere predisposte tecniche per garantire il rispetto della riservatezza dei dati personali, la sorveglianza umana del monitoraggio e dell'aggiornamento delle tecniche di *machine learning* (anche per assicurarne la spiegabilità), l'assenza di uso per decisioni automatizzate, profilazione o predizione di comportamenti[39]. Anche questi presidi devono essere resi trasparenti.

Trasparenza, spiegabilità, non discriminazione e non esclusività delle decisioni basata sull'intelligenza artificiale sono le garanzie procedurali che vanno assicurate anche nell'ambito di un procedimento amministrativo[40], come nel caso della comminazione di una sanzione per violazione della normativa fiscale.

A fronte di queste imprescindibili garanzie rafforzate riconosciute da quello che comincia ad essere definito il nuovo diritto amministrativo dell'amministrazione digitale[41], occorre però considerare anche la reale

[37] Coglianese, *Administrative law in the automated State*, *American Academy of Arts and Sciences*, cit., p. 104; S. Ranchordas, *Empathy in the Digital Administrative State,* 71 *Duke L. J.* 1342-1389, 2022.

[38] Regolamento sul trattamento dei dati personali nella gestione degli esposti del 2022, pubblicato nel sito della Banca d'Italia e in gazzetta ufficiale.

[39] Questi Gli Aspetti Particolarmente Apprezzati Dal Garante Della Privacy, Nel Parere Del 24 Febbraio 2022 Alla Banca D'italia Sullo Schema Di Regolamento Concernente Il Trattamento Dei Dati Personali Effettuato Nell'ambito Della Gestione Degli Esposti.

[40] Così varie pronunce del Consiglio di Stato a partire dalla sentenza n. 8472 del 2019.

[41] L. Torchia, *Lo Stato digitale e il diritto amministrativo*, in *Liber amicorum per Marco*

accessibilità e comprensibilità di queste informazioni. Non è detto che la pubblicazione di una disciplina dell'uso dell'intelligenza artificiale (e quello citato di Banca d'Italia costituisce la prima esperienza in Italia) nel sito istituzionale e in gazzetta ufficiale sia realmente accessibile a tutti i soggetti potenzialmente interessati, pur essendo questo approccio ineccepibile da un punto di vista giuridico (come evidenziato anche dal Garante della privacy).

Nel perseguimento di una trasparenza mirata rinsaldare la fiducia e assicurare una reale accessibilità e comprensibilità (tenendo conto dei limiti cognitivi delle persone) in alcuni contesti si comincia a ragionare di una resa di informazioni su più livelli. In questa direzione muove ad esempio l'esperienza della città di Helsinky, che mette a disposizione del pubblico un registro contenente tutti gli usi dell'intelligenza artificiale da parte dell'amministrazione municipale, le cui applicazioni sono descritte in forma molto semplificata anche attraverso il supporto di immagini, con possibilità di approfondimento tecnico per chi fosse interessato[42]. Un'interessante esemplificazione grafica su come rendere accessibili le informazioni relative agli usi dell'intelligenza artificiale, le ragioni dell'utilizzo, i dati utilizzati e i relativi rischi, con possibilità di approfondimento, è offerta dal rapporto del Regno Unito "Complete transparency, complete simplicity"[43].

Come emerso dal paragrafo 1, per alimentare la fiducia nelle amministrazioni che usano l'intelligenza artificiale rileva anche la fiducia del funzionario pubblico nell'intelligenza artificiale. Se dei *bias* cognitivi che possono gravare sul funzionario che interagisce con le indicazioni dell'intelligenza artificiale si è già detto, la fiducia del funzionario nei confronti delle indicazioni che provengono dall'uso di intelligenza artificiale è anch'essa un fattore imprescindibile. La fiducia attiene non solo alla correttezza delle indicazioni provenienti dall'intelligenza artificiale, ma anche alla *fairness* della procedura alla base di tali indicazioni, che può riguardare i dati utilizzati e modalità o trasparenza della loro elaborazione. Se la correttezza percepita del sistema di intelligenza artificiale collide con l'*individual fairness*[44], le indicazioni provenienti dall'intelligenza artificiale rischiano di essere ignorate o adattate alla luce dei valori degli "screen-level bureaucrats"[45].

D'Alberti, Giappichelli, 2022, p. 477.

[42] <https://ai.hel.fi/en/ai-register/>

[43] *BritainThinks: Complete transparency, complete simplicity*, 2021, pp. 1-51: p. 10-11.

[44] Centre for Data Ethics and Innovation, *Review into bias in algorithmic decision-making*, 2020, p. 4-68.

[45] R. Binns, *Human judgment in algorithm loops*, in *Regulation and Governance*, 2020, p. 10; D.S. Rubenstein, *Acquiring Ethical AI*, 23 *Florida L. Rev.* 747-819, 2021.

3. *Considerazioni conclusive*

Le amministrazioni italiane sperimentano e utilizzano l'intelligenza artificiale per esigenze conoscitive, a supporto di servizi al pubblico, nei processo decisionale per l'adozione di *policies*, regole, decisioni amministrative, così come nei controlli.

Numerosi sono i vantaggi potenzialmente ricavabili, sia in termini di risparmi di risorse economiche ed umane, che di interventi più mirati anche in termini prospettici, tanto che l'intelligenza artificiale può supportare l'obiettivo di una buona amministrazione.

Non mancano però i rischi, in ampia misura gestibili, se solo se ne riconosce l'esistenza e vengono affrontati attraverso previsioni normative che impongano adeguati vincoli agli sviluppatori e garanzie agli utilizzatori. Soprattutto quanto alle seconde, occorre però evitare di adagiarsi su facili soluzioni formali, che potrebbero minare la fiducia nell'amministrazione e nella stessa intelligenza artificiale. Nel caso della trasparenza, ciò significa assicurare una reale comprensibilità, laddove il ruolo riconosciuto all'intervento umano dovrebbe tener conto dei *bias* che potrebbero inficiare il pieno svolgimento[46].

Inoltre, nel rispetto delle previsioni normative che verranno adottate a livello europeo e nazionale, le singole decisioni sul se affidarsi completamente all'intelligenza artificiale oppure all'intelligenza umana supportata dall'intelligenza artificiale andrebbero effettuate comparando vantaggi e svantaggi a queste connessi con lo *status quo* (cioè la sola decisione umana). Una valutazione questa che non potrà essere presa una volta per tutte, ma dovrà essere rivista alla luce del monitoraggio, delle segnalazioni e dei ricorsi eventualmente presentati nei confronti degli esiti di decisioni parzialmente o completamente automatizzate[47].

[46] Per un tentativo di delineare meccanismi tesi ad arginarli, sia consentito rinviare a N. Rangone, *Le pubbliche amministrazioni italiane alla prova dell'intelligenza artificiale*, in *Studi parlamentari e di politica costituzionale*, n. 209, 2021, p. 11-29: p. 22-25.

[47] C. Coglianese e A. Lai, *Digital Versus Human Algorithms,* in *The Regulatory Review*, 25 aprile 2022.

Nicoletta Rangone

Italian public administrations facing artificial intelligence challenge

Le pubbliche amministrazioni italiane alla prova dell'intelligenza artificiale

Sommario: 1. Introduzione – 2. Pubbliche amministrazioni e nuove tecnologie: un ambito dei confini incerti – 3. Regolazioni, provvedimenti amministrativi, controlli: il ruolo dell'intelligenza artificiale- 3.1 Intelligenza artificiale e procedimenti di regolazione – 3.2 Intelligenza artificiale e procedimenti per l'adozione di decisioni amministrative – 3.3 Intelligenza artificiale e attuazione amministrativa – 4. Il complicato rapporto tra intelligenza artificiale e *human bounded rationality* –5. Nuove tecnologie e fiducia – 6. Considerazioni conclusive

Abstract: Italian public administrations are starting to use artificial intelligence in rulemaking, adjudication and enforcement. However, the exact perimeter and the scope of these experiences remain elusive, with a potential negative impact on the guarantees of the interested parties. The contribution intends to highlights the specificities of such moments of the decision-making process as regards to the problems linked to the use of artificial intelligence. Moreover, it addresses three determining factors for the success of innovation in public administrations: the transparency, the implementation of the principle of the human intervention (also in view of the limited rationality of individuals), and the trust in new technologies.

Abstract: Le pubbliche amministrazioni italiane cominciano ad utilizzare l'intelligenza artificiale nel processo decisionale pubblico volto all'adozione di regole o di decisioni amministrative puntuali e nell'attuazione amministrativa, ma resta sfuggente la portata di queste esperienze con un potenziale impatto negativo sulle garanzie degli interessati. Nel mettere in evidenza le specificità di tali momenti del processo decisionale quanto ai problemi legati all'uso dell'intelligenza artificiale, il contributo affronta tre fattori determinanti per il successo dell'innovazione nelle pubbliche amministrazioni: la trasparenza, l'attuazione del principio dell'intervento umano (anche in considerazione della razionalità limitata degli individui), la fiducia nelle nuove tecnologie

* Questo articolo è stato originariamente pubblicato in *Studi parlamentari e di politica costituzionale*, 209, 2021, pp.11-29. Il presente contributo è destinato alla pubblicazione nel *Liber amicorum*, dedicato a M. D'Alberti, in corso di pubblicazione. Le considerazioni svolte sono frutto della ricerca avviata nell'ambito del progetto PRIN 2017 "Governance of/Through Big Data: Challenges for European Law".

1. *Introduzione*

Il "Rapporto sui principali problemi della amministrazione dello Stato", trasmesso nel 1979 alle Camere dal ministro per la funzione pubblica Massimo Severo Giannini, evidenzia come "gli elaboratori elettronici, che erano all'inizio apparecchi di semplice registrazione di dati complessi, sono divenuti poi apparecchi di accertamento e verificazione, di calcolo, di partecipazione a fasi procedimentali di istruttoria, e infine di decisione"[1]. Quarant'anni dopo, un percorso simile caratterizza l'uso delle nuove tecnologie nelle pubbliche amministrazioni: da semplici algoritmi di calcolo di tipo lineare, fino alle prime applicazioni dell'intelligenza artificiale. Ieri "i sistemi informatici", oggi gli algoritmi e l'intelligenza artificiale, servono dunque "per amministrare, si proiettano cioè sempre più verso l'esterno"[2]. L'uso dell'intelligenza artificiale nei processi decisionali e la conseguente proiezione all'esterno dei dati così elaborati costituisce un aspetto particolarmente delicato, che presenta numerosi vantaggio, ma anche potenziali limiti per l'esercizio delle garanzie procedurali e processuali degli interessati, e per la stessa effettività dell'intervento pubblico.

Il presente lavoro delinea le principali applicazioni dell'intelligenza artificiale nel processo decisionale pubblico con l'obiettivo di affrontare alcuni fattori che restano problematici: la trasparenza, l'intervento umano e la fiducia nelle nuove tecnologie.

Il ragionamento è strutturato come segue: il paragrafo 2 delinea, in via introduttiva, i vantaggi e limiti di decisioni pubbliche *data-driven* concentrandosi, quanto ai secondi, sulla scarsissima trasparenza che caratterizza, nel momento in cui si scrive le amministrazioni italiane e su quello che questo comporta in termini di garanzie. Il paragrafo 3 si concentra, poi, sulle specificità dell'uso dell'intelligenza artificiale nei tre momenti del processo decisionale pubblico: l'adozione di regole (di seguito anche *rule-making*) e di decisioni amministrative (*adjudication*), l'attuazione amministrativa (*enforcement*). Da questa analisi emerge che tra le principali sfide connesse all'uso dell'intelligenza artificiale nell'*adjudication* e nell'*enforcement* vi è la garanzia dell'intervento umano. A questo riguardo, il paragrafo 4 mette in rilievo che questo presidio, i cui contorni non sono chiaramente definiti né da parte delle istituzioni europee[3], né a livello

[1] Punto 3.7, *Rapporto sui principali problemi della amministrazione dello Stato*, 1979.
[2] *Ibidem*.
[3] Punti 62-65, *Gruppo di esperti ad alto livello sull'intelligenza artificiale, Orientamenti etici per un'IA affidabile*, Commissione europea, 2019; comunicazione della Commissione,

nazionale[4], e che rischia di essere vanificato da possibili atteggiamenti di eccessivo o troppo limitato affidamento nei confronti delle indicazioni dell'intelligenza artificiale riconducibili alla *bounded rationality* degli individui, esiti che andrebbero fronteggiati attraverso adeguate azioni pubbliche. Vi è poi un ulteriore fattore, più sfuggente, che è suscettibile di giocare un ruolo determinante nella diffusione e nell'effettività dei processi decisionali che fanno uso dell'intelligenza artificiale: la fiducia dei decisori pubblici, dei funzionari, dei destinatari delle decisioni, cui è dedicato il paragrafo 5.

2. *Pubbliche amministrazioni e nuove tecnologie: un ambito dei confini incerti*

I vantaggi connessi all'uso dell'intelligenza artificiale "per amministrare"[5] sono potenzialmente di grande rilievo, consentendo non solo di informare i processi decisionali ad analisi empiriche accurate e diffuse con limitatissimo impiego di personale[6], ma anche di migliorare la qualità delle attività di prestazione[7].

Con riferimento alle decisioni pubbliche e alla regolazione, l'uso dell'intelligenza artificiale consente, in particolare, l'analisi di una enorme mole di dati o facilita la realizzazione di complicate valutazioni del rischio, supportando l'effettività delle regole, delle decisioni e del *law enforcement*. L'intelligenza artificiale può anche essere utilizzata in una fase preliminare, rispetto all'avvio (eventuale) di un procedimento, ad esempio nell'ambito

Creare fiducia nell'intelligenza artificiale antropocentrica, COM (2019)168 fin.; art. 14 proposta di regolamento europeo che stabilisce *regole armonizzate sull'intelligenza artificiale e modifica alcuni atti legislativi dell'Unione*, COM(2021) 206 fin.

[4] La "non esclusività della decisione algoritmica" di derivazione europea è menzionata anche dal Consiglio di Stato (VI, 13 dicembre 2019, n. 8472, punto 15.2).

[5] *Rapporto sui principali problemi della amministrazione dello Stato*, cit.

[6] Il risparmio di risorse (umane e monetarie) è, in particolare, la leva che ha portato alla diffusione dell'intelligenza artificiale nei paesi più avanzati in questa direzione, ma che potrebbe non essere altrettanto determinante in Italia stante l'importante *digital divide* che richiede, prioritariamente, importanti investimenti per il suo superamento.

[7] Seppure le attività di prestazione esulino dalla presente analisi, l'uso dell'internet delle cose per raccogliere dati poi rielaborati attraverso l'intelligenza artificiale consente, ad esempio, di offrire informazioni più puntuali (se non personalizzate) agli utenti e di adeguare l'organizzazione dell'offerta (OCSE, *Shaping the future of regulators. The impact of emerging technologies on economic regulators*, 2020). Con riferimento alle nuove tecnologie in sanità, alla telemedicina o alla robotica per le prestazioni sanitarie si veda M. De Angelis, *Alcune questioni giuridiche sulla regolamentazione del progresso tecnologico in sanità*, in *Diritto e questioni pubbliche*, n. 1, 2017, p. 197 ss.

di indagini o monitoraggi di determinati mercati o attività, oppure per analizzare e riorganizzare denunce, reclami e segnalazioni presentate da consumatori o imprese.

Questi impatti positivi non possono però ritenersi scontati e l'uso di nuove tecnologie è anche all'origine di rischi che vanno dall'opacità, all'errore, alla discriminazione come esito di interventi *data-driven*[8].

In primo luogo, la qualità dell'intervento *data-driven* è strettamente correlata alla qualità dei dati su cui si basa, che devono essere affidabili e significativi. L'affidabilità comporta, ad esempio, che i dati non siano raccolti da soggetti in conflitto di interessi[9], non siano basati su assunzioni discutibili[10] o esclusivamente su dati storici senza poter essere confutati dal diretto interessato[11]. Quanto alla significatività, se è pur vero che i big data hanno il vantaggio di fornire dati non limitati a campioni circoscritti di popolazione[12], vi è da chiedersi se i big data stessi possono considerarsi realmente rappresentativi[13] e se a tutti i dati possa essere riconosciuta la

[8] Vantaggi e svantaggi delle decisioni basate sull'intelligenza artificiale, ampiamente discussi in dottrina e dalla giurisprudenza del Consiglio di Stato, sono sintetizzati nell'*explanatory memorandum* alla proposta di regolamento europeo sulle regole armonizzate per l'intelligenza artificiale (COM(2021)206 def.).

[9] Come potrebbe essere il caso delle regioni italiane quanto ai dati trasmessi al governo per la determinazione del grado di rischio legato alla pandemia da COVID-19 (d.P.C.M. del 3 novembre 2020, approvato in Conferenza delle Regioni e Province autonome l'8 ottobre 2020, d.P.C.M. del 30 aprile 2020) ove il conflitto di interessi potenziale è connesso alla rilevazione e trasmissione di dati che poi incidono sulle aperture e dunque sullo sviluppo economico e il consenso. Per inciso, tale determinazione del livello di rischio non sembra riconducibile all'uso di intelligenza artificiale: la comunicazione istituzionale fa infatti riferimento all'uso di algoritmi che sembrerebbero di tipo lineare.

[10] Ad esempio, che chi acquista a rate non ha disponibilità di liquidi, indicatore diffuso tra gli istituti di credito ai fini del rilascio di un mutuo.

[11] Questo sembra essere il caso del "fitto figurativo" che il D.M. 24 dicembre 2012 prevedeva di attribuire a tutti coloro che non dispongano di un'abitazione di proprietà, locazione o uso gratuito nel comune di residenza e che, solo a seguito di un intervento del Garante per la protezione dei dati personali (provvedimento 21 novembre 2013, n. 515) concorre a determinare il maggior reddito accertabile esclusivamente se il contribuente non chiarisce la propria posizione o non si presenta al contraddittorio (Agenzia delle entrate, circolare n. 6/E dell'11 marzo 2014).

[12] N. Musacchio, G. Guaita, A. Ozzello, M.A. Pellegrini, P. Ponzani, R. Zilich, A. De Micheli, *Intelligenza Artificiale e Big Data in ambito medico: prospettive, opportunità, criticità*, in *The Journal of AMD*, 2018, vol. 21-3, p. 211-212.

[13] "If the training data only consist of information from certain population groups, then the tool might work less well for members of missing communities" (C. Coglianese, *A framework for governmental use of machine learning*, report for the Administrate Conference of the United States, 2020, p. 41).

stessa rilevanza (ad esempio, quelli ricavabili dai *social media*[14] rispetto ai dati custoditi da banche dati pubbliche). Vi è poi il rilevante ostacolo alla condivisione delle informazioni su un individuo o un'impresa raccolte da diverse pubbliche amministrazioni per fini differenti (ad esempio controlli previdenziali o tributari) a fronte di un'ambigua interpretazione dei principi di finalità della raccolta e di proporzionalità nel trattamento dei dati fornita dal Garante della privacy[15].

In secondo luogo, "le storture e le imperfezioni che caratterizzano tipicamente i processi cognitivi e le scelte compiute dagli esseri umani"[16], così come i pregiudizi (ad esempio, razziali o di genere) sono perpetrati dai supporti tecnologici (*bias in bias out*)[17]. Di tutto questo va tenuto conto, tanto nell'impostazione dell'algoritmo, quanto nel monitoraggio che dovrebbe essere realizzato per verificare periodicamente l'efficacia delle previsioni.

In terzo luogo, seppure da tempo varie amministrazioni italiane si confrontino con l'uso dell'intelligenza artificiale nei processi decisionali, esse restano accomunate da una diffusa mancanza di trasparenza che attiene diversi profili. Da un lato, manca una mappatura degli usi[18], che non è stata

[14] Il *social data mining* viene sempre più utilizzato dalle assicurazioni per definire tariffe *risk-based*, da banche per valutare l'affidabilità di un utente (T. Berg, V. Burg, A. Gombović, M. Puri, *On the Rise of FinTechs. Credit Scoring using Digital Footprints*, in NBER Working Paper n. 24551, 2018), da fornitori di servizi per praticare discriminazioni di prezzo (F. Di Porto, *La regolazione degli obblighi informativi. Le sfide delle scienze cognitive e dei big data*, Napoli, Editoriale Scientifica, 2017, p. 159).

[15] Parere del Garante Parere su uno schema di decreto concernente il Registro unico dei controlli ispettivi sulle imprese - 25 giugno 2015, n. 378, su cui si veda F. Blanc, M. Benedetti, C. Bertone, *L'informatica e il machine learning al servizio della semplificazione dei controlli sulle imprese: un equilibrio ancora da definire*, in questo numero.

[16] Cons. St., VI, 13 dicembre 2019, n. 8472, punto 7.1.

[17] Questo fenomeno, per il quale "the computer science idiom is garbage in, garbage out" (S. Mayson, *Bias in, Bias Out*, 128 *Yale L. J.* 2224, n. 23, 2019) è stato messo in evidenza da un'imponente letteratura. Ad esempio, A. Caliskan, J.J. Bryson, A. Narayanan, *Semantics derived automatically from language corpora contain human-like biases*, in *Science*, vol 356, 183, 2017, dimostrano che "standard machine learning can acquire stereotyped biases from textual data that reflect everyday human culture" e dunque "caution must be used in incorporating modules constructed via unsupervised machine learning into decision-making systems". Sul punto anche C. O'Neil, *Weapons of math destruction. How Big Data increses inequality and threatens democracy*, London, Penguin, 2016, p. 21.

[18] Vantaggi e limiti di decisioni *data-driven* sono stati ampiamente analizzati da una vasta letteratura che ha esaminato le quasi ventennali esperienze di applicazioni negli Stati Uniti, riorganizzate in modo problematico in un rapporto del 2020 (D.F. Engstrom, D.E. Ho, C.M. Sharkey, M.-F. Cuellar, *Government by Algorithm: Artificial Intelligence*

effettuata da soggetti pubblici (come AGID o la presidenza del Consiglio dei ministri), centri di ricerca o università, mentre la dottrina italiana si è prevalentemente concentrata sulla validità della decisione automatizzata[19], sui parametri costituzionali delle decisioni algoritmiche[20] o su indagini approfondite di applicazioni settoriali[21]. Manca, poi, da parte delle singole amministrazioni, una chiara indicazione delle applicazioni, delle tecnologie utilizzate (se algoritmo lineare di tipo deterministico o forme particolarmente avanzate di intelligenza artificiale) e delle modalità di sviluppo (interno o esterno all'amministrazione), delle ragioni dell'utilizzo, dell'impostazione o meno di un monitoraggio del funzionamento e dei relativi esiti[22]. Cosicché, molte delle informazioni a disposizione del pubblico derivano da notizie di stampa (come nei vari esempi di polizia predittiva), oppure sono riconducibili a comunicazioni istituzionali[23] o documenti che ne richiamano genericamente l'utilizzo[24], che non

in Federal Administrative Agencies, Report submitted to the Administrative Conference of the United States); quanto all'Australia si veda il più risalente rapporto del 2004 (Australian Administrative Review Council, *Automated Assistance in Administrative Decision-Making*, Administrative Review Council). Alcuni lavori non sistematici si riferiscono all'Europa, ad esempio, Algorithm Watch, *Automating Society Report 2020*, che analizza esperienza di sedici paesi.

[19] Sull'invalidità dell'atto amministrativo informatico, A. Masucci, *L'atto amministrativo informatico. Primi lineamenti di una ricostruzione*, Napoli, Jovene, 1993, p. 115 ss.; da ultimo, R. Cavallo Perin, I. Alberti, *Atti e procedimenti amministrativi digitali*, in R. Cavallo Perin, D.-U. Galetta (a cura di), *Diritto dell'amministrazione pubblica digitale*, Torino, Giappichelli, 2020, p. 119 ss.

[20] Cfr., per tutti, A. Simoncini, *Amministrazione digitale algoritmica. Il quadro costituzionale*, in R. Cavallo Perin, D.U. Galetta, *Il diritto dell'amministrazione pubblica digitale*, Torino, Giappichelli, 2020, p. 1 ss.

[21] M. Macchia, *Blockchain e pubblica amministrazione*, in *Federalismi.it*, 18 gennaio 2021.

[22] Il problema della limitata trasparenza resta attuale anche per numerose agenzie Nordamericane (D. Freeman Engstrom, D.F. Ho, *Algorithmic Accountability in the Administrative State*, 37 *Yale J. Regulation* 800, 2020) e non sembra sia stato dato seguito all'"Agency Inventory of AI use Cases" richiesto dall'Executive Order n. 13960/2000 (J.F. Weaver, *Everything Is Not Terminator. The Federal Government and Trustworthy AI*, in *Robotics, Artificial Intelligence* & L., vol. 4, 232, 2021.

[23] Ad esempio, la direzione studi e ricerche INPS, in base al sito istituzionale "fornisce supporto tecnico-scientifico all'elaborazione delle decisioni che l'Istituto assume nell'ambito delle proprie attività istituzionali attraverso (…) l'elaborazione di statistiche, di modelli di *data mining* e *machine learning*, anche in riferimento ai *big data*" (<https://www.inps.it/nuovoportaleinps/default.aspx?itemdir=53263>, visitato il 25 febbraio 2021).

[24] Come nel caso del cd. risparmiometro, menzionato nel Piano della performance dell'Agenzia delle entrate 2018-20, ma poi descritto solo in una decisione del Garante

consentono di andare oltre la notizia dell'uso di una determinata tecnologia (come nel caso delle circolari dell'INPS)[25].

In quarto luogo, tale mancanza di trasparenza, che ha ricadute sull'esercizio delle garanzie procedurali[26], potrebbe impattare negativamente anche su quelle processuali: là dove i soggetti interessati non siano al corrente del ricorso all'intelligenza artificiale nel processo decisionale e del suo ruolo nel decisione finale non possono contestarne né l'utilizzo, né gli esiti.

Vi è infine una criticità che, ad avviso di chi scrive, andrebbe affrontata: la mancanza, a monte, di una norma che consenta e disciplini il ricorso all'intelligenza artificiale nel procedimento amministrativo[27]. Ed invero questa, quantomeno per gli usi più avanzati, pur non traducendosi in decisioni automatizzate, è comunque scuscettibile di influenzare significativamente la decisione finale (attraverso una puntale definizione del supporto informativo che ne sarà alla base) e dunque difficilmente potrebbe ricondursi alla potestà organizzativa dell'amministrazione[28].

della Privacy del 20 luglio 2017, n. 321, *Sperimentazione di una procedura basata sull´utilizzo di informazioni fornite dall´Archivio dei rapporti finanziari e degli elementi presenti nell´Anagrafe tributaria per l´individuazione di profili di evasione rilevanti*.

[25] Ad esempio, INPS, circolare n. 23/2010, *Funzione di accertamento e verifica amministrativa - Attuazione del nuovo modello organizzativo delle strutture territoriali di produzione previsto dalla circolare n. 102 del 12/08/2009* e circolare n. 23/2010, *Funzione di accertamento e verifica amministrativa - Attuazione del nuovo modello organizzativo delle strutture territoriali di produzione previsto dalla circolare n. 102 del 12/08/2009*.

[26] "This century's automated decision making system combine individual adjudication with rulemaking while adhering to the procedural safeguards of neither" (D.K. Citron, *Technological Due Process*, 85 *Washington University L. Rev*. 1249, 2008.

[27] Alcuni usi trovano un fondamento normativo (ad esempio, art. 1, comma 682, legge di bilancio 2020 con riferimento al c.d. evasometro anonimizzato), altri sono basati su soft law (si vedano le varie circolari INPS o Agenzia entrate).

[28] Non è un passaggio di poco conto e la coerenza degli usi con il principio di legalità probabilmente avrebbe evitato la diffusa opacità che viene denunciata anche di seguito. Non si vuol con questo sostenere che l'adozione di una previsione normative generale sarebbe risolutiva ma, come evidenzia S. Civitarese Matteucci, "a more nuanced legal framework is desirable. One in which sector specific power-conferring rules, tailored to the context of any different policy and purpose, confer the relevant power to a public body" (*Public Administration Algorithm Decision-Making and the Rule of Law*, in *European Public Law*, vol. 27, n. 1, 2021, p. 129). Si veda anche S. Civitarese Matteucci, *Umano, troppo umano. Decisioni amministrative automatizzate e principio di legalità*, in *Diritto pubblico*, n. 1, 2019, p. 5 ss.; A. Masucci, *L'algoritmizzazione delle decisioni amministrative tra Regolamento europeo e leggi degli Stati membri*, in *Dir. pubblico*, n. 2, 2020, p. 943 ss.

3. *Regolazioni, provvedimenti amministrativi, controlli: il ruolo dell'intelligenza artificiale*

3.1 *Intelligenza artificiale e procedimenti di regolazione*

Le applicazioni nell'ambito di procedimenti di regolazione o definizione di *policies* possono avere riguardo alla scrittura di regole, così come alla raccolta e all'elaborazione di dati[29]. Ciò può avvenire attraverso tecnologie *real time* (come l'internet delle cose), oppure con l'analisi attraverso tecniche di *natural language processing*[30] di social media, di segnalazioni, denunce o reclami rivolti a una pubblica amministrazione, o ancora commenti presentati nell'ambito di consultazioni. La raccolta di dati può poi essere realizzata attraverso *distributed ledger technology,* ad esempio per registrare le transazioni che devono essere riportate ai regolatori dei mercati finanziari[31]. Le nuove tecnologie possono anche supportare i regolatori nello svolgimento di attività di *drafting*[32].

Queste applicazioni sembrano ancora poco diffuse nel nostro sistema istituzionale, seppure qualsiasi affermazione di questo tipo sconti il limite della mancanza di trasparenza lamentata nel paragrafo precedente. Quanto all'uso del *machine learning* per la lettura dei risultati di consultazioni con commenti di massa, può essere menzionata l'esperienza di due politiche "La buona scuola"[33] e "Rivoluzione@governo"[34], mentre un sistema intel-

[29] C. Coglianese, D. Lehr, *Regulating by robot*, 105 *Georgetown L. J.* 1160, 2017.

[30] Michael A. Livermore, Vladimir Eidelman, Brian Grom, *Computational assisted regulatory participation*, 93 *Notre Dame L. Rev.* 977, 2018.

[31] E. Micheler, A. Whaley, *Regulatory technology: replacing law with computer code*, in *European Business Organization Law Review*, 2020, vol. 21, p. 349 ss.

[32] W. Voermans, E. Verharden, *Leda: A Semi-Intelligent Legislative Drafting Support System,* in *Jurix*, 1993, p. 81 ss.; M. Zalnieriute, L. Burton Crawford, J. Boughey, L. Bennett Moses, S. Logan, *From Rule of Law to Statute Drafting. Legal Issues for Algorithms in Government Decision-Making*, in Woodrow Barfield (a cura di), *Cambridge Handbook on the Law of Algorithms,* Cambridge, Cambridge University Press, 2020, p. 256 ss.

[33] Nell'ambito della consultazione (che si è svolta dal 15 settembre al 15 novembre 2014) sono state inviate 6.470.000 risposte strutturate, analizzate attraverso la tecnica della linguistica computazionale per l'estrazione di concetti chiave (con il supporto della Fondazione Bruno Kessler).

[34] La consultazione pubblica svolta dal Governo nel 2014 ha utilizzato un account di posta elettronica per raccogliere osservazioni su 44 punti cardine della riforma della pubblica amministrazione. "In risposta alla consultazione, della durata di un mese (30 aprile – 30 maggio 2014), sono pervenute 39.343 mail, di cui 15.618 classificate come commenti di massa. Tali commenti sono stati il risultato di diverse campagne di mobilitazione

ligente di analisi automatica degli esposti è in corso di sperimentazione da parte della Consob[35]. L'Autorità per le garanzie nelle comunicazioni, nel quadro di un regolamento adottato dall'autorità[36] e dei codici di autoregolazione delle piattaforme da questa promossi, svolge un'attività di monitoraggio per individuare e prevenire la disinformazione *on line* attraverso un sistema di intelligenza artificiale che analizza sistematicamente i contenuti di articoli di quotidiani e siti web, trascrizioni di trasmissione televisive e radio, tweet/post[37].

L'uso di strumenti tecnologici avanzati nell'ambito di consultazioni o per processare segnalazioni presenta problemi specifici. Tale uso può portare, ad esempio, a non tenere conto di documenti che contengano errori ortografici o di commenti frutto di campagne di mobilitazione di massa[38], così da alterare il bagaglio informativo a disposizione del decisore pubblico, fino limitare le garanzie di partecipazione e compromettere la democraticità di un processo decisionale che si basi sulle consultazioni. Al contempo, l'uso dell'intelligenza artificiale per processare segnalazioni, denunce o esposti potrebbe, per questi stessi motivi, non far emergere comportamenti diffusi degli operatori economici a discapito dei concorrenti o dei consumatori; la parzialità dei dati raccolti potrebbe poi non mettere nella giusta evidenza un'eventuale esigenza di regolazione o di riforma delle regole esistenti, oppure di attivazione del *law enforcement* sotto forma di controlli o sanzioni.

Quanto alla scrittura di regole, occorre tener conto che qualsiasi traduzione in codici di una regolazione esistente o la scrittura in questa forma di una nuova regolazione rischia di essere eccessivamente semplificata, se non distorta[39], e porta - per questo stesso motivo - all'adozione di regole

promosse da altrettanti diversi soggetti organizzati" (C. Raiola, *Recensione. I commenti di massa nelle consultazioni pubbliche: l'analisi di S.J. Balla (et al.) sull'esperienza statunitense in materia ambientale*, in *Rassegna Trimestrale dell'Osservatorio AIR*, n. 4, 2020, p. 67-68).

[35] Primo rapporto ASTRID su *L'uso dell'intelligenza artificiale nel sistema amministrativo italiano*, a cura di E. Chiti, B. Marchetti, N. Rangone, *BioLaw Journal*, in corso di pubblicazione.

[36] Allegato B alla delibera n. 157/19/CONS del 15 maggio 2019, *Regolamento recante disposizioni in materia di rispetto della dignità umana e del principio di non discriminazione e contrasto all'hate speech*.

[37] Gli Esiti Di Questa Attività Sono Pubblicati Nel Bollettino Agcom, *Osservatorio Sulla Disinformazione Online*.

[38] S.J. Balla, A.R. Beck, E. Meehan, A. Prasad, *Lost in the flood?: Agency responsiveness to mass comment campaigns in administrative rulemaking*, in *Regulation and Governance*, 2020.

[39] "Policy is often distorted when programmers translate it into code. (…) This is, in part, because the artificial languages intelligible to computers have a more limited

dettagliate e poco flessibili (con i ben noti limiti che questo comporta)[40], oltre a risultati fuorvianti se un software non è tenuto al passo con la nuova normativa[41]. Inoltre questo approccio porta ad una tendenziale soprapposizione tra *adjudication* e *rule-making*, con perdita delle garanzie procedurali ad entrambi i livelli[42].

A fronte del limitato uso - al momento attuale - di nuove tecnologie a fini di regolazione, è interessante notare come la regolazione possa favorire o indurre i regolati a fare uso di intelligenza artificiale. Ad esempio, a seguito della regolazione che ha previsto il passaggio ai contatori elettrici di seconda generazione-*smart meter* 2g (che consentono novantasei letture giornaliere, a fronte delle tre mensili di quelli tradizionali)[43], i venditori si trovano a gestire un'enorme mole di dati relativi ai consumi, situazione che ha indotto alcuni a dotarsi di *data-base* di nuova generazione e ad elaborare questi dati attraverso *machine learning*[44]. Ciò comporta non solo una potenziale riduzione dei costi di sbilanciamento sostenuti dagli operatori, ma anche l'offerta ai clienti di informazioni che consentano di fare scelte più consapevoli nell'uso efficiente dell'energia[45].

vocabulary than human languages. Computer languages may be unable to capture the nuances of a particular policy. Code writers also interpret policy when they translate it from human language to computer code" (D.K. Citron, *Technological Due Process,* cit., p. 1261). Dunque questa impostazione porta verso regole dettagliate e poco flessibili, con i ben noti limiti che questo comporta (C.S. Diver, *The optimal precision of administrative law*, 93 *Yale L. J.* 65, 1983.

[40] C.S. Diver, *The optimal precision of administrative law*, cit.

[41] Come nel caso dell'*Arizona Correctional Information System*, volto calcolare le date di rilascio della popolazione carceraria dell'Arizona, che ha portato a numerosissimi errori, alcuni dei quali riconducibili alla mancata considerazione delle previsioni di una normativa del 2019 che ha introdotto importanti sconti di pena (*Un software gestisce la popolazione carceraria dell'Arizona: quali le conseguenze?*, in Osservatorio sullo Stato digitale, IRPA: https://www.irpa.eu/un-software-gestisce-la-popolazione-carceraria-dellarizona-quali-le-conseguenze/, visitato il 12 settembre 2021).

[42] "The dichotomy between these procedural regimes is rapidly becoming outmoded. This century's automated decision making systems combine individual adjudications with rulemaking while adhering to the procedural safeguards of neither" (D.K. Citron, *Technological Due Process,* cit., p. 1249).

[43] La normativa di riferimento è riconducibile alle direttive 2009/72/CE e 2009/73/CE, e all'art. 9, comma 3, d.lgs. n. 102/2014 (di attuazione della direttiva 2012/27, sull'efficienza energetica) che attribuisce ad ARERA il compito di definire le specifiche tecniche. ARERA inoltre approva i Piani di messa in servizio di *smart metering* 2G delle imprese distributrici.

[44] Primo rapporto ASTRID su *L'uso dell'intelligenza artificiale nel sistema amministrativo italiano*, cit.

[45] Dal momento che non tutti gli operatori sono necessariamente in grado di effettuare una

3.2 *Intelligenza artificiale e procedimenti per l'adozione di decisioni amministrative*

Con riferimento ai procedimenti per l'adozione di decisioni amministrative[46], le esperienze più note (perché più risalenti o per il contenzioso che ne è derivato) attengono all'uso di algoritmi con logica lineare. È questo il caso dell'attribuzione di sedi lavorative agli insegnati[47], del calcolo delle tariffe di energia e rifiuti dal parte di ARERA[48], delle sovvenzioni allo spettacolo[49]. Si tratta di tecnologie molto semplici, attraverso le quali si può gestire l'intera fase di procedimenti vincolati (come nel caso dell'e-procurement con aggiudicazione al prezzo più basso)[50]. Queste applicazioni estremamente semplici hanno portato giurisprudenza (e dottrina) a delineare uno statuto della decisione algoritmica, basato sulla trasparenza (accessibilità dell'algoritmo e del codice sorgente, comprensibilità della logica e del ragionamento, spiegabilità[51]),

siffatta elaborazione dei dati e di trarne conseguentemente vantaggio attraverso l'offerta di servizi avanzati agli utenti, ricadute di questo tipo dovrebbe essere valutate in sede di analisi preventiva di impatto della regolazione sia sull'innovazione, che sulla concorrenza.

[46] C. Coglianese, L.M. Ben Dor, *AI in Adjudication and Administration: A Status Report on Governmental Use of Algorithmic Tools in the United States*, in *Faculty Scholarship at Penn Law*, 2020.

[47] Come nel caso della mobilità del personale docente, che ha dato occasione di pronunciarsi sia a Tar Lazio (sez. III*bis*, n. 3742/2017 e n. 3769/2017) che a Consiglio di Stato (VI n. 8472/2019, n. 8473 e n. 8474 dello stesso anno, n. 881/2020). Anche con l'apertura dell'anno scolastico 2021/2022, numerosi errori dell'algoritmo nell'attribuzione delle cattedre vengono riportati dai quotidiani (si veda ad esempio, La Stampa, 10 settembre 2021).

[48] G. Avanzini, *Decisioni amministrative e algoritmi informatici. Predeterminazione, analisi predittiva e nuove forme di intellegibilità*, Napoli, Editoriale Scientifica, 2019, p. 53-57.

[49] D.M. 27 luglio 2017, *Criteri e modalità per l'erogazione, l'anticipazione e la liquidazione dei contributi allo spettacolo dal vivo, a valere sul Fondo unico per lo spettacolo di cui alla legge 30 aprile 1985, n. 163*. Cfr. anche G. Avanzini, *Decisioni amministrative e algoritmi informatici. Predeterminazione, analisi predittiva e nuove forme di intellegibilità*, cit., p. 51-53.

[50] Così il Consiglio di Stato che, nel parere allo schema che la valutazione delle offerte con sistemi telematici, ha ritenuto che "nelle sole procedure di affidamento aggiudicate secondo il criterio del prezzo più basso – ove possono essere eseguite in modo automatizzato, senza esercizio di potere discrezionale da parte del seggio di gara, sia la determinazione della soglia di anomalia dell'offerta economica sia l'elencazione dei ribassi d'asta (dal minore al maggiore) – meglio potranno essere sfruttate le potenzialità del sistema telematico" (sez. consultiva per gli atti normativi, adunanza 17 novembre 2020, n. 1322/2020). Sul punto, G. Fasano, *L'intelligenza artificiale nella cura dell'interesse generale*, in *Giornale di diritto amministrativo*, n. 6, 2020, p. 717.

[51] "La spiegabilità attiene alla capacità di spiegare sia i processi tecnici di un sistema

sul diritto alla partecipazione e all'intervento umano (non esclusività della decisione algoritmica)[52], sulla non discriminazione e sulla tutela giurisdizionale[53]. Si tratta di soluzioni che, proprio perché delineate con riferimento ad algoritmi semplici, per certi versi restano indeterminate e dunque difficilmente applicabili in concreto, come la spiegabilità, a causa dell'autoapprendimento caratteristico di alcune applicazioni del *machine learning*. Gli stessi limiti gravano sul requisito dell'intervento umano, che aprono una serie di questioni che vanno dal requisito minimo della partecipazione umana, al grado di partecipazione auspicabile, all'effettiva capacità degli individui di rapportarsi con le indicazioni dell'intelligenza artificiale senza essere influenzati da *bias* cognitivi (su cui si tornerà nel paragrafo 4).

di IA che le relative decisioni umane. (…) Tale spiegazione dovrebbe essere tempestiva e adeguata alle competenze del portatore di interesse in questione (un non esperto, un'autorità di regolamentazione o un ricercatore)" (*Orientamenti etici per una IA affidabile*, 2019, del Gruppo di esperti ad alto livello sull'intelligenza artificiale istituito dalla Commissione europea nel 2018, p. 20-21). Un riferimento al "diritto ad ottenere una spiegazione della decisione conseguita [con trattamento automatizzato] e di contestare la decisione" si rinviene nel considerando 71, del regolamento n. 679/2016/UE, che però non viene generalmente considerato fonte di un vero e proprio diritto alla "spiegabilità" (tra i tanti, S. Wachter, B. Mittelstadt, L. Floridi, *Why a right to explanation of automated decision-making does not exist in the general data protection regulation*, in *International Data Protection Law*, vol. 7, n. 2, 2017, p. 76 ss.).

[52] Cons. St. VI 4 febbraio 2020, n. 881; Cons. St. VI 13 dicembre 2019, n. 8472, 8473, 8474; Cons. St. VI 8 aprile 2019, n. 2270.

[53] Sulla "legalità algoritmica", D.-U. Galetta, J.G. Corvalán, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, n. 3/2019; R. Cavallo Perin, *Ragionando come se la digitalizzazione fosse data*, in *Dir. amm.*, n. 2, 2020, p. 305 ss.; E. Carloni, *I principi della legalità algoritmica. Le decisioni automatizzate di fronte al giudice amministrativo,* in *Dir. amm.*, n. 2, 2020, p. 273 ss.

3.3 *Intelligenza artificiale e attuazione amministrativa*

In ordine all'attuazione amministrativa[54], le esperienze più rilevanti attengono all'accertamento tributario sintetico da parte dell'Agenzia delle entrate, alla vigilanza bancaria per la rilevazione di anomalie indicative di possibili riciclaggio[55], alla sicurezza pubblica con le prime applicazioni di polizia predittiva[56]. Di grande interesse risultano poi le attuali sperimentazioni dell'uso del *machine learning* per la valutazione del rischio al fine della razionalizzazione dei controlli sulle imprese, che coinvolgono la provincia autonoma di Trento e le regioni Lombardia e Campania[57].

L'uso delle nuove tecnologie nel *law enforcement* può essere foriera di problematiche specifiche, come discriminazioni di tipo diretto, che riguardano individui, e indiretto, che interessano determinate attività, gruppi sociali, aree maggiormente sottoposte a controlli. Con riferimento alle prime, occorre valutare con attenzione il peso da riconoscere alla "storia" di un individuo o impresa (ad esempio, esiti di controlli, denunce, condanne) che comunque non dovrebbero mai avere un peso esclusivo nella definizione del grado di rischio riferita a quell'individuo o impresa. Al contempo, seppure le discriminazioni indirette appaiano meno pericolose di quelle dirette, il rischio che i dati vengano considerati predittivi del grado di rischio di commettere un illecito da parte di una persona è sempre presente, come già si è concretizzato in altri ordinamenti giuridici[58].

[54] Quanto ai controlli sulle imprese sia consentito rinviare a N. Rangone, *Semplificazione ed effettività dei controlli sulle imprese*, in *Riv. trim. dir. pubblico*, n. 3, 2019, p. 902-911.
[55] L'Unità di informazione finanziaria per l'Italia-UIF sta sperimentando l'uso del *machine learning* per l'individuazione di operazione sospette di riciclaggio o di finanziamento del terrorismo da sottoporre a ispezione. In generale con riferimento a controlli informati alla trasmissione di report periodici, in prospettiva si potrebbe superare la fase della *compliance* nella misura in cui controllori e controllati convergessero nell'uso di un *distributed ledger*, le cui attuali sperimentazioni attengono principalmente al FinTech. Sulle potenzialità (in termini di risparmi per regolatori e regolati e maggiore precisione dei controlli) e rischi, si veda E. Micheler, A. Whaley, *Regulatory technology: replacing law with computer code*, cit.
[56] M.B. Armiento, *La polizia predittiva come strumento di attuazione amministrativa delle regole*, in *Diritto amministrativo*, n. 4, 2020, p. 983 ss. Dello stesso autore si veda il contributo *Nuove tecnologie e "nuova" sicurezza urbana*, in questo numero.
[57] Al riguardo si rinvia al contributo di F. Blanc, M. Benedetti, C. Bertone, *L'informatica e il machine learning al servizio della semplificazione dei controlli sulle imprese: un equilibrio ancora da definire*, cit.
[58] Amnesty International, *Trapped in the matrix*, 2020.

Per evitare questi esiti e più in generale per assicurare che l'algoritmo funzioni, non solo questo "deve continuamente essere alimentato di dati e informazioni, ma deve anche essere in grado di modificare previsioni e parametri di rischio in precedenza assegnati. (…) L'algoritmo dunque va manutenuto, costantemente aggiornato e perfezionato, ma anche monitorato al fine di supervisionare periodicamente (per esempio ogni anno) l'efficacia delle sue previsioni" [59].

Infine, con riferimento alle applicazioni in alcuni ambiti (come i controlli tributari o quelli di polizia), vi è il rischio concreto che si verifichi un'eccessiva concentrazione sull'uso delle nuove tecnologie in funzione repressiva, là dove queste potrebbero essere messe a frutto anche in funzione di supporto mirato alla *compliance*[60].

4. *Il complicato rapporto tra intelligenza artificiale e* human bounded rationality

La complementarietà tra intelligenza umana e intelligenza artificiale viene richiesta da più parti per bilanciare gli eventuali *bias* se non errori dell'algoritmo[61] e per riconoscere la legittimità delle decisioni finali[62].

Questo presidio presenta, però, non pochi problemi attuativi con riferimento alle applicazioni di intelligenza artificiale.

In particolare, non è chiaro cosa si pretende dall'intervento umano. Resta, ad esempio, indeterminato il modo in cui dovrebbe e/o potrebbe in concreto svolgere l'interazione tra intelligenza umana - del funzionario - e intelligenza artificiale della macchina: si tratta di fasi separate oppure è realizzabile una sorte di interlocuzione, se non di controllo? Se questo fosse il caso, come sembra emergere dalla proposta di regolamento

[59] F. Blanc, M. Benedetti, C. Bertone, *L'informatica e il machine learning al servizio della semplificazione dei controlli sulle imprese: un equilibrio ancora da definire*, cit.

[60] In questa direzione muovono le applicazioni nei mercati finanziari del RegTech, vale a dire "la tecnologia applicata alla regolamentazione (…) che si concentra sulle tecnologie che possono agevolare il rispetto degli obblighi normativi in maniera più efficiente ed efficace rispetto alle capacità attuali" (comunicazione della Commissione UE sulla strategia in materia di finanza digitale per l'UE, COM 2020(591) fin.). Sulle limitate applicazioni in altri ambiti, come quello fiscale, sia consentito rinviare a N. Rangone, *Semplificazione ed effettività dei controlli sulle imprese*, cit., p. 896 ss.

[61] S.G. Mayson, *Bias In, Bias Out*, cit.

[62] B. Marchetti, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica,* in *BioLaw Journal*, n. 2/2021.

europeo quanto agli algoritmi ad alto rischio, vi è da chiedersi come una supervisione da parte del singolo sarebbe in concreto realizzabile, mancandone spesso il tempo e la capacità, senza considerare che nel caso di dati elaborati attraverso *machine learning* - che consente un apprendimento automatico e progressivo da parte della "macchina" - un controllo diventa addirittura impossibile. Il giudice amministrativo italiano non lo spiega e la definizione che ne dà il Gruppo di esperti ad alto livello sull'intelligenza artificiale istituito dalla Commissione europea è alquanto vaga[63].

Il presidio dell'intervento umano rischia poi di restare lettera morta se si considera la razionalità limitata degli individui, anche i più esperti, e le conseguenti due possibili opposte reazioni dell'intelligenza umana di fronte ai dati derivanti dall'intelligenza artificiale: eccessiva deferenza ("the algorithm make me do it" approach)[64], da un lato, diffidenza se non rifiuto, dall'altro.

Qui entrano in gioco i *bias* legati al rapporto uomo-macchina, la cui esistenza e il cui peso nel processo decisionale individuale non può essere assunto in astratto e andrebbero, invece, verificato in concreto con specifici esperimenti comportamentali. Ed invero, si parla in letteratura di reazioni di completo affidamento nei confronti di una presunta oggettività dell'indicazione derivante dall'intelligenza artificiale (*automation bias*)[65]. Se questo dovesse verificarsi essere il caso, è ovvio che la differenza tra un sistema completamente automatizzato e uno in cui ha un ruolo l'elemento umano si perderebbe completamente[66].

[63] "La sorveglianza [umana] può avvenire mediante meccanismi di governance che consentano un approccio con intervento umano (human-in-the-loop-HITL), [vale a dire] "la possibilità di intervento umano in ogni ciclo decisionale del sistema, che in molti casi non è né possibile né auspicabile" (*Orientamenti etici per una IA affidabile*, 2019, cit., p. 18).

[64] D.S. Rubenstein, *Acquiring Ethical AI*, cit., p. 123.

[65] L'*automation bias* viene descritto come "the use of automation as a heuristic replacement for vigilant information seeking and processing" (L.J. Skitka et al., *Automation Bias and Errors: Are Crews Better Than Individuals?*, in *The International Journal of Aviation Psychology*, 2000, vol. 10, n. 1, p. 85). Di particolare interesse a questo riguardo sono gli studi e gli esperimenti svolti con riferimento alle reazioni alle indicazioni delle "macchine" dei piloti di aereo (K.L. Mosier et al., *Automation Bias: Decision Making and Performance in High-Tech Cockpits*, in *International Journal of Aviation Psychology*, 1997, n. 8, vol. 1, p. 47 ss.) e in medicina (K. Goddard, A. Roudsari, J.C. Wyatt, *Automation bias: a systematic review of frequency, effect mediators, and mitigators*, in *J. Am. Med. Inform. Assoc*., 2012, vol. 19, p. 121 ss.). Peraltro, la proposta di regolamento europeo sull'uso dell'intelligenza artificiale assume l'esistenza di questo *bias* e senza dare indicazioni su come far sì che non ne venga conseguentemente neutralizzato lo *human oversight* (art. 14).

[66] "We are worried that if we simply thrust the human at the output end of the running

È peraltro anche possibile che si manifesti un atteggiamento opposto di rifiuto, che può derivare da sfiducia nella tecnica (*algoritm aversion*)[67], oppure da quello che è stato definito *illusion of validity*[68] (un'ingiustificata fiducia nel ragionamento umano che spesso caratterizza gli esperti)[69]. Può poi entrare in gioco una reazione molto diffusa, sia tra i regolati, che tra i decisori pubblici, definita *confirmation bias*, che induce a tener conto dei dati a disposizione (nel nostro caso quelli elaborati dalle macchine) solo quando a supporto dell'intuizione o convinzione formatasi dal funzionario ed a rifiutarli in caso di difformità[70]. Vi è peraltro da considerare che l'avversione potrebbe anche essere riconducibile a limitate competenze tecnologiche, che non portano a compiere neanche quel passo minimo per acquisire e utilizzare i dati elaborati dall'intelligenza artificiale[71].

Non è possibile dare indicazioni definitive riguardo al ruolo dei *bias* nel complicato rapporto tra intelligenza artificiale e intelligenza umana, sia

model, there is very little she can do to root out bias. The human becomes a rubber stamp for the machine, providing nothing more than a cosmetic reason to lull ourselves into feeling better about the results" (D. Lehr, P. Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, in *University of California*, vol. 51, 2017, p. 716). Sul punto anche D.K. Citron (*Technological Due Process*, cit., p. 1272) che evidenzia anche come "over time, human operators may lose the skills that would allow them to check a computer's recommendation".

[67] "The results of five studies show that seeing algorithms err makes people less confident in them and less likely to choose them over an inferior human forecaster. (…) Our findings do suggest that people will be much more willing to use algorithms when they do not see algorithms err, as will be the case when errors are unseen, the algorithm is unseen (as it often is for patients in doctors' offices), or when predictions are nearly perfect" (B.J. Dietvorst, J.P. Simmons, C. Massey, *Algorithm aversion: People erroneously avoid algorithms after seeing them err*, in *Journal of Experimental Psychology: General*, 2015, vol. 144, n. 1, p. 114 ss., 10-11 in Scholarly Commons).

[68] A. Tversky, D. Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, cit.

[69] D. Kahneman, G. Klein, *Conditions for Intuitive Expertise. A Failure to Disagree*, cit.

[70] Si tratta di un *bias* molto diffuso, che interessa gli individui in quanto destinatari delle regole e in quanto regolatori. Sul punto la letteratura è diffusa, si veda, per tutti, C.G. Lord, C. Taylor, *Biased Assimilation: Effects of Assumptions and Expectations on the Interpretation of New Evidence*, in *Social and Personality Psychology Compass*, 2009, n. 3, p. 827; E. Zamir, D. Teichman, *Behavioural law and economics*, Oxford, Oxford University Press, 2018, p. 399; S. Stern, *Cognitive Consistency: Theory Maintenance and Administrative Rulemaking,* 63 *University of Pittsburgh L. Rev*. 589, 2002.

[71] F. De Leonardis (*Big Data, decisioni amministrative e "povertà" di risorse della pubblica amministrazione*, in E. Calzolaio (a cura di), *La decisione nel prisma dell'intelligenza artificiale,* Milano, CEDAM- Wolters Kluwer, 2020, p. 159) rileva che l'età avanzata dei dipendenti pubblici italiani che potrebbe costituire un freno al *government by data*.

perché i numerosi studi non vanno tutti nella stessa direzione[72], sia perché manca una sperimentazione specifica riferita al contesto italiano. Non vi è dubbio, però, che queste diverse reazioni vadano prese in considerazione e affrontate con strumenti di *empowerment* cognitivo adeguati[73], in grado di rendere consapevoli gli interessati dell'esistenza di un determinato *bias* e dunque delle possibili reazioni individuali, non necessariamente razionali. In questa direzione potrebbe muovere, in particolare, una formazione mirata sui *bias* che intercorrono più frequentemente nel rapporto uomo-macchina[74]. Nello specifico, il funzionario dovrebbe poter comprendere le funzionalità e i limiti del sistema di intelligenza artificiale e interpretare correttamente i relativi esiti[75]; in quest'ottica è importante la partecipazione della struttura operativa alla progettazione dell'algoritmo, che sia questa effettuata all'interno o all'esterno dell'amministrazione. Al contempo, un sistema più o meno formalizzato di *accountability* (interna o esterna) aumenta la consapevolezza nella misura in cui porta il funzionario a individuare e indicare i dati generati dalla tecnologia che lo hanno indotto ad adottare una determinata decisione[76]. Più in generale andrebbe sempre

[72] Ad esempio, J.M. Logg, J.A. Minson, D.A. Moore (*Algorithm appreciation: people prefer algorithm to human judgment*, in *Organisational Behaviour and Human Processes*, 2009, n. 151 p. 99) evidenziano che nonostante gli operatori economici esitino a presentare un servizio come riconducibile alla sola intelligenza artificiale per paura dell'avversione del cliente, "one simple way to increase adherence to advise is to provide advice from an algorithms" (come nel caso dei Robo-advisors nei mercati finanziari).

[73] Sulla distinzione tra *nudging* and *empowerment* sia consentito rinviare a F. Di Porto, N. Rangone, *Behavioural Sciences in Practice: Lessons for EU Policymakers*, in A. Alemanno, A.-L. Sibony (a cura di), *Nudge and the Law: A European Perspective?,* Oxford, Hart Publishing, 2015, p. 29; A. Van Aaken, *Constitutional Limits to Paternalistic Nudging: A Proportionality Assessment*, in A. Kemmerer, C. Möllers, M. Steinbeis, G. Wagner (a cura di), *Choice architecture in democracy. Exploring the Legitimacy of Nudging*, Baden-Baden, Hart-Nomos, 2016, p. 161 ss.

[74] D.K. Citron, *Technological Due Process*, in *Washington University Law Review*, cit., p. 1306-1307. A questo riguardo è interessante notare come esperimenti cognitivi, seppure risalenti, hanno segnalato che la formazione sull'*automation bias* sarebbe efficace nel prevenire gli errori di "commissione" ma non quelli di "omissione" (L.J. Skitka et al., *Automation Bias and Errors: Are Crews Better Than Individuals?*, cit., p. 95).

[75] Così l'art. 14 comma 4, lett. a) e d) della proposta di regolamento europeo, già citata, che evidenzia anche che vanno predisposte misure affinché le persone cui è affidata al sorveglianza umana siano consapevoli "della possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio ("distorsione dell'automazione"), in particolare per i sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche" (lett. b).

[76] M. Hallsworth, M. Egan, J. Rutter, J. McCrae, *Behavioural Government. Using*

previsto un monitoraggio e una valutazione *ex post* dell'esito dell'uso dell'intelligenza artificiale nei processi decisionali pubblici alla luce di diversi parametri, che vanno dai risparmi, al rispetto dei tempi del procedimento, all'accettazione nel sistema organizzativo, al grado di soddisfazione degli utenti o dei cittadini, alla numerosità degli errori riscontrati[77].

5. *Nuove tecnologie e fiducia*

La fiducia è un fattore determinante del successo o meno dell'innovazione nelle pubbliche amministrazioni (in generale) e dell'uso dell'intelligenza artificiale (in particolare)[78].

Ci sono almeno tre dimensioni della fiducia nelle nuove tecnologie da prendere in considerazione. Fiducia dei decisori politici, fiducia dei funzionari, fiducia di cittadini e imprese.

La fiducia dei decisori finali nelle nuove tecnologie è fondamentale da vari punti di vista. Spesso il decisore politico è colui che contribuisce in modo determinante all'apertura dei processi decisionali all'intelligenza artificiale, attraverso assunzioni aperte ai *computer scientists*[79], l'acquisto delle tecnologie necessarie alla raccolta e stoccaggio di una mole di dati che possa alimentare il funzionamento dell'intelligenza artificiale[80] o la stipula

behavioural science to improve how governments makedecisions, The Behavioural Insights Team, 2018.

[77] D. Freeman Engstrom, D.F. Ho (*Algorithmic Accountability in the Administrative State,* cit.) propongono di impostare un "prospective benchmarking", vale a dire una comparazione tra casi decisi con il supporto dell'intelligenza artificiale e frutto di un processo decisionale tradizionale "This "human alongside the loop" approach provides critical information and a comparison set to help smoke out when an algorithm has gone astray, when encoding the past may miss new trends, when an algorithm may create disparate impact, or when "automation bias" causes excessive deference to machine outputs".

[78] F. Blanc, *Regulatory Delivery, Trust and Distrust. Avoiding Vicious Circles,* in M. De Benedetto, N. Lupo, N. Rangone (a cura di), *The Crisis of Confidence in Legislation,* Baden-Baden, Nomos-Hart, 2020, p. 307 ss.

[79] "Perhaps ironically, effective design and deployment of algorithms requires sufficient human capital. This includes the need for analysts and data scientists with technical skills, but also experts who can appreciate the specific challenges associated with the use of artificial intelligence in government" (C. Coglianese, *A framework for governmental use of machine learning*, cit., p. 38).

[80] "Algorithms are dependent upon an analytic infrastructure that includes a large volume of data as well as the hardware, software, and network resources needed to support machine learning analysis of such data" (*ibidem*, p. 41)

di accordi con altre amministrazioni per l'interconnessione di banche dati o l'uso del *federated learning*[81]. È ancora il decisore politico che dà seguito o meno alle indicazioni che provengono da un'istruttoria o preistruttoria nella quale ha avuto un ruolo l'intelligenza artificiale.

La fiducia del funzionario nei confronti delle indicazioni che provengono dall'uso di intelligenza artificiale è anch'essa un fattore imprescindibile per il successo di un procedimento *data driven*. La fiducia attiene non solo alla correttezza delle indicazioni provenienti dall'intelligenza artificiale, ma anche alla *fairness* della procedura alla base di tali indicazioni (che può riguardare i dati utilizzati o la trasparenza della loro elaborazione): se questa collide con l'*individual fairness*[82] o con l'imparzialità secondo l'interpretazione del canone costituzionale, le indicazioni provenienti dall'intelligenza artificiale nel settore pubblico vengono adattate - se non ignorate - alla luce dei valori degli "screen-level bureaucrats"[83].

La fiducia di cittadini e imprese nei confronti di regolazioni o decisioni amministrative (di *adjudication* o *enforcement*), che devono essere e apparire affidabili[84], è altrettanto fondamentale. In mancanza, aumentano le contestazioni sociali e il contenzioso, fattori che rischiano di alimentare una resistenza interna (da parte dei funzionari), così che da strumento per il buon andamento, l'intelligenza artificiale potrebbe diventare un ostacolo ad esso. Ci sono almeno due dimensioni, ad avviso di chi scrive, su cui si fonda la fiducia di cittadini e imprese. Una prima dimensione, di tipo tecnico-giuridico, attiene alla trasparenza in ordine agli usi dell'intelligenza artificiale (di cui i soggetti interessati dalle decisioni finali devono avere

[81] "The approach enables multiple organizations to collaborate on the development of models, but without needing to directly share secure data with each other. Over the course of several training iterations, the shared models get exposed to a significantly wider range of data than what any single organization possesses in-house. In other words, FL decentralizes machine learning by removing the need to pool data into a single location" (Open Data Science 3 aprile 2020).

[82] Centre for Data Ethics and Innovation, *Review into bias in algorithmic decision-making*, 2020, p. 4-68.

[83] R. Binns, *Human judgment in algorithm loops*, in *Regulation and Governance*, 2020, p. 10. "When a human-in-the-loop adjusts algorithmic scores to compensate for known algorithmic biases, the human interventions may not be fair, and may be illegal. For instance, if an algorithm used for government hiring exhibits bias toward men, to what extent (if any) can the algorithmic score be upwardly adjusted for women (without running afoul of equal-protection principles)? As this example illustrates, fairness depends on how it is defined—and there are many legitimate definitions" (D.S. Rubenstein, *Acquiring Ethical AI*, in *Florida Law Review*, cit., p. 27).

[84] Centre for Data Ethics and Innovation, *Review into bias in algorithmic decision-making*, 2020, p. 4.

contezza)[85], all'effettività delle garanzie procedurali, alla completezza della motivazione della decisione finale che di questi usi faccia menzione. La seconda dimensione, più sfuggente, attiene alla percezione di affidabilità delle indicazioni che provengono dalle applicazioni tecnologiche, di fondamentale importanza per alimentare la fiducia. Anche a questo riguardo, per alimentare la percezione di affidabilità è fondamentale la trasparenza[86]. In quest'ottica, sarebbe importante che venissero indicati nei siti istituzionali i procedimenti interessati dall'uso di algoritmi, che venissero esplicitate le tecnologie utilizzate, il loro sviluppo interno o esterno all'amministrazione, l'esistenza o meno di sistemi di monitoraggio del funzionamento/supervisione e revisione periodica[87]. Dovrebbero poi essere trasparenti i dati utilizzati e da questa trasparenza dovrebbe risultare la loro affidabilità (anche perché raccolti e letti da soggetti non in potenziale conflitto interessi) e provenienza (dati pubblici o forniti da soggetti terzi, social media). Più in generale, una trasparenza così estesa darebbe conto della *procedural fairness* nell'uso dell'intelligenza artificiale da parte delle pubbliche amministrazioni, che costituisce il fondamento della fiducia e alimenta il rispetto volontario delle regole e delle decisioni pubbliche[88].

6. *Considerazioni conclusive*

Le esperienze di utilizzo dell'intelligenza artificiale da parte delle pubbliche amministrazioni italiane sono accomunate da una notevole

[85] La proposta di regolamento europeo già citato evidenzia che le persone devono sapere di interagire con un sistema di intelligenza artificiale ad alto rischio, tra le quali possono essere comprese le varie attività di *law enforcement*, ma anche il risconoscimento di sovvenzioni (allegato III alla proposta di regolamento).

[86] Nel menzionato Executive Order n. 13960/200 titolato "Promoting the use of trustworthy artificial intelligence of Federal Government", transparency "in disclosing relevant information regarding their use of AI to appropriate stakeholders, including the Congress and the public" è un principio chiave.

[87] Questa trasparenza può però essere ostacolata se non resa impossibile nei frequenti casi di esternalizzazione dello sviluppo del sistema (D.S. Rubenstein, *Acquiring Ethical AI*, in corso di pubblicazione). Sui problemi specifici a questo legati, anche in termini di *rick of gaming* e *adversarial learning*, si rinvia a D. Freeman Engstrom, D.E. Ho, C.M. Sharkey, M.-F. Cuéllar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, Report submitted to the Administrative Conference of the United States, 2020, p. 86 ss.).

[88] T.R. Tyler, *Procedural Fairness and Compliance with the Law*, in *Swiss Journal of Economics and Statistics*, 1997, vol. 133, n. 2, p. 219 ss.

opacità. Non è chiaro il perimetro delle amministrazioni coinvolte, il tipo di tecnologie utilizzate, la produzione interna o esterna, il ruolo nel processo decisionale, i dati che vengono lavorati, l'esistenza o meno di sistemi di manutenzione. Complessivamente, questa mancanza di trasparenza, pur con i rischi a questa connessi[89], ha ricadute rilevanti sull'effettività delle garanzie procedurali e processuali degli interessati.

Mancano poi orientamenti chiari a supporto delle amministrazioni soprattutto quanto all'uso delle tecnologie più avanzate, che presentano problematiche specifiche con riferimento all'uso nella scrittura delle regole, alla lettura degli esiti delle consultazioni, alla programmazione dei controlli e più in generale quanto al rapporto uomo-macchina nell'*adjudication* e nell'*enforcement*. L'affidamento della definizione di questi parametri al solo giudice amministrativo (alla stregua di una *ex post regulation*) [90], non solo non ne assicura la stabilità, ma, guardando necessariamente al passato dunque ad impieghi non particolarmente sofisticati delle nuove tecnologie, rischia di non offrire risposte alle problematiche peculiari connesse all'uso dell'intelligenza artificiale.

Al contempo, andrebbe chiaramente sciolto il nodo della distinzione tra dati coperti dalla privacy e non, che al momento ostacola la condivisione di informazioni tra pubbliche amministrazioni, determinante per alimentare il funzionamento dei sistemi di intelligenza artificiale.

Le singole amministrazioni non dovrebbero essere lasciate sole nella soluzione di questi nodi, ma dovrebbero potersi muovere nel quadro di uno statuto dell'uso dell'intelligenza artificiale, basato su principi, declinati attraverso indicazioni flessibili, che possano essere adattate alle diverse realtà senza ostacolare l'innovazione[91]. In tale contesto, potrebbe, ad

[89] "With simpler forms of adversarial machine learning, adversaries can, for instance, exploit algorithmic tools to obtain favorable determinations, without changing the underlying characteristic the algorithm is designed to measure. At the extreme, regulatory targets can even gain access to the tool itself and feed it new data to corrupt its outputs" (D. Freeman Engstrom, D.E. Ho, C.M. Sharkey, M.-F. Cuéllar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, cit., p. 86).

[90] B. Marchetti, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica,* cit., p. 5.

[91] Magari facendo uso di specifiche tecniche come l'analisi di impatto sull'innovazione (prevista a livello europeo dal *Better regulation Toolbox*, tool 21) o la sperimentazione in un ambiente controllato (come previsto dalla proposta di regolamento europeo citata all'art. 53). Sulle *regulatory sandboxes* come strumenti di *better regulation*, R. Baldwin, M. Cave, *Taming the corporation. How to regulate for success*, Oxford, Oxford University Press 2021, p. 109 ss.; sui rischi per la concorrenza connessi a questo strumento, B. Knight, T. Mitchell, *The sandbox paradox: balancing the need to facilitate innovation*

esempio, essere richiesto che il ricorso all'intelligenza artificiale, così come la produzione interna o esterna, siano oggetto di una valutazione che tenga conto dei relativi vantaggi-svantaggi e vengano motivati anche alla luce degli esiti di queste ponderazioni[92].

with the risk of regulatory privilege, in *Mercatus Working Paper, Mercatus Center at George Mason University*, 2020.

[92] A questo riguardo C. Coglianese (*A framework for governmental use of machine learning*, cit., p. 65-66) suggerisce, tra le altre cose, "a qualitative (or "soft") benefit-cost analysis. This approach is also called multi-factor policy analysis. Basically, it calls for the public administrator to run through a checklist of criteria against which both the human-based status quo and the digital alternative should be judged. (…) The administrator compares how well each alternative will fare against each criteria, without converting them into a common unit. (…) The main question will be what criteria or factors should such a multi-factor analysis take into account".

Paolo Cavaliere, Graziella Romeo

From Poisons to Antidotes: Algorithms as Democracy Boosters

Abstract: Under what conditions can artificial intelligence contribute to political processes without undermining their legitimacy? Thanks to the ever-growing availability of data and the increasing power of decision-making algorithms, the future of political institutions is unlikely to be anything similar to what we have known throughout the last century, possibly with parliaments deprived of their traditional authority and public decision-making processes largely unaccountable. This paper discusses and challenges these concerns by suggesting a theoretical framework under which algorithmic decision-making is compatible with democracy and, most relevantly, can offer a viable solution to counter the rise of populist rhetoric in the governance arena. Such a framework is based on three pillars: (1) understanding the civic issues that are subjected to automated decision-making; (2) controlling the issues that are assigned to AI; and (3) evaluating and challenging the outputs of algorithmic decision-making.

1. *Introduction*

Disco Sour is a dystopian novel published in 2017 in which the author imagines a postnation-state world where elections are replaced by a Tinder-like app called Plebiscitum[1]. The book proved successful enough to inspire a cross-party conference at the European Parliament on "Democracy in the Age of Algorithms"[2] to discuss the ethical issues and threats that artificial intelligence (AI) can pose to democracy. The novel (and, more relevantly, the high-level conference that followed) is but one of many voices that have raised alarms in the last few years over the looming end of democracy at the hands of digital technologies and automation.

Thanks to the ever-growing availability of data and the increasing power of decision-making algorithms, the future of political institutions is unlikely to be anything similar to what we have known throughout

* This article was first published in *European Journal of Risk Regulation*, 2022, p. 1.

[1] G. Porcaro, *Disco Sour*, London, 2017.

[2] European Parliament, *Democracy in the Age of Algorithms*, Nov. 7, 2017, Brussels, <https://www.internetforum.eu/events/657-scaling-the-sharing-economy>.

the last century, possibly with parliaments handing over their traditional authority to largely unaccountable technologies. Critics have described this as a "toxic cocktail for democracy"[3]. This paper is framed as a (provocative but constructive) response to the concerns prevalent among academic circles and public opinion. We respond to these fears with a classic trope of dystopian literature: what if all is not lost?

In the same vein, the paper aims to make the case for harnessing the potential of digital technologies to increase the quality of democracy in times of rampant populism. In fact, we suggest that the current debate over the use of AI in the public sphere needs to be reframed against the backdrop of rampant populism, as opposed to an idealised concept of democracy.

The main focus of our discussion is on technology-enabled policymaking mechanisms and their potential to positively affect democratic representation and legitimisation. The theoretical background to our analysis involves democratic theories that reject the simplistic identification of democracy with the manifestation of popular will. This strand of literature characterises democracy as being linked to a certain level of protection of participatory rights that is functional for the realisation of the democratic principle. Such an approach directly challenges the rhetoric of populism that emphasises disproportionately the representative element at the expense of a more articulated notion of democracy as a regime that is supposed to protect individual and collective rights.

The paper suggests that algorithmic decision-making can contribute to an output-oriented democratic process in which fundamental rights can regain centrality – one that can be challenged on substantive grounds by voters who are self-aware and have experience with technology. The key shift that an AI government may have in store for us is the elimination of the need to cater to voters' often irrational and detrimental concerns from the process of policymaking. This shift is likely to be a positive one.

While not discounting existing concerns, the paper discusses the merits of algorithmic decision-making in the public sphere from the perspective of democratic theories, as elaborated by Robert Dahl, Charles Beitz and Fritz Scharpf, and it offers a change of perspective from the currently dominant narratives. Instead of accepting the narrow view of technology as a binary alternative to traditional representative democracy, where participation is mainly channeled through periodic voting, we look at the

[3] C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York, Crown Books, 2016.

current dynamic through the prism of the protection of fundamental rights and output legitimacy.

With a view to supporting our claim, the paper proceeds as follows: it will firstly engage with the concept of populism with a view to exposing the incomplete (and partially distorted) concept of democracy it purports to represent (Section II); it goes on to propose an alternative reading of the relationship between algorithms and democracy, which relies on the notion of output legitimacy and its fundamental rights implications (Section III). The article then builds an agenda for an algorithmic decision-making framework based on a methodological framework for the analysis of the relationship between algorithms and democracy (Section IV). The framework is centred on three main pillars: (1) understanding civic issues that are subjected to automated decision-making; (2) controlling issues that are assigned to AI; and (3) evaluating and challenging the outputs of algorithmic decision-making (Sections IV.1, IV.2 and IV.3). Finally, the paper offers some concluding remarks on the way in which algorithmic decision-making can be framed to be (1) compatible with democracy and (2) an effective alternative to populist rhetoric.

2. *The narrative of "the people"*

The rise of populism has made the headlines and has been a topic of discussion in academic and public contexts in recent years. Yet, despite its renewed popularity, the concept itself is far from new. Among early commentators from the midtwentieth century, the notion of populism emerged as markedly "characterized by a peculiar negativism [and] great doses of blind hatred"[4] towards multiple facets of society. Isaiah Berlin was among the first to identify a distinct element common to all types of populism, namely the centrality of the notion of the people, invariably characterised as the have-nots, in stark opposition to an enemy – a dominant group said to have caused them damage, usually identified as the capitalists, the bureaucracy or some other group identified on the basis of an ethnic, racial or national otherness.[5] Along the same lines, Hugh Seton-Watson described populism as the "idolization and worship of the people :

[4] I. Berlin et al., *To Define Populism*, in *Government and Opposition*, vol. 3, 137, 1968, p. 169.
[5] *Ibid*, at 175.

: : contrasted with the vices of the elite"[6]. Other observers have pointed out the emphasis on the popular will as the ultimate source of legitimisation for political authority. Lloyd Fallers, most influentially, suggested that both populistic and nationalistic ideologies share this same trait[7]. Even with the lack of a generally agreed upon definition, the relevant literature has been in agreement in seeing populism as a pathological feature of democracy – a "syndrome" rather than a specific doctrine[8].

In keeping with this tradition, recent works have almost universally framed populism in negative terms. Yet identifying precisely its practical impact has proven less than straightforward. Paul Taggart discusses current forms of "new" populism as the "rejection of the political agenda, institutions, and legitimacy of the modern welfare state model of mixed economy capitalism"[9]. David Landau has expressed a largely similar concern while also providing an explanation of the progressive unfolding of the detrimental effects of populistic institutions across three consecutive phases, such as "undermining the existing institutional order, constructing a new order built on a critical vision of the old one, and consolidating power in the hands of populist leaders"[10]. The dynamic between constitutional democracy and populism has been described as a form of "parasitism" where the latter exploits certain features and fragilities of the former to tip the balance between the rule of majority and the rule of law on which constitutional democracy is traditionally founded[11]. From this perspective, the practical threat that populism poses to democracy is understood as a toxic rhetoric that, over time, manipulates both of these pillars by proposing a procedural vision of democracy that disproportionally prioritises the rule of the majority, in pursuit of aims incompatible with the spirit of democracy and its values, such as political pluralism, transnational solidarity and the protection of minority rights[12].

Crucially, the prioritisation of the rule of majority is made possible

[6] *Ibid*, 156.
[7] L. Fallers, *Populism and Nationalism*, in *Comparative Studies in Soc'y and History*, vol. 6, 445, 1964, p. 447.
[8] P. Wiles, *A Syndrome, Not a Doctrine: Some Elementary Theses on Populism*, in G. Ionescu and E. Gellner (eds), *Populism: Its Meanings and National Characteristics,* London, 1970, p 166.
[9] P. Taggart, *Populism*, New Delhi, Viva Books, 2002, p 75.
[10] D. Landau, *Populist Constitutions*, in *University of Chicago L. Rev*. vol. 85, 521, 2018, 523.
[11] T. Fournier, *From Rhetoric to Action, a Constitutional Analysis of Populism*, in *German Law Journal*, vol. 20, 362, 2019, p. 364.
[12] *Ibid*, 381.

by the vision, intrinsic to populistic rhetoric, of the majority as the sole legitimate ruling entity[13].13 Oran Doyle elaborates on this concept by offering a convincing account of populism as a form of constituent power that, while delegitimising non-majoritarian institutions (such as those that typically provide checks and balances), twists the constitutional system towards serving the needs of a specific part of the population[14].

The most immediate effect of populist rhetoric is thus the de-legitimisation of certain traditional aspects of democratic procedures, most typically all of those that provide some forms of checks and balances; and this, in turn, unleashes a further consequence on the quality of democratic governance and policymaking. Populism's first victims are usually the procedural mechanisms typical of parliamentary politics, such as multiparty competition and democratic checks and balances that, in the view of populist rhetoric, "illegitimately restrict the will of the people by checking the power of the majority and empowering minorities" in a way that time and again throughout history has led to "the crisis of parliamentary institutions"[15]. The proceduralist conception of political legitimacy traditionally places a strong emphasis on parliamentary deliberation and constitutional-level decision-making rules. But rising populist movements have attacked the role of institutions underpinned by those procedures – traditional parties in the first place, as well as all other institutions normally overseeing governments' operations – and, lastly, are undermining procedural legitimacy.

However, by disrupting the legitimate standing of those bodies whose specific function is to counter the view of hegemonic majorities, to contribute the perspectives of diverse groups and to provide mechanisms for the oversight of governments' decisions, populist stances cause practical and direct consequences at the governance level. Empirical studies have confirmed that populist political forces, when they are in government, typically cause a decline in the enjoyment of civil liberties and political rights[16], lead to significant surges in corruption[17] and tend to perform proportionately less effectively than non-populist governments in domains

[13] *Ibid*, 380.
[14] O. Doyle, *Populist Constitutionalism and Constituent Power*, in *German Law Journal*, vol. 20, 161, 2019.
[15] M.P. Saffon and N. Urbinati, *Procedural Democracy, the Bulwark of Equal Liberty*, in *Political Theory*, vol. 41, 441, 2013, p.454.
[16] J. Kyle and Y. Mounk, *The Populist Harm to Democracy: An Empirical Assessment*, Tony Blair Institute for Global Change, 2018, p 18.
[17] *Ibid*, 19.

such as basic welfare, gender equality and local democracy[18].

The complexity of this dynamic can be tentatively summarised as follows: while populism primarily and most directly affects and undermines the procedural legitimacy of democratic decision-making, its most practical and perceivable impact is, instead, on the quality of policy decisions. Both the inputs (procedures) and outputs (actual policy decisions) of democratic governance are negatively affected by populism; however, solutions proposed so far have struggled to capture this ambivalence, as they largely focus on correcting the quality of outputs alone, disregarding the centrality of procedures and their perceived legitimacy in leading to such outcomes. In fact, a common reaction all across Europe to the surge of populist movements has been a renewed appeal for technocratic governments and their promise of better governance[19]; meanwhile, resorting to algorithmic decision-making is often decried on the grounds of its innate inability to achieve fair and non-discriminatory results. In both cases, the acceptance of technocratic governance and the rejection of AI-assisted solutions are based on a similar inability to reconcile procedural and substantive elements of democracy. We suggest instead that, under certain circumstances, AI can be deployed as a useful tool to increase the quality of democratic decisions – not in and of itself, but by enabling new mechanisms for democratic procedures to regain their legitimacy.

3. *The counterclaim: the notions of input and output legitimacy and their implications for algorithmic decision-making*

Democratic legitimacy is a potentially elusive locution, which deserves to be clarified for the present purposes. It can be defined as the principles and procedures through which collective decisions are accepted and deemed to be binding by those who have not directly participated in making them[20]. The locution thus refers to that particular kind of

[18] A. Silva-Leander, *Populist Government and Democracy: An Impact Assessment Using the Global State of Democracy Indice*s, in *Global State of Democracy in Focus*, no. 9, 2020, p 7.
[19] See H. Kudnani, *Technocracy and Populism After the Coronavirus*, in R. Youngs (ed.), *How the Coronavirus Tests European Democracy*, Brussels, Carnegie Europe, 2020.
[20] P. Rosanvallon, *Democratic Legitimacy*, NJ, Princeton University Press, 2011. For a European perspective, see S. Bartolini, *The Nature of the EU Legitimacy Crisis and Institutional Constraints: Defining the Conditions for Politicisation and Partisanship*, in O. Cramme (ed.), *Rescuing the European Project: EU Legitimacy, Governance and Security*,

justification for the exercise of power under a democratic framework that is normative in the sense that it is objectively based on normative principles and procedures and informed by political theory.

Towards the end of the twentieth century, a cluster of different phenomena led scholars in law and political science to question the meaning of democratic legitimacy[21]. First of all, the progressive particularisation of governmental functions, which went hand in hand with the emergence of new rights, determined the articulation of the administrative apparatus in a series of agencies and bodies taking the responsibility for important policy decisions[22]. Second, the ideological clash between capitalism and socialism led some scholars to explore different models of legitimisation of political institutions by focusing on the interplay between the terms of political participation and the desirability of certain political results[23]. Third, judicial activism in the field of fundamental rights urged intellectuals to address the issue of democratic legitimacy by advancing the need to frame democracy in light of a "strong principle of equality"[24]. Finally, the establishment of supranational levels of government, in Europe and globally, forced scholars to find a conceptual frame for the exercise of public powers beyond state borders, which happened via centralised decision-making processes, only marginally involving national democratic circuits[25].

Thus, the literature on democratic legitimacy has grown mainly with the goal of deepening the understanding of democracy in the context of multi-layered and complex decision-making processes within which democratic deliberation is one step in the articulated procedures of political actions.

The main contribution of those doctrinal studies was the identification of different kinds of democratic legitimisation. Scholars such as Joseph Weiler went as far as proposing a typology within which it is possible to

London, Policy Network, 2009, p 57.

[21] J.H.H. Weiler, *The Transformation of Europe*, in *Yale L. J.*, vol. 100, 1991, p. 2403.

[22] See S. Rose-Ackerman, *American Administrative Law Under Siege: Is Germany a Model?*, in *Harvard L. Rev.*, vol. 107, 1994, p. 1279, arguing that "democracies need to strike a balance between popular control and expertise"; and R.H. Pildes and C.R. Sunstein, *Reinventing the Regulatory State*, in *University of Chicago L. Rev.* vol. 62, 1, 1995, p. 3. The authors delve into the challenges that the role of expertise poses to political decision-making.

[23] C.R. Beitz, *Political Equality: An Essay in Democratic Theory*. Princeton, NJ, Princeton University Press, 1990.

[24] R. Dahl, *Democracy and Its Critics*, New Haven, CT, Yale University Press, 1991.

[25] J. Lodge, *Transparency and Democratic Legitimacy*, in *Journal of Common Market Studies*, vol. 32, 343, 1994.

distinguish between: (1) input or process legitimacy; (2) output or results legitimacy; and (3) legitimacy based on telos, which means that legitimacy is gained by promising a desirable result[26]. Weiler's classification builds on the work of prominent political philosophers who engaged in defending democracy. Robert Dahl, for example, reflected on democracy by focusing on three components: voting, political participation and understanding of civic issues. According to Dahl, individuals subjected to collective decisions should be able to have their interests equally taken into consideration, as well as to gain control over the matters that reach the decision-making agenda[27]. Therefore, Dahl built his democratic theory upon a strong understanding of the principle of political equality, which is functional for the protection of a complex set of collective and individual social, cultural and economic interests[28]. Within Dahl's democratic theory, the essential core of democratic legitimisation lies in its ability to reflect the popular will in a meaningful way, such as through means of effective control over (1) the selection of political options and (2) political decisions.

Charles Beitz started by acknowledging that political equality is the main challenge for any model of democratic legitimacy. At the same time, he warned that the output ("result" in his words) of democratic processes should not be underestimated. In his conception of democracy, institutions gain legitimacy when they maximise the expected values of an independently specified social welfare function. In this context, only fair terms of political participation are likely to produce the most desirable results. According to Beitz, though, this understanding of democratic legitimacy is not equivalent to an outcome-oriented theory because the logic of the maximisation of expected values works within the framework of a social function that is concerned about alternative political outcomes, which essentially means that the social function is drawn by political preferences and, therefore, by inputs[29].

Almost ten years after Dahl and Beitz, Fritz Scharpf identified two frames of input and output legitimacy by conflating the normative and social legitimacy of a given political regime[30]. Input legitimacy refers to the bottom-up process through which the people make political choices concerning how they want to be governed. Within the "input frame",

[26] Weiler, *supra*, note 21, at 2405.
[27] Dahl, *supra*, note 24, at 322.
[28] *Ibid*, 92.
[29] Beitz, *supra*, note 23.
[30] F. Scharpf, *Governing Europe: Effective and Democratic*, Oxford University Press 1999, pp. 11–12.

political choices are legitimate to the extent that they reflect the will of the people. The latter, in turn, expresses a collective selfdetermination of preferences that are expected to be addressed by the representatives.

Input legitimacy resorts to a kind of consensus rhetoric whereby the people define preferences on the basis of a minimum agreement on some values. It comes as no surprise, then, that populist movements escalated the input legitimacy rhetoric by drawing a direct connection between the will of the people and the accomplishment of an authentic democratic model of government. In other words, according to populist rhetoric, democracy exists insofar as it consists of the realisation of people's determinations without the need for further political assessment or appreciation[31].

On the contrary, under an output-oriented model, legitimacy concerns are addressed by focusing on the effective promotion of a constituency's common welfare through a number of political actions designed to solve problems of a collective nature. The output perspective takes into consideration a political environment of articulated needs and preferences in which shared values do not necessarily express a common identity that is translated into a unitary political will. To that extent, an output legitimacy frame suits the reality of pluralist societies where decision-making processes are located at more than one level of government. The output perspective engages with the substantive content of democracy rather than with its procedural meaning[32].

Against this backdrop, focusing on outputs means considering democracy as a political regime in which a given set of individual and collective needs are recognised in the form of rights with a view to promoting social welfare and ensuring peaceful coexistence. By contrast, the input frame is more concerned with fulfilling democracy from the standpoint of democratic deliberation (majority rule), while it does not address the effectiveness of democratic values in a given constituency.

[31] Fournier, *supra*, note 11.

[32] Democracy is an ambiguous term: it can refer to majority rule or it can capture a broader meaning whereby a political regime can be identified as democratic only if a set of fundamental rights is guaranteed. The conceptualisation of a substantive meaning of democracy has been offered by T. M. Franck in *The Emerging Right to Democratic Governance*, in *American J.l of International L.*, vol. 86, 46–91, 1992. Constitutional theory is also concerned with the substantive meaning of democracy and generally reluctant to identify the latter merely with a procedural rule. Under a constitutional model, democracy is framed by substantive constraints deriving from common shared values enshrined in constitutional texts, which in turn define the boundary of legitimate democratic choices: see B. Ackerman, *We the People: Foundations*, Cambridge, MA, Harvard University Press, 1991, pp. 3–33.

Output legitimacy, however, does not necessarily rule out the bottom-up element of democracy. In looking at ways to reconcile the quality of outputs with the a most influential manner, emphasised the requirement that political decisions produce rational outcomes[33]. Different authors have discussed the requirement of rational outcomes from different perspectives, some looking at the consistency of preferences and choices, others at the reasoning provided during deliberation. Under a model of legitimacy defined as "rational deliberative proceduralism", requirements include the political equality of all of the individuals comprised in the polity and called to participate in the political process and the rationally justified outcomes of collective decision-making[34].

Input and output legitimacy tend to coexist in mature democracies. Through an input framework, needs and priorities are identified, legitimising the pursuit of some goals; in parallel, outputs measure the results and legitimise political choices that may not represent the specific content of democratic deliberation but nevertheless effectively achieve those desired goals.

From this perspective, algorithmic decision-making can be seen as an instrument of output legitimacy when looking at algorithms through the lenses of democratic theories stressing the importance of individuals' chances of understanding and assessing political choices. The perspective we suggest for this, however, is different from merely resorting to technocratic governance; in fact, as we discuss below, neither algorithms, which would certainly not be capable of ensuring "better" output per se, nor technocracy as a model of governance – whether AI- or human-led – fully address the issue of the de-legitimisation of democratic procedures that lies at the root of the problem.

[33] K. Arrow, *Social Choice and Individual Values*, New York, 1951.
[34] F. Peter, *Democratic Legitimacy and Proceduralist Social Epistemology*, in *Politics, Philosophy & Economics*, vol. 6, 329, 2007, pp. 335–37.

4. *Algorithms and democratic legitimization: a framework for analysis*

A compelling interpretation of populism and technocracy suggests that the two tendencies are fundamentally complementary to one another[35]. The complementarity resides in the idea that both populism and technocracy can be understood as reactions to the system of government by political parties and its provision of political mediation and the procedural conception of political legitimacy to democratic governance. In this sense, populism and technocracy share a similar approach in that they both constrain and eventually void political antagonism, whereas procedural democracy offers a system to channel it. From this perspective, both populism and technocracy undermine two key features of parliamentary democracy, which are political mediation among different social groups and a procedural conception of political legitimacy.

If technocratic government seems unfit to offer a valid solution to the issue of delegitimised democratic procedures, so do algorithms if they are expected to reach substantive political decisions independently. In fact, the application of AI and algorithms to the public sector has been framed in slightly different terms by different regional and international institutions in recent years. The European Union (EU) defines AI as "systems that display intelligent behaviour by analysing their environment and taking action – with some degree of autonomy – to achieve specific goals"[36]; the Organisation for Economic Co-operation and Development (OECD) defines AI as "machine-based systems that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments"[37]; and the United Nations Educational, Scientific and Cultural Organization (UNESCO) defines AI as "technological systems which have the capacity to process information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control"[38]. Even with the lack of a generally agreed upon definition, all of

[35] C. Bickerton and C.I. Accetti, *Populism and Technocracy: Opposites or Complements?*, in *Critical Review of International Social and Political Philosophy* , vol. 20, 2017, p. 186.
[36] EU Commission, Communication From the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee of the Regions on *"Coordinated plan on Artificial Intelligence"*, COM(2018) 795 final, 7.12.2018.
[37] OECD, *Recommendation of the Council on Artificial Intelligence*, C/MIN(2019)3/FINAL adopted on 22 May 2019.
[38] UNESCO, *First draft of the Recommendation on the Ethics of Artificial Intelligence*,

these formulations focus on some common features, such as the ability of these technologies to process information and perform tasks intelligently. Interactions between public bodies and AI can take several forms, including financing and developing technologies, harnessing the data to feed into the algorithms, imposing regulatory standards and, finally, utilizing the technology to provide a range of services[39]. We are particularly interested in the intersection between these last two functions – more specifically, how algorithms can help boost the democratic legitimisation of the public bodies that utilise them.

In the context of public decision-making, algorithms can function in several ways. Let us attempt to provide a non-exhaustive classification by looking at the steps of policymaking and decision-making procedures. AI can: (1) represent the world by, for example, proxying demographic data; (2) predict or test the desirability of a given course of action in light of the results that the latter may determine, such as immigration detention risk assessment[40]; (3) reach a decision, generally on the assumption of the inherent reliability of the result and/or efficiency of the process, by, for example, selecting individuals who will benefit from an allocation decision that had been made within the traditional political process[41]; and (4) act as an algorithm-manager by supervising and controlling public servants who are required to make complex decisions[42]. In each of these scenarios, algorithms play different roles. Representation and prediction functions, described under (1) and (2), imply that algorithms can provide decision-

SHS/BIO/AHEGAI/2020/4REV.2, 7 September 2020.

[39] B. Ubaldi et al, *State of the Art in the Use of Emerging Technologies in the Public Sector*, OECD Working Papers on Public Governance No. 31, 2019.

[40] Let us imagine a situation in which an algorithm is used to predict the likelihood that a migrant would face torture or inhuman and degrading treatment if they are sent back to their country of origin. In this particular case, algorithm settings will probably consider the safety of the country of origin on the basis of a number of features whose consistency and accuracy may be open to question. See M. Nofferi and R. Koulish, *The Immigration Detention Risk Assessment*, in *Georgetown Immigration Law J.l* vol. 29, 45, 2014.

[41] This is the case for using an algorithm in the context of welfare allocation decisions. There are a number of such examples: see *infra*, note 54.

[42] Let us take as an example the use of algorithms in the context of judicial proceedings. The Council of Europe addressed the potentialities of the use of AI in judicial decisions: see *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*, European Commission for the Efficiency of Justice, Strasbourg, 3–4 December 2018. See also M. Bovens and S. Zouridis, *From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control*, in *Public Administration* Review, vol. 62, 174, 2002.

makers with accurate information concerning a given political option. Therefore, algorithms do not replace political choices, but rather they create the conditions for a political choice to be confronted with concrete outputs. The selection function mentioned under (3) uses algorithms to speed up procedures that would otherwise require careful and lengthy examination. Algorithms can thus ensure the efficiency of the selective process and the consistency of results. In the scenarios depicted under (3) and (4), the algorithm makes decisions instead of the decision maker identified via the ordinary political process. Computer science is used to replace political deliberation with possible benefits in terms of the efficiency of a democratic system.

It is worth noting that none of these scenarios envisage algorithms independently defining preferred outcomes of political decisions, which would completely exclude any grounds for democratic procedural legitimacy; instead, we focus on the opportunities they bring for disentangling political legitimisation from the merely formulaic input-centric aspect of the will of the majority. Algorithms offer the opportunity to refocus on output legitimacy by connecting inputs and outputs and making such connections rationally appraisable.

From this perspective, a notable element of the rational deliberative procedural model is that, when the rationality test is applied to the outcomes in order to assess their consistency with the premises of any political decision (defined as "the reasons given during deliberation in favour of or against certain alternatives")[43], it may be the case that the rational outcome(s) is different from what voters would have immediately chosen if asked directly. In this way, rational deliberative proceduralism separates legitimacy from input-centrism by offering alternative procedural grounds.

In light of this theoretical model, we suggest that algorithms offer a viable way to implement both of these conditions of (rational deliberative procedural) democratic legitimacy by offering the possibility of equal democratic participation and a system by which to rationally assess the consistency between premises and outcomes.

In translating the model from theoretical to more practical perspectives, the impact of algorithmic decision-making on democratic governance and legitimacy can be discussed with respect to its contribution to policymaking, and particularly through the prism of consolidated findings in policy studies where, over the last few decades (largely following seminal works through the 1960s to the 1980s), a direct influence of

[43] Dahl, *supra*, note 24, at 335.

the content of public policy decisions on the character of democracy has become clear. More specifically, distinct features of the policy cycle – such as the framing of issues, the construction of targets, the structure of implementation and delivery systems and the transparency of the whole process – impact foundational elements of democratic life, such as citizens' trust in the government's ability to solve public problems, public support for the government's action and public accountability. Helen Ingram and Anne Schneider have helpfully suggested a framework through which to understand and evaluate the effectiveness of public policies in contributing to democracy: by building on core literature in the field, they suggested that policies contribute to democratic governance inasmuch as they allow it to expand its franchise, scope and authenticity of democratic governance[44]. These three domains respectively refer to the number of participants, the number of domains of life directly and legitimately affected by political decisions and the possibility for citizens to exercise "substantive, informed and competency-engaged" democratic control. The three domains exist in delicate balance with one another, as, for instance, expanding the scope of democratic governance could come at the cost of superficial deliberation, thus undermining authenticity. In practical terms, the first two domains refer to the processes of the identification of issues that belong in the political discourse and the terms of the debate regarding which actions public authorities would be expected to take (the authors exemplify this dynamic by describing the switch in the discourse over water policy in the USA from water as a public good to the conceptualisation of water as a privatised commodity, and criminal policies built around the idea of crime as a violation against an individual as opposed to an offence against the society or state). The terms and the results of such definitory processes depend naturally on how open and inclusive these are with respect to their ability to include those who are directly affected as well as those with a specialised knowledge in the field. The process of identifying the issues leads to two relevant effects: by defining any societal issues as a question of political relevance, the process identifies public authorities as competent decision-makers; and at the same time, by "pitching" the debate at a more or less specialised level, the process inevitably defines its boundaries by making it more or less accessible to ordinary citizens.

A crucial aspect of this consists in the creation of open public forums where citizens can discuss policy problems openly and directly. The way

[44] H. Ingram and A. L. Schneider, *Policy Analysis for Democracy*, in R.E. Goodin, M. Moran and M. Rein (eds), *The Oxford Handbook of Public Policy*, Oxford, Oxford University Press, 2008, p 169.

a policy is framed and designed is directly responsible for the appearance of such forums. The process of policymaking, in other words, needs to be framed in a way that facilitates the emergence of such opportunities for civic discussion.

The model, as explained in those terms, does not fail to consider a substantive change in how accountability ought to be construed in the current context of the decentralisation of political power and the increasing distances between citizenries and those in government, which, in turn, calls for more direct citizen involvement in holding governments accountable and, crucially, a different approach to assessing the specificities of public governance by evaluating aspects that are not limited to mere effectiveness and efficiency. Ingram and Schneider suggest that government action instead be measured by "its ability to intervene strategically in the complex networks of policy delivery systems to encourage better access to information, to correct for power imbalances and damaging stereotypes and social constructions among stakeholders, and to create arenas and spheres of public discourse"[45].

It is therefore useful to examine algorithmic decision-making by carefully disentangling the cluster of issues surrounding it through the lenses of democratic theories. By unpacking the logic of democratic legitimacy and still adhering to the input/output framing, it is possible to identify at least three different, though related, cruxes: (1) the problem of understanding and selecting civic issues that deserve to be addressed by political institutions; (2) the problem of controlling which issues reach the democratic institutions; and (3) the problem of evaluating and challenging the results of a given course of political action. The inherent complexity of political processes makes these three aspects particularly sensitive with respect to algorithmic decision-making. Algorithms work on the basis of instructions related to data processing; the selection and the organisation of data derive from human input or they can be learnt by the computational machine itself according to a logic that is completely artificial.

Therefore, tackling the three aforementioned problems in the context of algorithmic decision-making involves different consecutive steps. The first issue to address is determining the extent to which the instruction algorithms reflect political preferences that have legitimately reached democratic institutions. This issue can also be framed as a problem of understanding the "democratic soundness" of algorithmic decision-making by clarifying how computers process data or even how they learn to select

[45] *Ibid*, at 184.

and process data. The understandability of the process and the selection of data are key to ensuring that participation is broad enough to let democracy expand its franchise.

The second issue is the ability of the political community to control the algorithmic decision-making process by deciding which choices can be dealt with by AI. Control over the algorithm here means that the political community has the opportunity to select the issues that are allocated to AI and, therefore, to own the scope of democratic governance.

Finally, the third issue is the opportunity to challenge algorithmic decision-making, understood as the opportunity to assess, question and potentially change the outcome of any given non-human decision.

With a view to applying the output legitimacy frame to algorithmic decisionmaking and to test its impact on democratic processes, the paper next explores these three different dimensions of the interplay between algorithmic decision-making and democracy.

4.1. *Understanding civic issues under an algorithmic decision-making framework*

For a democratic process to be authentically based on political freedom, citizens should be able to understand civic issues; that is to say that the citizenry should, with a reasonable intellectual effort, understand the justification for decisions taken collectively. Dahl identified understanding civic issues as one of the conditions under which democracy can properly function as a reliable means for protecting and promoting the goals of persons that are subjected to collective decisions[46]. To be more precise, according to Dahl, the opportunity to understand matters that reach the decision-making agenda enables people to properly enjoy political equality, which means to have their interests equally addressed by legislators. In democratic theory, this need is commonly expressed by an emphasis on education and participatory rights. In such a context, understanding means being able to have a sufficiently clear idea of the functioning of the institutions as well as (at least) a basic understanding of the language of the public discourse[47]. Understanding is essential to building trust in political institutions; without it, people are left with no clear point of reference regarding the reliability of democratic processes.

[46] Dahl, *supra* note 24, at 322.

[47] See R. Dworkin, *What Is Equality – Part 4: Political Equality*, in *University of San Francisco L. Rev.*, vol. 22, 1, 19871.

When it comes to algorithmic decision-making, understanding civic issues that are processed through algorithms requires a second level of comprehension concerning the peculiar mechanism of AI, especially when it is performed by machine learning and with limited involvement from human intelligence. This is to say that for any algorithmic decision-making process to be compatible with democracy it should be coupled with the guarantee of a minimum level of technological education, as the EU has recently recognised[48]. Algorithmic decision-making therefore needs to be grounded in a "cultural cognition" of what an AI decision-making process is about.

By "cultural cognition" we refer here to the shared view or frame of the world and/or of a given portion of reality of a group of people[49]. Such a concept builds on Mireille Hildebrandt's idea of enabling citizens to counter-profile AI. According to Hildebrandt, if citizens are equipped with proper intellectual instruments by which to understand the AI ecosystem, then they may be able to challenge the modelling that they have been represented with or even to question the set of instructions used by algorithms[50]. Counter-profiling requires citizens to engage actively with AI, an attitude that can be stimulated by spreading a cultural cognition of AI decision-making.

By assuming this perspective, we can imagine a situation in which citizens may approach algorithm decisions in order to make sense of the unfolding of the political decision-making process.

Let us consider the example of immigration. AI can be employed across the whole spectrum of procedures related to migration, from the screening of applicants for asylum protection to the risk assessment decisions related to individual migrants who may pose a threat to national security. Examples include the USA and Canada, where governments considered using algorithms developed through the analysis of wide swaths of data to realise trend studies and to make predictions as to the influx of migrants in a particular context[51]. Such use comes with many concerns, including risks

[48] See, for example, the European Commission Strategy on Artificial Intelligence, which states that modernisation of education in light of the impacts of AI on social dynamics should be a priority for governments: see COM(2018) 237 final, at 4 and 12.

[49] The concept of "cultural cognition" is borrowed from J. K. Sax, *The Problems with Decision-Making*, in *Tulsa L. Rev*., vol. 56, 39, 2020.

[50] M. Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*, Cheltenham, Edward Elgar Publishing, 2015, pp. 100–03.

[51] In Canada, this is the result of a broader strategy to implement ethically aware artificial intelligence decision-making processes: see the Pan Canadian Artificial Intelligence

of human rights violations and forms of direct or indirect discrimination deriving from implicit biases in algorithm setting or simply from the absence of reasonable assessments of results[52]. Those risks can be reduced by increasing the sophistication of the instructions given to the machine or through a careful human check on biases. These issues, though important, are not the focus of our attention here. Instead, we are attempting to look at the problem from a different angle. Let us imagine using an algorithm, or even a machine learning algorithm, to predict the likelihood that a migrant would face torture or inhuman and degrading treatment if they are sent back to their country of origin. Let us also imagine that the algorithm's settings consider the safety of the country of origin on the basis of a number of features derived from decisions made by competent authorities in the past. What the algorithm is doing in this situation is predicting how many people, given the instructions on which it operates, will be denied entry, eventually by also classifying different levels of risks of inhuman or degrading treatment.

A decision ordering expulsion when the risk of inhuman or degrading treatment exists can have devastating effects on the right to life of the individual concerned if it is made based on incorrect assumptions. As such, there can be reasonable agreement on the fact that this is a technical issue that impacts the general public and also signals a state's commitment to human rights protection. It comes as no surprise, then, that scholars urge governments to avoid using AI in such a context on the basis of the assumption that relying on the empathetic assessment of human beings can produce fairer results[53]. Even when the emphasis is not strongly on empathetic assessment but on implicit biases, such as automatic decisions on the level of risk that an individual migrant may pose to national security, scholars insist that AI does not tailor decisions on well-assessed grounds[54].

Strategy, available at <https://policyoptions.irpp.org/magazines/august-2018/responsibly-deploying-ai-in-the-immigration-process/>.

[52] See D. Robinson & K. Vold, *Immigration Decisions Are Complex with High Stakes for the People Involved. The Government Must Tread Carefully on Using AI in the Screening Process*, available at <https://policyoptions.irpp.org/magazines/august-2018/responsibly-deploying-ai-in-the-immigrationprocess/>.

[53] Nofferi & Koulish, *supra* note 40.

[54] When looking closely at both arguments, though, a fundamental logical flaw emerges: any decision based on some kind of generally applicable criteria, as decisions based on laws generally are, faces the risk of being insufficiently tailored to a peculiar case and thus proving unfair with respect to an individual situation while still pursuing an objectively fair result on the whole. Therefore, the problem is not a problem of AI, but rather one of the general or blind application of non-empathetic intelligence, which may lack

In such a context, it is possible to imagine legal solutions ranging from the right to challenge the automatic decision, as codified in Article 22 of the GDPR 8[55], or the duty of double-checking on the part of public authorities. In both cases, a second step involving human intelligence can correct the unfair, unreasonable results of an algorithm or simply place a second deliberation before the final decision so that the decision-making process is still not determined by an algorithm.

From an alternative viewpoint, algorithms such as the one used in the example above can perform an informative function by making people aware of the actual numbers of individuals risking death or torture because of expulsion decisions made under a certain set of criteria reflecting a given political choice. Moreover, people may be able to frame those pieces of information against the backdrop of broader national or supranational policies in order to better understand their implications (and eventually to better understand their preferences or priorities).

Along these same lines, algorithms can enhance the understanding of the real dimensions of a problem and expose populist rhetoric by confronting it with facts and data. This is a kind of "use" of algorithms that, when properly accessible and transparent, equips voters with fact-checking instruments concerning policies and their effectiveness.

The relationship between data processing and participation can be further explored by addressing the ways in which data processed by algorithms are fed into the political decision-making process to generate informed political choices.

A key contribution in the field of data science is that extremely fragmented, individual data collected by technologies are often insignificant if taken by themselves. At a later stage, technologies re-aggregate data in the process of generating outputs and "fold [the data] back into the experience of everyday life". This portion of the process offers an opportunity to harness data within governance procedures in a way that can effectively contribute to a more meaningful and thorough understanding of civic issues by decision-makers[56].

reasonableness or be fundamentally biased. See J. Dickinson, *Legal Rules: Their Function in the Process of Decision*, in *University of Pennsylvania L. Rev.*, vol. 79, 835, 1931.

[55] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[56] N. Couldry & A. Powell, *Big Data from the Bottom Up*, in *Big Data & Society*, vol. 1, 2014, pp. 1–5.

Examples of such processes have been identified and discussed with respect to specific policy sectors, such as environmental policies. It has been discussed how citizens' participation in the data-led delivery of public services within the context of smart cities allows local governments to tune in to the everyday experiences of citizens and their use of space and urban infrastructure. Citizens' sharing of individual data performs a function that is substantively equivalent, for its contribution to democratic engagement, to participatory media, to the point of eventually becoming a "constitutive [practice] of citizenship"[57]. Even beyond the specific cases of smart cities and environmental policies, the positive impact of harnessing data in public decisionmaking processes has been identified as twofold; it enables public action to become better informed and more responsive to citizens' needs, which, in turn, places a specific responsibility on public decision-makers to develop legal and policy frameworks for the responsible sharing and processing of data to feed democratic processes[58].

From these doctrinal perspectives, data-sharing emerges as a practice that enables citizens to connect with public decision-making processes in a way that continues and complements the more traditional role of the news media to build and inform political agendas. From this premise, a further reflection can follow regarding the opportunity that such new modalities of citizenship offer to feeding the policymaking process with qualitative and reliable information that neither mainstream media[59] nor social media[60] seem capable of contributing in the current climate due to the policy and legal frameworks in which they operate[61]. By contrast, citizen data could break the cycle of biased, sensationalistic narratives cycled into the political agenda.

Whereas concerns regarding the impacts of the mass collection of individualized data on privacy and data protection rights have (quite rightfully) been at the centre of the academic literature, this alternative perspective speaks to the current rise of populism in a way that has often

[57] J. Gabrys, *Programming Environments: Environmentality and Citizen Sensing in the Smart City*, in *Environment and Planning D: Society and Space*, vol. 32, 2014, pp. 30–48.
[58] W. J. Mitchell & F. Casalegno, *Connected Sustainable Cities*, Cambridge, MA, MIT Mobile Experience Lab Publishing, 2008.
[59] G. Mazzoleni, *Populism and the Media*, in D. Albertazzi and D. McDonnell (eds), *Twenty-First Century Populism: The Spectre of Western European Democracy*, New York, Palgrave Macmillan, 2008, pp 49–64.
[60] T. Flew and P. Iosifidis, *Populism, Globalisation and Social Media*, in *International Communication Gazette*, vol. 82, 2020, p. 7–25.
[61] D. Freedman, *Populism and Media Policy Failure*, in *European Journal of Communication*, vol. 33, 2018, pp. 604–18.

been overshadowed by the more prominent conversation surrounding the risks connected to such practices. Without dismissing the valid concerns that the academic literature has unveiled and continues to discuss to date, re-centring the conversation in a way that acknowledges this potential opportunity can instead be key to making progress in countering the rise of populism.

For data to feed into the political process, they need to be discernible from those in charge of shaping public policies. This idea is based on the observation that social analytics provide information and can perform a role in the policy cycle akin to the role traditionally played by content in the media sphere. But whereas media content can be understood and interpreted semantically, data convey information that can be made sense of in different ways[62]. Algorithms can provide decision-makers with several alternative outputs that depend on how a given set of instructions elaborates data. From a democratic standpoint, then, concerns are raised by the reliability of the particular elaboration or modelling of the available dataset. The literature has already addressed this problem in the context of automated data processing, with a view to controlling and directing machine learning processes. In particular, Hildebrandt advocates for "agonistic machine learning", which she identifies as a requirement that "companies or governments that base decisions on machine learning must explore and enable alternative ways of datafying and modelling the same event, person or action"[63]. Such an agonistic frame would provide decision-makers with several accounts of processed data, which, in turn, would enable them to detect biases or incorrect assumptions in machine learning processes. An agonistic approach, however, benefits algorithmic decision-making more generally. In fact, it helps public institutions make sense of data by confronting multiple models or alternative readings of the reality. Public institutions can then couple this piece of information with citizens' inputs to make informed choices that are based on a thorough understanding of civic issues.

4.2. *Controlling and selecting civic issues that are assigned to algorithm decision-making*

It may well be pointed out that no decision-maker would like to

[62] Couldry and Powell, *supra* note 56, at 5.

[63] M. Hildebrandt, *Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning*, in *Theoretical Inquiries in Law*, vol. 83, 2019, 83–121.

expose their choices to a fact-checking system that is the result of a non-empathetic assessment not connected to a genuine political appreciation of their choices. This is where the relevance of the second element of the algorithmic decision-making frame comes in: the opportunity to exercise control over the issues that can be delegated to algorithmic decision-making processes.

Algorithmic decision-making is often praised for being efficient, while critical voices have been raised concerning the discriminatory biases that algorithms often incorporate and reproduce[64]. Therefore, allocating public choices to AI can result in a blank check, even when precise instructions have been set up. This is especially true in the context of machine learning when human intelligence's contribution is limited and the machine is capable of setting autonomous premises for logical deductions[65]. An uncontrolled use of AI increases people's scepticism regarding algorithmic decision-making, thus leading them to question its legitimacy in a liberal democracy grounded on moral values that have legal significance. As Tom Tyler has argued, people's perceptions of legitimate authority are more likely to be met if authorities ensure the existence of procedural safeguards as well as the opportunity to be heard for individuals[66]. To avoid the blank check effect, then, the processes through which AI reaches decisions need to be complemented by procedural safeguards that should, at least, enable people, through their representatives, to exercise full control over determining the issues that are delegated to an AI process. In this respect, the role of parliaments is crucial. Algorithmic decision-making can contribute

[64] F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, MA, Harvard University Press, 2015, pp. 1–17 (arguing that data analyses about consumers are often not disclosed and even hidden from legal process). See also A. Chander, *The Racist Algorithm*, in *Michigan L. Rev*., vol. 115, 1023, 2017. Studies have been specifically conducted on the credit sector to demonstrate biases on the allocation of loans: L. Rice and D. Swesnik, *Discriminatory Effects of Credit Scoring on Communities of Colour*, in *Suffolk University L. Rev*., vol. 46, 935, 2013, showing that credit-scoring systems in the USA systematically discriminate against communities of colour vis-à-vis White Americans. See also B. Birnbaum, *Credit Scoring and Insurance: Costing Consumers Billions and Perpetuating the Racial Divide*, Boston, MA, National Consumer Law Center, 2007, maintaining that discrimination for residents in minority communities has been demonstrated by studies on insurance scores calculated by algorithms that had eliminated factors such as income, education or unemployment status.

[65] C. Riegler, *The Moral Decision-Making Capacity of Self-Driving Cars: Socially Responsible Technological Development, Algorithm-Driven Sensing Devices, and Autonomous Vehicle Ethics*, in *Contemporary Readings in Law and Social Justice*, vol. 11, 15, 2019.

[66] T. Tyler, *Why People Obey the Law*, Princeton, NJ, Princeton University Press, 1990, pp. 96, 137–38.

to democratic legitimisation to the extent that parliaments are enabled to control the issues that are delegated to AI, the procedures and conditions under which the issues are evaluated and the checks on the fairness of the overall decision-making process. Such an approach requires that parliamentary procedures are in place for (1) selecting issues that can be assigned to AI decision-making and (2) proceduralising this particular kind of decision-making process in a way that enables individuals to evaluate and challenge its outcomes, as will be further explained in Section IV.3.

As for the selection of issues, this essentially means that parliaments should assess whether certain issues can be dealt with by AI and the extent to which it can do so. The principle that statutory law is required to regulate AI decision-making is not included in existing international legal instruments, and such a principle is uncertain or weak in national jurisdictions as well. Neither is Article 22 GDPR diriment – it has not been interpreted in such a way as to demand that the law should explicitly authorise resorting to automated decision-making. For example, the UK's Information Commissioner Office (ICO) has stated that when either legislation or common law authorises the exercise of power, the choice of means to achieve purposes consistent with legitimate powers includes resorting to automated decisions, with no need for an ad hoc authorisation by law[67].

Let us consider the case of welfare benefits allocation choices, which in recent years have been delegated to AI to process the identification of beneficiaries as well as the sums they were entitled to receive[68]. There are hardly any cases of formalised delegation from democratically legitimised public authorities to automated processes of such substantive decisions.

Most often, automated decisions pertain to technical tasks to be performed downline of a political decision (eg whether to establish a certain welfare benefit or not) made elsewhere, on the assumption of the efficiency of AI.

Concerns regarding the impacts of automated decision-making on socioeconomic rights have been expressed by academics and civil society

[67] See <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-dataprotection-regulation-gdpr/automated-decision-making-and-profiling/when-can-we-carry-out-thistype-of-processing/#id1>.

[68] Examples include resorting to automated decisions in relation to eligibility to receive welfare benefits such as Medicaid and food stamps at the state level in the USA: see V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*, New York, St Martin's Press, 2018. Local authorities also use automated decision-making procedures to assign welfare benefits in the UK: see L. Dencik, J. Redden, A. Hintz and H. Warne, *The 'Golden View': Data-Driven Governance in the Scoring Society*, in *Internet Policy Rev.*, vol. 8, 2019, <https://doi.org/10.14763/2019.2.1413>.

organisations. Among the most high-profile and influential critical voices is the report on "Extreme Poverty and Human Rights" released by the UN Special Rapporteur on extreme poverty and human rights in 2019[69]. The report expresses concerns regarding the application of algorithmic decision-making to welfare services for at least three separate ranges of reasons: (1) lack of privacy and data protection (especially in the phase of identity verification), even veering towards systemic surveillance (exploited on the pretext of preventing and detecting fraud); (2) system failures (while assessing receivers' eligibility, benefit calculations and payments); and (3) unfairness (from basing decisions affecting individual rights on group-based predictions), lack of transparency and the risk of reinforcing existing inequalities and discrimination, particularly in relation to risk scoring and need classification. Such concerns question the appropriateness of AI decision-making on socioeconomic rights, irrespective of the constitutional or legislative status they enjoy in a particular country. In fact, the danger of discrimination and inequality predates the development and possible uptake of AI. Where the relevant rights are granted at the legislative level, the risk is one of violating due process guarantees; by contrast, where they enjoy constitutional status, the problem is even greater in terms of the limitation or curtailment of fundamental rights. The choice of assigning a given issue to automated decision-making should therefore be carefully assessed within the political process and not made on the simplistic assumption of the efficiency that AI may bring into the public institutions.

The Commission White Paper on Artificial Intelligence engages with similar issues[70]. Indeed, the Commission states the need for the EU to implement the use of AI in public sectors to the advantage of citizens, while at the same time it remarks on the relevance of a regulatory framework ensuring the protection of European values and fundamental rights. The White Paper's concern for fundamental rights, however, is not exclusively devoted to protecting privacy and avoiding discrimination. In fact, the Commission is preoccupied with envisioning the use of AI in a context in which citizens trust institutions because they perceive that automated decisions are encapsulated in a system that takes into consideration human dignity and all of the fundamental rights protected at the EU level[71]. The relevance of the principle of trust is also reflected in the Commission's

[69] United Nations General Assembly, *Extreme Poverty and Human Rights*, Report of the Special Rapporteur on extreme poverty and human rights, 11 October 2019.

[70] On Artificial Intelligence – A European Approach to Excellence and Trust, COM(2020) 65 final.

[71] See also Communication on Building Trust in Human-Centric AI, COM(2019) 168.

approach to the assessment of risks connected to AI. In particular, the White Paper stresses the importance of evaluating the impacts of AI not only from an individual perspective, but also from a societal one. In doing so, the Commission favours the idea of AI as a functional tool to enhance equality of opportunity, socioeconomic welfare and democracy. The Commission has advanced a similar approach in the Communication to the European Parliament titled "Fostering a European Approach to Artificial Intelligence", in which it has identified a strategy built on several requirements for an AI ecosystem that also tackles public responsibility, including societal and environmental wellbeing[72].

Neither the White Paper nor the AI strategy, however, addresses openly how the use of AI can become entrenched in the democratic process. Rather, at its present stage of evolution, EU law conceives of trust in the AI ecosystem as a principle that is fulfilled by substantive rules protecting fundamental rights and ensuring oversight and accountability mechanisms. Such an approach is also confirmed by the Commission's Proposal for a Regulation on Artificial Intelligence[73]. Taking stock of the White Paper, the Proposal focuses on addressing risks associated with the use of algorithm technology by advancing "a legal framework for trustworthy AI". To that effect, the proposed regulation identifies different levels of risks connected to algorithmic technology and establishes transparency obligations for those systems that interact with humans or with their emotions, as well as for those generating or manipulating contents. The legal framework is completed by setting up a governance system for AI that establishes a system based on the interplay between the EU and national authorities. At the EU level, the Commission proposes the creation of a European Artificial Intelligence Board with the task of guaranteeing the effective implementation of the regulation. At the national level, Member States will be required to designate competent authorities to apply the regulation, including a national supervisory authority. Those authorities will be designed to supervise AI systems providers. In particular, those bodies will set out to monitor and report obligations regarding use, incidents or malfunctioning of AI technology[74]. While the Proposal is mainly concerned with guaranteeing fundamental rights and the functioning of the internal market, nothing in the Proposal suggests that EU institutions in fact are specifically concerned with the broader approach involving the political

[72] COM(2021) 205 final.
[73] COM(2021) 206 final.
[74] COM(2021) 206 final, Titles VI, VII and VIII.

decisions of resorting to AI in decision-making process in the first place.

While the proposal put forward by the EU authorities focuses on a (legitimate and timely) concern for the concrete impacts of AI on fundamental rights, it appears to disenfranchise representative political processes from the governance of AI. However effective it may be to set up dedicated agencies to monitor the impacts of AI on fundamental rights, proceduralisation of AI is not a guarantee per se[75], and the current European framework disregards the political value of the choice of authorising the use of AI in decision-making processes.

4.3. *Evaluating and challenging algorithmic decision-making*

Following on from this analytical framework, we now look at the potential for algorithms to increase the accountability of public policies by making relevant information more easily accessible, providing information more accurately and facilitating public debate among citizens.

Looking at the impact of algorithms from this specific angle raises apparent challenges, even at first glance. The academic debate on the topic has traditionally discussed the two issues of algorithmic transparency and accountability as being closely intertwined. In the academic literature, the issue of the chronic lack of meaningful ways to subject algorithm-led decision processes to thorough scrutiny has been connected to characteristics such as complexity and secrecy, leading to the now popular expression "algorithmic black box"[76].

It may be that, due to the very nature of such technologies, black-box issues cannot find satisfactory solutions. It has been emphasised that the level of sophistication of current AI techniques and deep learning oftentimes makes the details of decisional rules and patterns impossible to explain, even for the developers of the technology, and when decision-making processes in the public sector involve a certain degree of discretion it may be necessary to always require human intervention[77]. More optimistically, other academic authors have already suggested

[75] R. Koulu, *Proceduralizing Control and Discretion: Human Oversight in Artificial Intelligence Policy*, in *Maastricht Journal of European and Comparative Law*, vol. 27, 2020, 720–35.

[76] Pasquale, *supra*, note 64.

[77] H-W Liu et al, *Beyond State v Loomis: Artificial Intelligence, Government Algorithmization and Accountability*, in *International Journal of Law and Information Technology*, vol. 27, 122, 2019, p. 139.

that algorithmic decisionmaking processes could, both in theory and in practice, lead to "fairer and more objective decisions, grounded in data that are representative of the community where the decisions apply"[78] thanks to the implementation of a range of possible technical solutions (eg deploying participatory processes to include diverse and local voices in co-designing algorithms and vetting procedures to allow for the sharing of data and algorithm templates)[79]. Other authors have expressed optimistic views that research is underway to promote more transparency, ideally to the point of providing an explanation of decision models sufficient to allow a human to understand how inputs relate to predicted results[80].

However, one question that remains is how to make algorithms transparent; another related yet different question is how to make sure that algorithms can contribute to the accountability of policy decisions and democratic legitimacy. While technology and law hopefully both progress on parallel tracks to make "algorithmic boxes" less "black", we turn to the question of whether and how algorithmic decisionmaking can be subjected to thorough evaluation and, when necessary, challenged in a way that boosts democratic legitimacy.

Several regional and international organisations have considered this issue and recommended policy frameworks that expressly connect requirements such as transparency and explicability with accountability. UNESCO's draft recommendation is particularly straightforward in making the connection between explicability (defined as "making intelligible and providing insight into the outcome of AI systems", with specific regard for the "input, output and behaviour of each algorithmic building block and how it contributes to the outcome of the systems") and transparency, as well as the connection between these two principles and accountability and trustworthiness, since making processes and outcomes transparent and traceable is conducive, in turn, to keeping AI actors responsible and liable for their decisions. Appropriate oversight and due diligence mechanisms throughout the life cycle of AI systems are thus recommended[81]. Similarly, the EU's Ethics Guidelines for Trustworthy AI require that automated

[78] B. Lepri et al, *Fair, Transparent, and Accountable Algorithmic Decision-Making Processes. The Premise, the Proposed Solutions, and the Open Challenges*, in *Philosophy & Technology*, vol. 31, 611, 2018, p. 622.
[79] *Ibid*, 623.
[80] B. Goodman and S. Flaxman, *European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation'*, in *AI Magazine,* vol. 38, 2016, DOI: 10.1609/aimag.v38i3.2741.
[81] UNESCO Recommendation, paras 40–41.

processes be "transparent" and any decisions "explained to those directly and indirectly affected" "to the extent possible" so that they can be "duly contested". When black-box or other technical difficulties make full explicability impossible, then other back-up measures should be provided (eg "traceability, auditability and transparent communication on system capabilities"), proportionately to the context and severity of the consequences of a possible erroneous policy outcome[82]. The OECD recommends "transparency and responsible disclosure regarding AI systems" through the provision of "meaningful information, appropriate to context and consistent with the state of art" to "enable those affected by an AI system to understand the outcome" and "enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information of the factors, and the logic that served as the basis for the prediction, recommendation or decision"[83].

Transparency and explicability are also the core principles in Articles 15 and 22 GDPR, which have made the rights to an explanation and to contest decisions parts of a fundamental right to data protection. Any references to Articles 15 and 22 GDPR need to be interpreted, in this context, only lato sensu, as in fact in many cases public policy decisions may be based on anonymised data, which in any form do not involve the right to personal data protection in a direct way. These provisions ultimately appeal to the idea that the opacity of "black-box" decisions could be countered by a right to an explanation and to "open" the box.

Other suggestions have been advanced in the literature to make algorithms explicable and accountable. Doshi-Velez and Kortz, for instance, considered the possibility of legally mandating that governments explain the ways in which inputs affect outcomes, concluding that such an obligation would most likely be technically feasible but onerous[84]. The requirement of a "human in the loop" has also proven to be popular; however, concerns have been raised over this possibly being counterproductive, for it would exclude from the scope of application of GDPR's safeguards a large number of decisions in high-risk areas[85].

[82] High Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 2019, p 13.
[83] OECD Recommendation, Principle 1.3.
[84] F. Doshi-Velez and M. Kortz, *Accountability of AI Under the Law: The Role of Explanation*, Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, 2017.
[85] S Wachter, B Mittelstadt and L Floridi, "Transparent, Explainable, and Accountable AI for Robotics" (2017) 2(6) Science Robotics eaan6080.

The right to an explanation is thus not enough in and of itself to enable a meaningful right to challenge algorithmic decisions. Yet some recently developed legal frameworks evidently focus on the former but not the latter; an example of this is France's Digital Republic Law[86]. This law requires large public-sector bodies to make publicly available, in an easily accessible format, the algorithmic processes that were used to reach decisions affecting individuals. This provision takes a step forward compared to the GDPR in that it applies to partially automated decisions as well; it does not, however, provide for a specific procedure to challenge such decisions. While the original bill was debated, however, the advisory commission had recommended introducing a new individual right to oppose profiling, to require human intervention and to oppose discriminatory decisions[87].

Another suggestion for extending the scope of this newly established right was put forward by researchers working in the French Prime Ministerial task force for open data and open government, Etalab. The researchers identified a set of discrete (yet interconnected) obligations that public authorities should meet when making decisions based on algorithms, such as acknowledging the (entire or partially) automated nature of the process, explaining the functioning of the algorithm, offering a justification for choosing that particular algorithm, making the relevant source code and documentation public and, finally, providing mechanisms for challenging the outcomes[88].

The idea of turning the principles of explicability and accountability into actual legal rights certainly marks a step in the right direction. However, Edwards and Veale noted that when automated processes affect large groups, individuals are usually unlikely to successfully mount challenges, and representative bodies may be in a better position to exercise this function, provided they are equipped with certain technical requirements, such as access to data used for training and modelling purposes, amongst others[89].

It may seem far more appropriate for a representative body to accept and mount challenges than each individual user when it comes to public policy decisions. Following on from Edwards and Veale's suggestion that

[86] Law no. 2016-1321 du 7 octobre 2016 pour une République numérique.

[87] <https://www.assemblee-nationale.fr/14/pdf/rapports/r3119.pdf>.

[88] S. Chignard and S. Penicaud, *'With Great Power Comes Great Responsibility': Keeping Public Sector Algorithms Accountable*, Etalab Working paper on algorithmic accountability, 2019.

[89] L. Edwards and M. Veale, *Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'?*, in *IEEE Security & Privacy*, vol. 46, 2018, 16.

representative bodies step up, it is useful to recall here two principles included in the Council of Europe Recommendation on the human rights impacts of algorithmic systems[90], namely the principle of democratic participation and awareness (urging state authorities to "foster general public awareness of the capacity, power and consequential impacts of algorithmic systems") and the expectation that adequate institutional frameworks are set up providing for "general or sector-specific benchmarks and safeguards".

In the context of public policies, the best-suited bodies seem to be – once again –parliaments rather than data protection authorities as a way of reprising what is ultimately a traditional role of parliaments in keeping governments accountable, though possibly in a different form. As much as populist rhetoric aims to cut parliaments out of their traditional intermediary role between the people and decision-making circles, implementing algorithms – in a way that is mindful of the principles included in the policy and legal frameworks illustrated above and the findings from the literature –offers a way to reinstate them in the policymaking cycle. If parliaments – along with individuals – were granted access to specific pieces of information, as described above (the functioning and justification of the algorithm, including the source code), and to a dedicated mechanism for challenging the outcomes, this could reinforce the democratic accountability of algorithmic decision-making and even fill in a longstanding gap in the field of policy evaluation.

The debate surrounding the assessment of the outcomes of policy decisions and the kind of metrics that would be best suited for this purpose in fact predates the rise of algorithmic decision-making. Especially after the advent of "new public management", emphasis has progressively grown on finding ways to quantify and measure government performances. Yet the relationship between performance and its quantification has proven a difficult one: different alternative metrics (such as targets, rankings and intelligence) have all been found to present shortcomings and only work effectively under specific circumstances[91]. The advent of digital technologies, and big data in particular, has offered new solutions and raised several new concerns in turn. Academic commentators have questioned the possibility of completing thorough evaluations by meaningfully connecting data

[90] Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, adopted by the Committee of Ministers on 8 April 2020.

[91] C. Hood, *Public Management by Numbers as a Performance-Enhancing Drug: Two Hypotheses*, in *Public Administration Rev.*, vol. 72, 2012, p. S85–S92.

collected through all of the phases of the policy cycle[92] and warned about the risk of an excessive focus on data in evaluation processes distracting from underlying substantive problems[93]. There are thus inherent difficulties in choosing the method, or even the relevant data, to enable a thorough assessment of government performance.

It is important to consider that for parliamentary scrutiny to contribute to the procedural legitimacy of policymaking decisions scrutiny itself ought to focus on the effects and impacts of such decisions rather than on their processes. This is only apparently a contradiction; substantive scrutiny on the outcomes can well be a fundamental aspect of procedural legitimacy. In fact, if black-box issues can challenge parliaments' ability to be effectively capable of performing thorough scrutiny of the technicalities of the process, it may then seem paradoxical that algorithms could contribute to the legitimacy of democratic governance while their own internal legitimacy and transparency are disputed. The two grounds, however, are to be kept separate.

In his strong critique of algorithmic decision-making and its underlying neoliberal mind-set, Ari Ezra Waldman has contended that there is a lack of effectiveness of any such mechanisms aimed at correcting the process of algorithmic decision-making based on the traditional equation of fair processes with fair results[94]. However, in the author's view, boosting transparency and other procedural safeguards cannot overcome the underlying characteristic of algorithmic decision-making as "agnostic about its sociopolitical and economic implications"[95]. In the context of decisions made by AI, the "emphasis on efficiency … undermines the effectiveness of procedures"[96]. In turn, Waldman suggests that scrutiny of algorithmic decisions should focus on substantive outcomes and their impacts on fundamental rights rather than on their mere procedural fairness.

In our framework, however, the emphasis on results is a means to an end; focusing on the results helps to create a place for public discussion where discursive legitimacy is built by offering the opportunity for rational debate. A set of legal provisions granting parliaments (as well as

[92] DF Kettl, "Making Data Speak: Lessons for Using Numbers for Solving Public Policy Puzzles" (2016) 29(4) Governance 573–79.
[93] P. White and R. S. Breckenridge, *Trade-Offs, Limitations, and Promises of Big Data in Social Science Research*, in *Review of Policy Research*, vol. 31, 2014, pp. 331–38.
[94] AE Waldman, "Algorithmic Legitimacy" in W Barfield (ed.), Cambridge Handbook of the Law of Algorithms (Cambridge, Cambridge University Press 2020) p 107.
[95] *Ibid*, 116.
[96] *Ibid*.

individuals) access to specific pieces of information and the opportunity to challenge the outcomes of decisions against specific benchmarks, such as their impacts on fundamental rights, can streamline the process and offer a more robust underpinning to it (focusing on outcomes rather than data per se).

In line with the rationalist deliberative proceduralist approach illustrated above[97] and the focus on output legitimacy, empowering parliaments to scrutinise the outcomes of automated decisions can offer an opportunity for institutional scrutiny in order to minimise the potential negative impacts of technology. At the same time, algorithmic decision-making can offer a ground for parliamentary scrutiny and enable the public debate to refocus on rational discourses and their ability to produce rationally justified outcomes through fair procedures.

5. *Concluding remarks*

Algorithmic decision-making has started to be used in many policy areas in recent years, and it is now on trial in many others, such as welfare benefits and immigration management. Although critics point out the discriminatory effects, the human rights violations and the distorting effects on political communities as a whole of such systems, algorithmic decision-making seems to be "here to stay"[98]. As is often the case with instruments of scientific progress, their mere existence makes the case for their use. Without dismissing those relevant and salient concerns, we focused on a complementary perspective to discuss the opportunities and the potentialities of algorithmic decision-making.

The model suggested in this article tested the hypothesis that algorithms can contribute to increasing democratic legitimacy at times of rampant populism, provided that their use takes place within a framework that maximises political equality and rational decision-making by enabling wider participation, consideration of diverse social issues and oversight of

[97] F. Peter, *Democratic Legitimacy and Proceduralist Social Epistemology*, in *Politics, Philosophy & Economics*, vol. 6, 2017, pp. 329–53.

[98] As has been shown by the recent study "Understanding Algorithmic Decision-Making: Opportunities and Challenges" of the European Parliamentary Research Service EPRS | Scientific Foresight Unit (STOA), PE 624.261, March 2019, available at <https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf>.

the decisions made.

We suggest that these conditions are met when citizens are able to: (1) understand the civic issues assigned to AI; (2) control the agenda of AI decision-making; and (3) evaluate and challenge the outcomes. For these conditions to be met, we suggest that frameworks such as a right to explanation, parliamentary oversight and opportunities to challenge the elements of an algorithm's processes, both judicially and in public debate, are put in place.

The efficiency as well as the representation of reality that an algorithmic process may be able to produce expose the fallacies or the "easy truth" of populist rhetoric. From such a viewpoint, algorithmic decision-making is not good because the output is intrinsically trustworthy. It is good as long as it is embedded in a democratic frame that enables both the represented and representatives to exercise choices and control over the decision-making process. In this way, algorithms can expose the populist rhetoric by being an instrument of knowledge and therefore a tool to read the reality and solve its problems.

Acknowledgements. The authors wish to thank the anonymous reviewers for their precious comments and suggestions. The usual disclaimers apply.

Competing interests. The authors declare none.

Fabiana Di Porto

Algorithmic Disclosure Rules

Abstract: During the past decade, a small but rapidly growing number of Law&Tech scholars have been applying algorithmic methods in their legal research. This Article does it too, for the sake of saving disclosure regulation failure: a normative strategy that has long been considered dead by legal scholars, but conspicuously abused by rulemakers. Existing proposals to revive disclosure duties, however, either focus on the industry policies (e.g. seeking to reduce consumers' costs of reading) or on rulemaking (e.g. by simplifying linguistic intricacies). But failure may well depend on both. Therefore, this Article develops a 'comprehensive approach', suggesting to use computational tools to cope with linguistic and behavioral failures at both the enactment and implementation phases of disclosure duties, thus filling a void in the Law & Tech scholarship. Specifically, it outlines how algorithmic tools can be used in a holistic manner to address the many failures of disclosures from the rulemaking in parliament to consumer screens. It suggests a multi-layered design where lawmakers deploy three tools in order to produce optimal disclosure rules: machine learning, natural language processing, and behavioral experimentation through regulatory sandboxes. To clarify how and why these tasks should be performed, disclosures in the contexts of online contract terms and privacy online are taken as examples. Because algorithmic rulemaking is frequently met with well-justified skepticism, problems of its compatibility with legitimacy, efficacy and proportionality are also discussed.

---

* This article was first published in *Artificial Intelligence and Law*, 31, 2023, pp. 13-51. This article has been written during the research period spent at and benefitted from the generous contribution of the Hebrew University, Israel. I am deeply grateful for insightful discussions to a number of people: Omri Abend, Ittai Bar Siman Tov, Netta Barak-Corren, Tamar Berenblum, Antonio Davola, Lex de Lange, Orr Dunkelman, Catalina Goanta, David Hay, Renana Keydar, Marco Lippi, Michael Livermore, Boaz Matan, Astorre Modena, Giorgio Monti, Monica Palmirani, Marilena Rizzo, Yuval Shany, Limor Shmerling Magazanik, Michal Shur-Ofri, Daniel D Sokol, Keren Weinshall, and Eyal Zamir; the participants to the Workshops 'Law, AI and Data Science: Challenges and Opportunities' held at Bar Ilan University, 18–19 December, 2019; the 'HUJI Federmann Cyber Security Summit Meeting' (13 May 2020) 'Private and Commercial Law Meeting' (22 June 2020), both held at the Hebrew University; the Conference 'Should Data Shape Private Law? Between Stereotypes and Personalization', organized jointly by the Universities of Tilburg, Maastricht and Osnabrück, 4–5 June 2020. I am thankful to Tatjana Grote for her wonderful research assistance. All mistakes remain my own.

1. *Introduction*

Ms Schwarz had just finished editing her article, and was about to email it to a journal, when her laptop stuck: she was asked to choose among three different layouts of her browser's tab, so that next time she would open the app, the tab would look exactly like her preferred option, between an 'Inspirational', 'Informational' and 'Focused' appearance.

Here are the layouts:

Inspirational Informational Focused

You can always change it later in Page Settings

The browser tab is an example of algorithmically 'targeted disclosure', through which a company[1] uses Information Technology[2] to convey information to consumers in a way that suits their preferences. It also provides a nice way to visualize the goal of this article: How can NLP and ML algorithms[3] be used to target disclosure rules at clusters (not individuals) of consumers to reduce the rules' failures? How could that be done in ways to ensure that disclosure rules be implemented automatically by the industry, thus significantly decreasing the cost of complying? And how could disclosure be differentiated to target the different informational preferences of consumers? Could that be done by rulemakers? Taking the economic ground for producing disclosure rules as a given, how should

[1] The tab layouts are by Edge, the new Microsoft's web browser (2020).

[2] Here human–computer interaction.

[3] While the history of automation in legal science can be traced back to the early 1950s, the rise of highly performative NLP tools and ML algorithms has substantially widened the spectrum of possible applications: See F. Fagan, *Big data legal scholarship: toward a research program and practitioner's guide*, in *Virginia. J. L. Technol.*, vol. 20, 1–81, 2016, and M. Medvedeva, M. Vols, M. Wieling, *Using machine learning to predict decisions of the European Court of Human Rights*, in *Artif Intell. L.*, vol. 28, 237–266, 2019, <https://doi.org/10.1007/s10506-019-09255-y> for review.

rulemakers proceed?

We take online contract terms and online privacy disclosures as examples to illustrate our proposal. In fact, both are massively produced by providers of websites, are usually not subject to face-to-face negotiations with consumers but rather accepted on a take-it-or-leave-it basis. Far from reducing information asymmetry and increasing the bargaining power of consumers, the information conveyed through such duties not only increases obfuscation (Bar-Gill 2014), but is often 'weaponized' by the industry (Luguri and Strahilevitz 2021; Stigler Center 2019) to steer individuals towards behaviors that maximize the industry's profits (Thaler 2018). For instance, through lengthy and obscure contract terms, companies may increase the acceptance rate of online offers, or induce individuals to buy tied insurance services. Similarly, by framing the choice of lenient cookie policies in a more prominent way, users are induced to provide more personal data than they would according to their rational preferences.

That disclosure regulation online is prone to failure is nothing new (Ben-Shahar and Schneider 2014; Bakos et al. 2014; Marotta-Wurgler 2015). And its detractors are as numerous as are attempts to revitalize it. Among the latter, the most promising are the Law&Tech scholars (Ashley and Kevin 2017; Livermore and Rockmore 2019), some of whom suggest using NLP and ML tools to personalize, simplify and summarize disclosures. Some authors (Ayres and Schwartz 2014) propose to automatically detect unexpected or unfavorable terms in privacy disclosure policies, and presenting them in a separate warning box. Others suggest to frame terms of contracts differently to increase readability of online privacy policies, based on behavioral evidence (Plaut and Bartlett 2012); still others propose to base the decision of which parts consumers should pay special attention to on the requirements imposed by privacy law and consequently focus on choice provisions (Mysore Sathyendra et al. 2017). Moreover, others recommend to employ bots to highlight the content of platforms' privacy disclaimers or help educate consumers (Harkous et al. 2018). Finally, an important literature is emerging that identifies "probabilistic disclosures" as superior to discrete yes/no type disclosures (Levmore 2021)[4]. Although the proposed solutions are certainly promising, and come with the great benefit of only marginally intervening with the consumer's autonomy, they

[4] Demonstrating that for some addressees, information provided in probabilistic terms may be more informative than generic one, and suggesting that law not only allows disclosure to be provided in diverse formats, but also safe harbors for probabilistic disclosures.

present some critical shortcomings.

First, they suggest intervening mostly on the implementation phase, namely to adjust fallacies at the firm-level disclosure policies, assuming that the sole reason disclosure regulations fail is the prohibitive cost of reading (Bartlett et al. 2019). For instance, a great work has been done using algorithms to show that online privacy policies are often incomplete (Contissa et al. 2018a, b; Lepina et al. 2019) or some of their clauses are linguistically imprecise (Liu et al. 2016). With regard to online contracts, an algorithm has been developed to automatically detect clauses that are potentially unfair under EU law (Lippi et al. 2018).

However, the source of failure may well depend *also* on how disclosure rules are formulated in the first place. For example, goals may be self-defeating: the GDPR, Article 12 requires data controllers to provide individuals with the information on their rights (stipulated in Articles 13 and 14) 'in a concise, transparent, intelligible and easily accessible form, using clear and plain language'. Being concise and intelligible at the same time can be two conflicting goals.

Second, the solutions proposed make assumptions about what consumers need to be warned of that are based on singular surveys. In this sense, they are static, instead of being continuously revived and dynamically updated with the support of live data. However, this would be necessary in order to react to shifts in both consumer preferences and business behavior. Designing legal solutions on evidence that is gathered through ad hoc experiments may be limiting, as new evidence may come about showing opposing results. For instance, in the US, the recently proposed 'Algorithmic Justice and Online Platform Transparency Act'[5] establishes new disclosure duties on online platforms[6]6 with regard to the algorithmic processes they utilize to 'promote content to a user' (i.e. personalized product/service). Although these data are made available to the regulator[7], the FTC would only access past (not real time) information. In other words, this solution may not allow capturing in due time if consumers become reactive to some piece of information and not to other, or if companies adapt their disclosures to behavioral changes of consumers

[5] H.R. 3611, 'Algorithmic Justice and Online Platform Transparency Act', 117th Cong. (2020–2021) of 28 May 2021.

[6] Section 4(1)(a) H.R. 3611, 117th Cong. requires that, for each process, users are informed of: (i) their personal data; (ii) in what way they are collected or created by the platform; (iii) how the latter uses them; and (iv) what methods are used to prioritize, assign weight, rank personal data to deny, amplify, recommend, or promote content to a user.

[7] Section 4(a)(2)(C) H.R. 3611, 117th Cong. (above, note 5).

(as is the case with dark patterns)[8] or to counter changes in the law. Therefore, repeated experiments using real time data would be preferable in order to sustain regulatory intervention. Apart from that, no evidence is available of the number of consumers actually using, or the efficacy of the bots like CLAUDETTE (Lippi et al., at 136–137) and darkpatterns.com (Calo 2014; Stigler Center, at 28).

Before this background, this Article innovates in a relevant regard: it argues that algorithmic tools can and should be used that 'comprehensively' consider solutions at the drafting stage jointly with the implementation phase of disclosure regulation. This 'comprehensive approach' conceptualizes disclosure regulation as a process composed of rulemaking and implementation, and therefore suggests using algorithms to tackle the fallacies affecting each step singularly and all of them together (part one). As a unique contribution, this article elucidates on how to implement this 'comprehensive approach' in practice. From a technical point of view, lawmakers should deploy three tools in order to produce optimal disclosure rules: machine learning (ML), natural language processing (NLP), and behavioral experimentation through regulatory sandboxes (part two).

Lawmakers should begin (Phase One) by creating a dataset composed of disclosure rules, firm-level disclosure policies, and the case law pertaining to both (Sect. 3.1.1). Next, they should deploy NLP and other techniques to map out the causes of failure, rank the disclosures accordingly and find the best matches of law and implementation on the basis of such ranking (Sect. 3.1.2). The goal would be to develop an ontology of self-implementable rules that produces good outcomes in terms of readability, informativeness, and coherence, which could be dynamically updated. This ontology is called HOD or Hypothetically Optimal Disclosure (Sect. 3.1.3), which would only include disclosure rules that fail the least (according to our library of measurable failure indexes), and therefore give rise to the least disputed issues (out of the relevant case-law). HOD raise nonetheless questions of efficacy, legitimacy and proportionality that need to be addressed (Sect. 3.1.4).

For Phase Two suggests exploring the potential of regulatory experimentation in sandbox, as a viable solution. We propose testing HOD with regulatory experimental methods, as a unique solution. Regulatory sandboxes are thus presented as a means to pre-test of different layouts of HOD disclosures with stakeholders in a collaborative (co-regulatory) fashion; to ensure transparency and participation; target disclosures to

[8] Section 2 (Findings) H.R. 3611, 117th Cong.

increase efficacy; and cluster individuals (Sect. 3.2.1).

Section 3.2.2 explains how the sandbox is organized, both from a governance perspective and a technical one. The final outcome to sort with, once behavioral data from the sandbox are integrated, is the Best Ever Disclosures, or BED: an algorithm producing legal notices that would be targeted at clusters of consumers; updated continuously with rules, caselaw and behavioral data, and that would also be automatically implementable.

Section 3.2.3 explains how automatic implementation of BED on large scale works, both at the very first launch on the market, and successively, when amendments are needed.

Lastly, a discussion of possible drawbacks and wider effects of BED algorithmic disclosures on stakeholders is presented (Sect. 3.2.4) before concluding.

2. *Part one: The case for a 'comprehensive approach'*

2.1. *Disclosure regulation in online markets: a failing strategy in need of a cure*

Traditionally, Disclosure Regulation serves function of reducing information asymmetries that plague consumers (Akerlof 1970) and their unequal bargaining power (Coffee 1984; Grossman and Stiglitz 1980). In online consumer transactions we are literally flooded with transparency and disclosure duties and policies. For instance, the EU Consumer Rights Directive (CRD)[9] contains several rules mandating the provision of information to the point of limiting the freedom to design e-commerce websites[10]. The CRD also relies on pre-contractual information requirements to protect consumers: online marketplaces must inform consumers about the characteristics of a third party offering goods, services, or digital content in the online marketplace[11]; state if the provider

[9] Directive 2011/83/EU, amended by Directive (EU) 2019/2161 of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU as regards the better enforcement and modernisation of Union consumer protection rules, OJ L 328, 18 December 2019, 7–28.

[10] CRD Art. 8(2) sets out clear requirements for the design of buttons in online consumer transactions: they may only state 'order with obligation to pay' or similarly unambiguous formulations.

[11] CRD Art. 6a(1)(a).

is a trader or not (in which case, other and less protective laws would apply) (Di Porto and Zuppetta 2020)[12]; or break down key information which involve costs[13]. Similarly, in the online privacy field, disclosure duties have flourished: 'cookie banners' (or more precisely 'consent management platforms', CMP) appear at any first website access requiring user consent for personal data processing, based on legal requirements in both the EU[14] and the US[15].

However, the appropriateness of such duties to provide effective protection to consumers is knowingly poor. Online contract terms and online privacy policies are unilaterally designed by the platform, and are mass-marketed online, essentially on a take-it-or-leave-it basis (Bar-Gill 2014). In this scenario, the platform can determine the 'choice architecture' in which consumers act, thus deliberately exploiting irrational consumer behavior to increase its profits (sludging) (Thaler 2018). For instance, one may accept to provide more personal data than she would deem reasonable according to her preferences, or accept to buy more quantities of a given service.

Personalization has boosted the manipulative power of platforms, (Zuboff 2019) and digital firms have become skilled at developing 'dark patterns'. (Brignull 2013) through which the most vulnerable consumers are especially targeted (Stigler Center 2019).

Disclosure duties can do little to intercept or counter these practices or educate consumers. This because of the disproportionate informational disadvantage of which regulators suffer vis-à-vis the industry. Regulators do not possess granular and real-time data about users' behavior, nor can they observe changes in privacy policies made by the industry as a response. They can certainly run experiments and collect data, but do not have enough resources to do so on a regular basis, as can digital firms (e.g. by running A/B testing). In addition, educational campaigns for consumers do not seem a viable solution, not only because they suffer from a collective action problem (Bar Gill 2014), but also because they are costly for the industry.

[12] If the user is not a consumer, then the relationship is one of one of Business-to-Business and hence covered by the (less protective) Regulation (EU) 2019/1150 of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services, OJ L 186, 11 July 2019, 57–79.

[13] CRD Art. 6(1)(e).

[14] See the General Data Protection Regulation (GDPR) art. 7 and Alinea 32 (requiring 'informed' consent to data treatment).

[15] The California Consumer Privacy Act (CCPA) (requiring businesses to give consumers information about the data they collect and the way they use it, at the time or before they start collecting it, lit. in a 'notice at collection').

Despite all this evidence, rule-makers continue to employ disclosure regulation massively in both online contract terms and the privacy contexts. To quote a few: traders of online marketplaces like Amazon shall inform consumers if their prices are personalized[16]; to ensure that reviews originate from real customers or are not manipulated, platforms must provide with 'clear and comprehensible information' about the 'main parameters determining the ranking' in research queries[17]; online general search engines (like Google, Edge and the like) must provide a 'description of the main ranking parameters and of the possibilities to influence such rankings against remuneration'[18].

In the US, as seen, the information a consumer needs to be provided with must be given in a 'in conspicuous, accessible, and plain language that is not misleading'[19]. All these new requirements do not innovate in terms of disclosure strategy, which remains based on long duties to provide information to some impersonal non-differentiated addressee (e.g. the average consumer). Rather, they rest on traditional and disproved assumptions: that individuals will read the disclosures by just using plain and intelligible language, or by putting the information in a given place of the platform's websites. For general terms and conditions this is requested by Articles 3 and 5, EU Regulation 2019/1150, and for rankings parameters by Article 6, CRD.

Based on previous massive evidence, however, we should expect that this avalanche of new information duties would not escape failure. For this paper argues that computer science solutions should be used 'comprehensively', that is: to tackle failures at the rule-making phase jointly with the implementation phase. Before illustrating how to implement this 'comprehensive approach' in practice (Phase Two), in the following we elucidate on our methodological approach.

[16] CRD Alinea 45, and Art. 6(1)(ea) Lit. 'based on automated decision-making and profiling of consumer behavior.'.

[17] CRD Art. 6a(1)(a). The rationale is of course to ensure that reviews, on which rankings are based, originate from real customers, real purchasing experiences, no sponsorship nor contractual ties are supporting the reviews, and that no technical manipulation of the results occurred (Alinea 47). Such information should be 'made available in a specific section of the online interface that is directly and easily accessible from the page where the offers are presented'. (Ibid). The omission of such information may amount to an Unfair Commercial Practice (UCP) under Art. 7(4a) UCP Directive No. 2005/29/CE (Annex I, Nos. 23b, 23c).

[18] Art. 5, CRD.

[19] Section 4(a)(1)(A), H.R. 3611, 117th Cong.

2.2. *Tackling failures at rulemaking and implementation stages*

The idea behind the 'comprehensive approach' is to address failures at all levels through a two-step methodology. First we use text analysis to tackle failures of rules and policies; then we employ behavioral testing.

There is a reason if we separate this in two phases. Given the current state of art, algorithmic tools exist that allow intervening on texts and helping measuring failures pertaining specifically to the drafting of disclosures rules as well as firm-level disclosure policies in terms of readability, informativeness, and coherence (Phase One).

On the other hand, text analysis is not (yet) a good tool to identify and measure consumer behavior, nor the industry reaction to it. However, behavioral data is needed to assess how consumers and firms interact with the respective documents, and assess disclosures effectiveness with a view to overcome these types of failures. Hence, we will address the issue of how behavioral data can be generated and used in an inclusive and efficient way in a different part of the Article, by taking inspiration from the 'regulatory sandbox' model (Phase Two).

For completeness, one should mention a strain of literature that seems to consider both text and behavioral elements of disclosure, by addressing changes in the privacy disclosure based on rules from the GDPR as well as consumer expectations (behavioral element). More specifically, Schwartzhneider et al. (2018) claim that big disorders (like the Cambridge Analytica scandal) depend on the 'mis-alignment' between privacy notice and consumers' expectations (a behavioral element) regarding those notices and contend that such failure could be avoidable if both (i) a 'coherent flow' of information was identifiable between rules (principles level) and disclosure policies, and if (ii) the (average) consumer was not 'overwhelmed by the legal[istic] language.' Another noteworthy example is the work of Gluck et al. (2016) who link textual failures of overly lengthy privacy policies with behavioral elements (like the negative framing of disclosures offered to consumers).

In both cases, however, what lacks is a full picture, capable of capturing and measuring all failures of disclosures (not just length or legalistic language) at all different stages: the drafting of disclosure rules, their firm-level implementation, and behavioral failures when consumers are exposed (as well as their interactions).

3. *Part two: Implementing the 'comprehensive approach'*

3.1. *Phase one: Getting to hypothetically optimal disclosures (HOD)*

3.1.1. *Mapping texts*

Lawmakers should begin by creating three datasets composed of disclosure rules (Sect. 1), firm-level policies (Sect. 2), and the case-law pertaining to both (Sect. 3). Again, the domains of online privacy and online contract terms are taken as examples.

1. Disclosure rules as dataset: the *De Iure* disclosures

For the sake of simplicity, we term *De Iure disclosures* all rules where disclosure duties are set. In the privacy context, requirements to platforms to disclose information to individuals regarding their rights, how their data are collected and treated would fit this category. Just as examples we may quote: Sec. 1798.100(a) CCPA, which stipulates the duty of 'a business that collects a consumer's personal information [to] disclose to that consumer the categories and specific pieces of personal information the business has collected'. GDPR Art. 12 requires data-controllers to provide similar information to data subjects.

Technically speaking, these rules can be understood as datasets (Livermore and Rockmore), that can be retrieved and analyzed through NLP techniques (Boella et al. 2013, 2015), easily searched (e.g. via the Eur-lex repository), modelled (e.g. using LegalRuleML) (Governatori et al. 2016; Palmirani and Governatori 2018), classified and annotated (e.g. through the ELI annotation tool)[20] For instance, the PrOnto ontology has been developed specifically to retrieve normative content from the GDPR (Palmirani et al. 2018).

While rules may be clear in stating the goals of required disclosure, it may well be that convoluted sentences or implied meaning appear that make the stated goal far from clear. Also, the same rule may sometime prescribe a conduct with a nice level of detail (if X, then Y), but it may include provisions that require, for instance, that information about privacy shall be given by platforms in 'conspicuous, accessible, and plain language'[21]. Even if governmental regulation is adopted specifying what these terms mean, they would not escape interpretation (Waddington

[20] The European Legislation Identifier (ELI) is available online at: https://eur-lex.europa.eu/eli-register/resources.html.

[21] Section 4(1)(a) H.R. 3611, 117th Cong. (above, note 5).

2020), and thus possible conflicting views by the courts[22].

To help attenuate these problems, proposals have been made to use NLP tools to extract legal concepts and linking them to one another, e.g. through the combination of legislation database and legal ontology (or knowledge graph). Boella et al. (2015) suggest using the unsupervised TULE parser and a supervised SVM to automate the collection, classification of rules and extraction of legal concepts (in accordance with Eurovoc Thesaurus). This way, the meaning of legal texts will be easier to understand, making complex regulations and the relationships between rules simpler to catch, even if they change overtime. Similarly, LegalRuleML may be used to specify in different ways how legal documents evolve, and to keep track of these evolutions and connect them to each other.

2. Firm-level disclosure policies as dataset: the *De Facto* disclosures

The second dataset is that of firm-level disclosure policies, that we term *De Facto disclosures*. The latter include but are not identical to the notices elaborated by the industry to implement the law or regulations. We refer to the overly-famous online Terms of Services (commonly found online and seldom read). With regard to privacy policies, pioneering work in assembling and annotating them was undertaken by Wilson et al. (2016), resulting in the frequently used 'OPP-155' corpus. Indeed, ML is now standard method to annotate and analyze industry privacy policies (Sarne et al. 2019; Harkous et al. 2018).

3. The thinking role of case-law

The case-law would play an important role, serving as the missing link between legal provisions and their implementation. Indeed, courts' decisions help detect controversial text and provide clarification on the exact meaning to give both *De Iure* and *De Facto* disclosures. It follows that case outcomes and rule interpretation should be used to update the libraries with terms that can come out as disputed, and others that can become settled and undisputed[23].

A good way to link the case law with rules is that proposed by Boella et al (2019) who present a 'database of prescriptions (duties and prohibitions),

[22] See below §3.

[23] For instance, the 1985 US Supreme Court Zauderer v. Office of Disciplinary Counsel case established a rational basis review standard triggered by a provision requiring "factual and uncontroversial information" in the disclosure regulation. This is one example of how case law can link terms in the *de iure*disclosures and the corresponding provisions in the *de facto*-disclosures. (Brannon 2019).

annotated with explanations in natural language, indexed according to the roles involved in the norm, and connected with relevant parts of legislation and case law'.

In the EU legal system, a question might arise if only interpretative decisions by the European Courts or also those of national jurisdictions should be included in the text analysis, given that the first would provide uniform elucidation that binds all national courts (having force of precedent), but most case-law on disclosures originates from national controversies and does not reach the EU courts. We know, for instance, that the EU jurisprudence saves to global platforms only a minor part of the costs they spend in controversies with consumers; the paramount ones are those platforms bear for litigations hold before national jurisdictions[24], where there is no binding precedent, and the same clause can be qualified differently.

Moreover, differently from the US[25], in Europe, only the decisions by the EU Courts are fully machine-readable and coded (Panagis et al. 2017)[26], while the process to make national courts' ones also so is still in the making (it is the European Case Law Identifier: ECLI)[27], although at a very advanced stage. Nonetheless, analytical tools are already available that allow to link the EU to national courts' cases. For instance, Agnoloni et al 2017 introduced the BO-ECLI Parser Engine, which is a Java-based system enabling to extract and link case law from different European countries. By offering pluggable, national extension, the system produces standard identifier (ECLI or CELEX) annotations to link case law from different countries. Furthermore, the EU itself is increasingly conscious of the need to link European and national case law, resulting, for instance, in the EUCases project which developed a unique pan-European law and case

[24] One easy way to measure this is to estimate the costs platforms spend to insure themselves against the risk of lost controversies (and distinguish between EU and national ones).

[25] See the 'Caselaw Access Project', providing (free) access to the published decisions from nearly all US State and Federal Courts: https://case.law/.

[26] Texts of all judgments of the European courts can be downloaded for free from EUR-Lex (https://eurlex.europa.eu/homepage.html).

[27] European Case Law Identifier (ECLI) is a computer readable and processable code that can be assigned to every judicial decision from every national or European court. Having an ECLI code assigned ensures that the database is indexed by the ECLI Search Engine, which is based on XLM, on an open source basis. ECLI 'facilitates automated linking of judgments to each other, to other legal sources or to academic writings'. See http://www.bo-ecli.eu/ecli/current-implementation.

law Linking Platform[28].

As shown by Panagis (cit.), of algorithmic tools, citation network analysis in particular, can be extremely useful in addressing not only the question of which is the valid law but also which preceding cases are relevant as well as how to deal with conflicting interpretations by different courts. The latter is especially relevant in systems where there is no binding precedence (i.e. most national EU legal systems) and where, consequently, differing interpretations of certain ambiguous terms might arise. By combining network analysis and NLP to distinguish between different kind of references, it might be possible to assess which opinions are endorsed by the majority of courts and could thus be considered the 'majority opinion'. While other methods to analyze citations in case law might establish the overall relevance of certain cases in general, only the more granular methodology suggested by Panagis et al. seems well fit to assess which interpretations of certain ambiguous terms are "the truly important reference points in a court's repository". In this way, case law can be used to link the general *de iure* disclosures and the specific *de facto* disclosures while duly taking into account different interpretations of the former by different courts.

3.1.2. *Mapping the causes of failure*

To measure the causes of failure of both *de iure* and *de facto* disclosures is not an easy task (Costante et al. 2012). Nevertheless, quantitative indices are indispensable to conduct the following analysis, to make the information they store easily accessible and readable for machines and algorithms. Also, such indexes guarantee the repeatability and objectivity required for the sake of scientific validity.

In line with our 'comprehensive approach', for each stage, failures must be identified, mapped and linked with the failures at other stages, since these are inherently intimately related.

Therefore, we propose defining a standard made of three top-level categories of failure that can be used for both *de iure* and *de facto* disclosures:

1- Readability. *Length of text* can be excessive leading to *information overload.*

2- Informativeness. *Lack of clarity* and simplicity can lead to *information overload*. But also the *lack of information* can result in asymmetry.

[28] EUCases LLOD, available at: http://www.eucases.eu/start.html.

3- Consistency. *Lack of same lexicon* and *cross reference* in the same document or across documents that may lead to incoherence.

Based on these three framework categories, we establish golden standard thresholds and rank clauses as optimal (O) or sub-optimal (S–O) (Contissa et al. 2018a). This way, we would for instance, rank as S–O a privacy policy clause under the 'length of text' index, if it fails to achieve the established threshold under the goal of 'clarity' as stated in the GDPR Article 12. At the same time, however, Article 12 or some of its provisions—as seen—may score S–O under other failure indexes, such as lack of clarity (vagueness). The case-law might help clarify whether this is the case.

	Relevant failure index	**Proxy**	**Methodology and Ranking (O/S–O)**
Readability	Information overload	Length of text	No. of polysyllables on the basis of the length of the text *Rank*: e.g. if longer that X words (golden standard), then rank S–O *ALGO*: SMOG; Dale–Chall readability formula; Gunning Fog Index *Major Ref.* Bartlett et al. (2019)
Informativeness	Information overload	Complexity of text	*Syntactic*: No. of certain grammatical structures (nodes) containing complex text (e.g. conjunctive adverbs—however, thus, nevertheless—passives, modal verbs—could, should, might) *Rank*: E.g. if number of nodes containing complex tokens in clause is higher than X per sentence of a Y length (golden standard), than rank S–O *Major Ref.* Botel and Granowsky (1972) or Szmrecsanyi (2004) *Semantic*: use of complex, difficult, technical or unusual terms called 'outliers' (e.g. 'as necessary', 'generally') or of two or more semantically different CI parameters in information flows *Rank*: E.g. if clause contains more outliers than the number set in golden standard, then rank as S–O *ALGO*: LOF, CI in information flows *Major Ref.* Bartlett et al (2019) Shvartzshnaider et al. (2019)
	Information asymmetry	Lack of information	Presence of all information required by the law (e.g. identity of data controller, types of personal data collected; goals of treatment, etc.) *Rank* E.g. if clause omits more elements than all those necessary according to golden standard, then rank as S–O *Major Ref.* Liepina et al (2019) and Costante et al (2012) /or Contissa et al. (2018a, b)
Consistency	Internal and External	Interaction amongst clauses within the same text and across texts	Recurrence of same lexicon and cross reference between different clauses in the same document and across documents *Rank*: E.g. if a clause scores lower than the citation network gold standard for cross-reference links or evaluation of textual similarity, then rank as S–O *Major Ref.* Panagis et al. (2017) [citation net]; or Nanda et al. (2019) [similarity models]

In the following, we elaborate the methodology for designing a detailed system of indexes to capture the main causes of failure. Furthermore, we provide ideas on how to translate each indicator into quantitative, machine-readable indices. Table 1 summarizes our findings.

1. Readability. Information overload: length of text

The first quantitative index is readability. It is mainly understood as non-readership due to information overload, and measured in terms of 'length of text'. There is a large variety of readability scores (Shedlosky-Shoemaker et al. 2009), based on the length of text which are frequently highly correlated, thus 'easing future choice making processes and comparisons' between different readability measures (Fabian et al. 2017).

Among the many, we take Bartlett et al. (2019) proposing an updated version of the old (1969) SMOG. Accordingly, annotators establish a threshold of polysyllables (words with more than 3 syllables) a sentence may contain, in order to be tagged as unreadable by the machine[29], and hence S–O. The authors suggest 'a domain specific validation to verify the validity of the SMOG Grade'.

This is especially relevant to make our proposal workable. Not all domains are the same and an assessment of firm-level privacy policies would clearly require to be made in each sector. For instance, the type of personal data a provider of health-related services collects would be treated differently from those of a manufacturer retailer dealing with non-sensitive data.

Under the Readability-Length of text index, sub-optimal disclosure clauses use more polysyllables than those established in the golden standard, set and measured using the revised version of SMOG proposed by Bartlett et al (2019).

2. Informativeness. Information overload: complexity of text

Lack of readability of disclosures may also depend on the complexity of text. The scholarship has suggested to measure it from both a semantic and syntactic points of view.

(a) Syntactic complexity

While most analyses of readability focus on the number of words in a

[29] A policymaker might decide to attach a legal effect, e.g. by establishing that the consumer would be bound only if the number of polysyllables is lower than a fixed threshold. *Ibid.* p. 9.

specified unit (e.g. a sentence, paragraph, etc.) as a proxy for complexity, only few authors focus on analyzing the syntactic complexity of a text separately (Botel and Granowsky 1972). Although some scholars search for certain conditional or relational operators, they usually do so with the aim of detecting sentences that are semantically vague or difficult to understand (e.g. see Liepina et al. (2019): see next para.).

Going back to Botel and Granowsky, they propose a count system which designates a certain amount of 'points' to certain grammatical structures, based on their complexity (the more complex, the more points). For instance, conjunctive adverbs ('however', 'thus', 'nevertheless', etc.), dependent clauses, noun modifiers, modal verbs ('should', 'could', etc.) and passives will be assigned one or two points respectively, whereas, for instance, simple subject-verb structures (e.g. 'she speaks') receive no points.[30] The final complexity score of a text is then calculated as the arithmetic average of the complexity counts of all sentences.[31]

An alternative approach is that of Szmrecsanyi (2004), who proposes an 'Index of Syntactic Complexity', which relies on the notion that 'syntactic complexity in language is related to the number, type, and depth of embedding in a text', meaning that the more number of nodes in a sentence (e.g. subject, object, pronouns), the higher the complexity of a text.[32] The proposed index thus combines counts of linguistic tokens like subordinating conjunctions (e.g. 'because', 'since', 'when', etc.), WHpronouns (e.g. 'who', 'whose', 'which', etc.), verb forms (finite and non-finite) and noun phrases.[33].

Although this might be 'conceptually certainly the most direct and intuitively the most appropriate way to assess syntactic complexity', it is pointed out that this method usually requires manual coding.[34]

Since at least the last two measures seem to be highly correlated,[35] choosing among them might in the end be a question of the computational effort associated with calculating such scores.

[30] *Ibid.*, p. 515.
[31] *Ibid.*, p. 515.
[32] *Ibid.*, p. 1034.
[33] These features are used to calculate the final complexity score as follows: ISC*(u)* = 2 × *n(u, SUB)* + 2 × *n(u, WH)* + *n(u, V F)* + *n(u, NP)*, which has been called a rather ad-Hoc-solution by the author himself. *Ibid.*, p. 1035.
[34] *Ibid.*, p. 1031. The article cited was published 16 years ago, therefore, some progress in the automatization of measuring syntactic complexity might have been made in the meanwhile.
[35] *Ibid.*, p. 1037.

Under the Informativeness-Syntactic complexity Index, S–O disclosure clauses (of a given length) use a number of complexity nodes that is higher than the standard, defined and measured using Botel and Granowsky (1972) or Szmrecsanyi (2004).

(b) Semantic complexity

Semantic complexity (or the use of complex, difficult, technical or unusual terms called 'outliers') is analyzed by Bartlett et al. (2019) who use the Local Outlier Factor (LOF) algorithm (based on the density of a term's nearest neighbors) to detect such terms.

Approaching the issue of semantic complexity from a slightly different angle, Liepina et al. (2019) evaluate the complexity of a text based on four criteria: (1) indeterminate conditioners (e.g. 'as necessary', 'from time to time', etc.), (2) expression generalizations (e.g. 'generally', 'normally', 'largely', etc.), (3) modality ('adverbs and non-specific adjectives, which create uncertainty with respect to the possibility of certain actions and events') and (4) non-specific numeric qualifiers (e.g. 'numerous', 'some', etc.). These indicators are then used to tag problematic sentences as 'vague'.

In a similar vein, Shvartzshnaider et al. (2019) base their assessment of complexity/clarity on tags, however, in a different manner. They analyze the phenomenon of 'parameter bloating', which can be explained as follows: building on the idea of 'Contextual Integrity' or CI and information flows,[36] the description of an information flow is deemed (too) complex (or bloated) when it 'contains two or more semantically different CI parameters (senders, recipients, subjects of information, information types, condition of transference or collection) of the same type (e.g., two senders or four attributes) without a clear indication of how these parameter instances are related to each other'[37] This results in a situation where

[36] An information flow denotes the transmission of information from one actor to another. The concept of 'contextual integrity' (CI) is based on the notion that the assessment of an information flow's implications requires information on the full context of the flow, with the latter being operationalized by five CI parameters (senders, recipients, subjects of information, information types, condition of transference or collection).

[37] To illustrate this issue, consider the following fictive clause: 'Advertisers, app developers and specified partners <*three senders*> can request information on the content uploaded by you and your friends <*two subjects*> as well as your interactions with other pages <*two types of data*> '. While this clause might seem straightforward at first, a plethora of different information flows are conceivable based on the large number of different parameter values provided, thus keeping the consumer in the dark concerning the precise flow of her information. Shvartzshnaider et al. (2019), 164.

the reader must infer the exact relationship between different actors and types of information, which significantly increases the complexity of the respective disclosure (at 164).

Therefore, the number of possible information flows might be used as a quantitative index to measure the semantic complexity of a clause.

Under the Informativeness-Semantic complexity Index, a S–O disclosure clause contains more outliers or semantically different CI parameters in information flows than the number set in the golden standard, defined and measured using Bartlett et al (2019) or Shvartzshnaider et al. (2019) respectively.

3. Informativeness. Information asymmetry: lack of information

Another failure index of information asymmetry is the completeness of the information provided in a disclosure. Comprehensiveness has been investigated mainly at the firm-level disclosure policies, rather than the rulemaking (Costante et al. 2012).

It must be noted that the requirement of completeness does not automatically counter readability. While an evaluation of the completeness of a disclosure clause is merely concerned with the question whether all essential information requested by the law is provided, readability problems mostly arise from the way this information is presented to the consumer by the industry. Therefore, a complete disclosure is not per se unreadable (just as an unreadable disclosure is not automatically complete) and the two concepts thus need to be separated.

Several authors suggest tools to measure completeness, especially in the context of privacy disclosure. However, nothing impedes to transfer the approaches presented in this section to disclosures like terms and conditions of online contracts.

For instance, based on the above-outlined theory of CI, Shvartzshnaider et al. define completeness of privacy policies as the specification of all five CI parameters (senders, recipients, subjects of information, information types, condition of transference or collection). Similarly, Liepina et al. (2019) consider a clause complete if it contains information on 23 pre-defined categories (i.e. '<id> identity of the data controller, <cat> categories of personal data concerned, and <ret> the period for which the personal data will be stored'). If information that is considered 'crucial' is missing, the respective clause is tagged as incomplete. Manually setting the threshold would then help define if a clause scores as optimal or not.

A similar, but slightly refined approach is presented by Costante et al. (2012, at 3): while they also define a number of 'privacy categories' (e.g.

advertising, cookies, location, retention, etc.), their proposed completeness score is calculated as the weighted and normalized sum of the categories covered in a paragraph.120F.

For our purposes, a privacy or online contract disclosure clause could be ranked using the methodology suggested by Costante et al (2012) and Liepina et al., or alternatively, by Contissa et al. In both cases, however, corpus tagging would be necessary.

Under the Informativeness-Lack of information Index, in sub-optimal disclosure clauses (of a given length) the number of omitted elements is higher than the pre-defined minimum necessary standard, defined and measured using Liepina et al (2019) and Costante et al (2012) or Contissa et al. (2018a, b).

4. Consistency of documents

One of the two root causes of failure identified above concerns the misalignment of the regulatory goals behind the duty to disclose certain information (as stated in the *de iure* disclosures) and the actual implementation thereof in the *de facto* disclosure. The general criterion that can be derived from this is that of *consistency*, which can be translated into two sub-criteria: internal and external. However, since their measure and computability are identical, they will be treated together.

Internal consistency denotes the recurrence of the same lexicon in different clauses of the same document as well as the verification of cross-references between different clauses within the same document. External coherence means the crossreferences that refer to clauses contained in different legal documents. External coherence too can be understood both as the recurrence of the same lexicon across referred documents and the verification of the respective cross-references. For instance, one rule in the GDPR might refer to others both explicitly (e.g. Article 12 recalling Article 5) or implicitly (like the Guidelines on Transparency provided for by the European Data Protection Board)[38]; or a privacy policy might refer to a rule without expressly quoting its article or alinea in the article.

Unfortunately, there is no common, explicit operationalization of internal and external coherence in the literature.

A first attempt to analyze cross-references in legal documents is made by Sannier et al. (2017), who develop a NLP-based algorithmic tool to automatically detect and resolve complex cross-references within legal

[38] See for instance the Edpb's Guidelines on Transparency under the GDPR of 11 April 2018, available at https://edpb.europa.eu/our-work-tools/general-guidance/guidelines-recommendations-best-practices_en.

texts. Testing their tool on Luxembourgian legislation as well as on regional Canadian legislation, they conclude that NLP can be used to accurately detect and verify cross-references (at 236). However, their tool would allow to construct a simple count measure of unresolved cross-referenced, which might serve as a basis for the operationalization of internal and external coherence, both in terms of the lexicon used (see above, 2nd cause of failure: complexity of text), and the correct referencing of different clauses (see above, 3rd cause of failure: lack of information). Nevertheless, this is far from the straightforward, comprehensive solution one might wish for.

A solution could be to rely on more complex NLP tools such as 'citation networks', as proposed by Panagis et al. (2017) or 'text similarity models', as suggested by Nanda et al. (2019).

The citation network analysis tool by Panagis et al. (2017), seems particularly straightforward, since it uses the Tversky index to measure text similarity. Therefore, using a tool such as theirs would automatically cover both the verification of crossreference links as well as an evaluation of the textual similarity of the cited text.

Another promising option to capture text similarity is the model proposed by Nanda et al. (2019), who use a word and paragraph vector model to help measure the semantic similarity from combined corpuses.[39] After manually mapping the documents (rules provisions and respective policy disclosures), the corpuses are automatically annotated helping to establish the gold standard for coherence. Provisions and terms in the disclosure documents would then be represented as vectors in a common vector space (VSM) and later processed to measure the magnitude of similarity among texts.

This last two models especially come with the advantage of capturing the distance in implementation of rules-based disclosures by the industry policies. They seem therefore very promising in the aim of measuring both the distance in lexicon as well as the presence of cross-reference within the same disclosure rule or policy (and define the gold standard).

Under the Consistency Index, a sub-optimal disclosure clause scores lower than the gold standard for cross-reference links or lexicon similarity, measured using either the citation network tool by Panagis et al. (2017) or the similarity model by Nanda et al. (2019).

[39] Although it is meant to measure legal transposition of EU Directives by national legislations (especially those of Italy, Ireland and Luxembourg), the model can be adapted to capture similarities between disclosure duties and their transpositions.

3.1.3. *Getting to hypothetically optimal disclosures (HOD) through ontology*

Preparing the texts in the *de iure* and *de facto* data sets means that we process the disclosures in each domain to rank them, thus collecting those that score optimal for each failure index. More specifically, per each clause or text partition of the disclosures in each (*de iure* and *de facto*) dataset, processing for the five analyzed indexes will provide a score, allowing to identify a set of optimal disclosure texts (see Table 2). So for instance, we should be able to select the optimal disclosure provision in the GDPR as far as its 'readability' index is concerned. The same should be for the clause of a privacy policy implementing that provision in a given sector (like e.g. the short term online home renting): imagine that is 'Clause X' of AirBnB disclosure policy. The two would form the 'optimal pair', under readability, of *de iure* and *de facto* privacy disclosures in the short-term online home renting sector. The same should be done for all clauses and each failure index.

The kind of coding (whether done manually or automatically) and training to employ, clearly depends on the methodology that will be chosen to perform for each of the failure indexes sketched above. In any event, labelling the disclosures might require some manual work by legal experts in the specific sector considered.

The next step is to link the two selected 'optimal pair' of *de iure* and *de facto* disclosures in the data sets, to reach a sole dataset of what we should term Hypothetically Optimal Disclosure, or HOD.

While, theoretically, a simple, manually organized, static database could be used to do so, the Law & Tech literature suggests a significantly more effective and flexible solution: the use of an ontology/knowledge graph (Shrader 2020; Sartor et al. 2011; Benjamins 2005)10 (Table 3).

As discussed above (I.A.1), legal ontologies are especially apt in this purpose, because they allow automating the extraction and linking of legal concepts, and to keep them up to date even if they change overtime (Boella et al 2015). Another reason is that some ontologies allow to link legal norms with their implementation practices, a feature that is relevant to us.

A good model for linking texts through ontology is provided for by the Lynx project[40] (Montiel-Ponsoda and Rodríguez-Doncel 2018). Lynx has developed a 'Legal Knowledge Graph Ontology', meaning an algorithmic technology that links and integrates heterogeneous legal data sources such as legislation, case law, standards, industry norms and best practices.[41]

[40] http://lynx-project.eu/.

[41] http://lynx-project.eu/doc/lkg/ The Knowledge Legal Graph ontology reuses sources already available on an open access basis, as well as their metadata (such as the afore-

Lynx is especially interesting as it accommodates several ontologies able to provide the flexibility required to include additional nodes anytime rules or policies change.

To adapt the Lynx ontology to our needs, manual annotation to establish structural and semantical links of *de iure* and *de facto* disclosure datasets would nonetheless be needed. That should be done taking into consideration the results of the ranking process, upon which optimal disclosure pairs are selected (Table 2, above). Hence, manual annotation in ontology would consist in functionally linking of only the latter texts, based on semantic relations between their contents.

In our model, nodes will be represented by the failure criteria sketched above. These nodes are already weighted as Optimal/Sub-Optimal and thus given a specific relevance, which allows an analytically targeted and granular nuancing of the ontology.

Table 2 – Example of ranking of disclosure pair leading to HOD, based on failure index

Failure criteria → ↓ Disclosure pair			Readab (length)	Informativeness			Consistency	Ranking and HOD
				Syntactic compl	Semantic compl	Lack of info		
Rule	De iure disclosure	De facto disclosure						
CRD Art. 6	Partition X, Art. 6a(1)(ea) (info on personalized prices)	AirBnB Portion Y ToS policy	Optimal (score 1)	Optimal (score 1)	Optimal (score 1)	Optimal (score 1)	Optimal (score 1)	HOD
		Expedia Port. W	Optimal (score 1)	Optimal (score 1)	Optimal (score 1)	S–O (score 0)	Optimal (score 1)	Not incl. in HOD
		Booking Port. Z	S–O (score 0)	Optimal (score 1)	S–O (score 0)	S–O (score 0)	S–O (score 0)	Not incl
		VRBO Port. XY	S–O (score 0)	S–O (score 0)	S–O (score 0)	S–O (score 0)	S–O (score 0)	Not incl

A further step consists in the assessment of the overall 'coherence' of HOD ontology. Coherence in this context is understood as a further failure index, consisting of *Lack of cross-reference between the Optimal principles-level rule and the corresponding Optimal implementing level policy* (Table 3).

Table 3 Using ontology to get to the hypothetically optimal disclosures (HOD)

Once De Iure and De Facto datasets prepared: Matching (linking) and Ranking through Knowledge Graph/Ontology, leading to 'HOD'		
Coherence/overall	Cross-validation amongst clauses across datasets	Verification of cross-referencing between principles-level (*de iure*) and application level (de facto) leading to incoherence *Rank* If 'clause-pair' scores lower than the gold standard for cross-reference links, then rank S–O *ALGO*: Lynx Legal Knowledge Graph Ontology + manual annotation *Major Ref.* Alschner and Skougarevskiy (2015)

mentioned ELI codes of EU case law) and other ontologies (a full list of which is available here: http://lynx-project.eu/data2/refer-enceontologies).

After manual annotation, to cross-validate amongst clauses across datasets, this process would help to further verify if there is cross-reference between the optimal pairs, or between the principles-level of the *de iure* disclosure and the application level of the *de facto* disclosures, given that they might come from policies drafted by different firms.

A solution could be to rely on 'citation networks', as proposed by Alschner and Skougarevskiy (2015). Focusing on the lexical component of coherence, citation network would help to calculate the linguistic 'closeness' between different, cross-referenced documents[42] and to assess their coherence.[43]

This way, we will be able to give evidence to the overall optimal linked disclosures (i.e. showing the highest scores assigned to each and every pair per single sector domain) and hence to validate the overall coherence of HOD per given domain.

In conclusion, out of the linked data ontology HOD, we should be able to select the texts that fail the least, under a comprehensive approach. These are linked texts, made of the optimal rules (disclosure duties), linked to their optimal implementations (policies), whose terms are clarified through the case law and that score optimal for each and every failure index.

HOD are self-executing algorithmic disclosures, which specifications can be used by the industry to directly implement their content. This however opens a plethora of legal and economic questions regarding their efficacy, legitimacy and proportionality.

3.1.4.*Mapping the causes of failure*

HOD are selected that are the optimal available algorithmic disclosures, but they are still prone to failure. We do not know how effective they might be in leading to behavioral change; how well they could inform real consumers and have them make a sensible choice (for a skeptical take: Zamir and Teichman 2018), given their diverse preferences (Fung et al. 2007). We do not have evidence if the optimal disclosure text regarding a given clause will perform well or not. For instance, imagine we are ranking disclosures in the short-term online renting sector, and that the HOD regarding information provision on the service ranking indicates that the optimal pair is "CRD Art. 5"—"AirBnB Terms, Clause X": what do we know about its efficacy? The HOD cannot tell.

[42] In Alschner and Skougarevskiy's work citation network allowed to compute the 'textual distance' between 1623 Bilateral Investment Treaties.

[43] Which the authors define as 'close mutual distances' between two treaties.

Moreover, since the comprehensiveness of the proposed approach implies that HOD might complement or even partially substitute tasks that would normally be executed or at least supervised by democratically elected representatives, concerns of legitimacy arise. In the example done, once the optimal pairs identified through the HOD, the idea is that "CRD Art. 5"- "AirBnB Terms Clause X" would be automatically implementable. However, that would be problematic under legitimacy terms.

Lastly, HOD may lack proportionality, since they are addressed to undiversified, homogeneous consumers (the average ones), based on assumption of homogenous reading, understanding, evaluation, and acting capabilities (Di Porto and Maggiolino 2019; Casey and Niblett 2019). However, the same disclosure may well be excessively burdensome for less cultivated consumers, while being effective for well-informed, highly literate ones.

In the following, we explore these three issues separately.

1. Untested efficacy of HOD

Although they are hypothetically optimal inter-linked texts, constantly updated with new rules, industry policies, and case-law, easily accessible and simplified, not so costly to read and understand, the overall efficacy of HOD remains untested.

On this land stand the enthusiasts, like Bartlett et al. who purport that the use of text analysis algorithmic tools, which summarize terms of contracts and display them in graphic charts, 'greatly economize[s] on [consumers'] ability to parse contracts' (Bartlett et al. 2019). However, they do not provide proof that this is really so (if one excludes the empirical evidence supporting their paper). Paradoxically, the same holds for those who oppose the validity of simplification strategies and information behavioral nudging, like Ben-Shahar (2016). They consider that 'simplification techniques…have little or no effect on respondents' comprehension of the disclosure.' But again, this conclusion refers to the 'best-practice they surveyed'.

2. Legitimacy deficit of HOD

HOD suffer from a deficit of legitimacy. Because an algorithm is not democratically elected, nor is it a representative of the people, it cannot *sic et simpliciter* be delegated rulemaking power (Citron 2008, at 1297).

While in a not so far future it may well be that disclosure rules become fully algorithmic (produced through our HOD machine), a completely different question is whether disclosure we have selected as the

hypothetically optimal might also become 'self-applicable', or, in other words, whether their adoption can become one step only, without any need for implementation. This is surely one of the objectives of HOD. By selecting the optimal rules together with the optimal implementation and linking them in an ontology, we aim at having self-implementing disclosure duties.

Hence, it is necessary to re-think of implementation as a technical process, strictly linked (not merged) with the disclosure enactment phase. But especially, we need to ensure some degree of transparency of the HOD algorithmic functioning and participation of the parties involved in the production of algorithmic disclosures.

Self-implementation of algorithmic rules is one of the least studied but probably the most relevant issues for the future. A lot has been written on the need to ensure accountability of AI-led decisions and due process of algorithmic rule-making and adjudication (Crawford and Schultz 2014; Citron 2008; Casey and Niblett 2019; Coglianese and Lehr 2016). However, while some literature exists on transparency and explicability of automated decision-making and profiling for the sake of compliance with privacy rules (Koene et al. 2019), the question of due process and disclosure algorithmic rule-making has been substantially neglected.

However, a problem might exist that the potential addressees of self-applicable algorithmic disclosure rules do not receive sufficient notice of the intended action. That might reduce their ability to become aware of the reasons for action (Crawford and Schultz at 23), respond and hence support their own rights.[44] Also comments and hearings are generally hardly compatible with an algorithmic production of disclosures; while they would be especially relevant, because they would provide all conflicting interests at issue to come about and leave a record for judicial review. The same goes for expert opinions, which are often essential parts of the hearings: technicians may discuss the code, how it works, what is the best algorithm to design, how to avoid errors, and suggest improvements.

In the US system, it is believed that hearings would hardly be granted

[44] See Citron (supra note 152) p. 1284 (noting that the black box nature of algorithms can make their decisions non predictable, or non-fully compatible with the guarantees of due process. In Italy, an algorithm was used by the Ministry of Education to decide upon the allocation of high school chairs among teachers who had won a public selection in 2017. The decision being entirely delegated to an algorithm, it has been challenged before administrative courts and further annulled both on first instance and on appeal on discriminatory grounds (Tar Rome, Decision no. 9230/2018, on appeal Council of State, dec. 13 December 2019, no. 8472).

in the wake of automated decisions because they would involve straight access to 'a program's access code' or 'the logic of a computer program's decision', something that would be found far too expensive under the so-called Mathews balancing test (Crawford and Schultz at 123, Citron at 1284).

In Europe too, firms would most probably refuse to collaborate in a notice and comment rulemaking, if they were the sole owner of the algorithm used to produce disclosures, since that might imply to disclose their source codes, and codes are qualified as trade secrets (thus, exempt from disclosure).

Moreover, as (pessimistically) noted by Devins et al., the chances for an algorithm to produce rules are nullified, because 'Without human intervention, Big Data cannot update its "frame" to account for novelty, and thus cannot account for the creatively evolving nature of law.' (at 388).

Clearly, all the described obstacles and the few proposals thus far advanced are signs that a way to make due process compatible with an algorithmic production of disclosure rules is urgent and strongly advisable.

3. Lack of proportionality of HOD

Although it is undeniable that general undiversified disclosures may accommodate heterogeneous preferences of consumers (Sibony and Helleringer 2015), in practice, they may put too heavy a burden on the most vulnerable or less cultivated ones, while not generating outweighing benefits for other recipients or the society. In this sense, they may become disproportionate (Di Porto and Maggiolino 2019).

On the other side, also targeting disclosure rules at the individual level (or personalizing) (Casey and Niblett), as suggested by Busch (2019), may be equally disproportionate (Devins et al 2017) as can generate costs for the individuals and the society. For instance, if messages are personalized, the individual would not be able to compare information and therefore make meaningful choices on the market (Di Porto and Maggiolino 2019). That, in turn, would endanger policies aimed at fostering competition among products, which are based on consumers' ability to compare information about their qualities.[45] Also, targeting at the singular level requires necessarily to obtain individual consent to process personal data (for the sake of producing personalized messages) and also show one's 'own' fittest disclosure.[46]

[45] *Ibid.*
[46] *Ibid.* p. 23.

3.2. *Phase two: Integrating behavioral data into HOD: getting to the best ever disclosures (BED)*

3.2.1. *Experimental sandboxes to pre-test HOD*

One way to possibly overcome the three claims (ensure transparency, participation, proportionality and efficacy of HOD disclosures) would be to integrate real-time behavioral data into the HOD algorithm and have it produce targeted, yet dynamic (i.e. fed by real-time data) self-implementable disclosures.

To achieve that, we suggest exploiting the potential of 'regulatory sandboxes'. In the following, we articulate how this tool could be used to conduct pretrial tests of HOD algorithmic disclosures. Such experiments serve the triple function of ensuring legitimacy of the algorithmic rulemaking by allowing participation and transparency; producing targeted disclosures to test their efficacy, and granting proportionality by clustering.

1. Regulatory sandboxes

Regulatory sandboxes are not new (Tsang 2019; Mattli 2018; Picht 2018). They exist in the Fintech industry, where new rules are experimented in controlled environments (thanks to simulations run over big data) before being implemented at large scale.[47] For instance, the UK's Financial Conduct Authority adopted a regulatory sandbox approach to allow firms 'to test innovative propositions in the market, with real consumers'.[48] Regulatory sandboxes can be conceptualized as venues for experimenting with co-regulation, in the sense that they foster collaboration between the regulator (which takes the lead) and the stakeholders to experiment with new avenues for rule production (Yang and Li 2018). Given their increasing relevance, they are being disciplined by the forthcoming EU Regulation on Artificial Intelligence (Article 45 ff.).

We argue for a regulatory sandbox model where, under the auspices of the regulator, stakeholders come together to pre-test the HOD algorithm to develop self-implementable targeted disclosure rules for consumers.

2. Pre-testing HOD to meet legitimacy claims

The main takeaway of the above discussion on having an algorithm

[47] See the joint report by ESMA, EBA, and EIOPA, JC 2018 74, FinTech: Regulatory Sandboxes and Innovation Hubs (2019).

[48] Financial Conduct Authority, Regulatory sandbox, 10 February 2020, https://www.fca.org.uk/firms/innovation/regulatory-sandbox (last accessed 16 June 2020).

legitimately producing disclosure rules, is that the human presence is irrepressible. That implies that a straight suppression of any transparency and participation guarantees (for the humans) in algorithmic rulemaking is not admissible.

Using regulatory sandboxes might remedy legitimacy concerns, as stakeholders will participate in real, and contribute to the regulatory process. Of this participation (i.e. of reactions, comments, etc.) data are tracked that feed the algorithm. Indeed, in the sandbox, the regulator sets up an agile group (of consumers, digital firms, legal experts, data scientists) for the ex-ante testing of HOD algorithmic disclosures in the course of a co-regulatory process. As real individuals interact with each other in the sandbox and their true responses to legal notices are registered and fed into the algorithm, they may constitute a good substitute for both notice and comment.

3. Pre-testing HOD for targeting and gather evidence of efficacy

Another reason why HOD disclosures need to be tested with real people in the sandbox is to check if they may actually change the behavior of addressees in the real world: e.g. if optimal acceptance of cookies by those adversely affected increases.

However, as said, to overcome what we consider the main limitation of the current scholarship on disclosure, we deem that experiments should not be occasional, but conducted on a 'real-time basis' and repeated. The sandbox mode is a proxy for real-time evidence of the recipients' actual reaction to the disclosures. The latter will be gathered later, when algorithmic disclosures will be implemented on the market (see below 1.1.3). Nonetheless, the sandbox mode would still greatly increase our understanding of what does not work, but most importantly, would provide behavioral data for reuse in the HOD algorithm to target the messages.

Elsewhere we purported that 'targeted disclosure' helps increase its effectiveness, as it allows to tackle the different groups of consumers showing homogeneous understanding capabilities and preferences with different messages. For instance, we might expect that consumers participating in the sandbox testing may react differently to the HOD-produced privacy disclosure and show different click-through attitudes. This might depend on their literacy, time availability, framing, and other bias. Exposing them to differentiated layouts instead of just one might increase their ability to overcome click-though.

But this needs to be tested. And the reactions of consumers traced

by the algorithm. An example might clarify: only to the extent that targetization of privacy disclosure layouts also becomes optimal, meaning that it helps most consumers in a cluster overcome click-though, can HOD become really optimal, or Best.

4. Clustering to meet proportionality claims

Thus, targeting disclosures at 'clusters' is preferable. However, clustering is not an easy task, since clusters should be made of individuals showing similar preferences (e.g. all those who prefer detailed, long boilerplate of fine-print terms vs those who prefer synthetic warning messages). And humans are nuanced. A criterion should be set to form clusters, that can be either descriptive (what consumers in that group typically want to know) or normative (what they ought to know). Either way it should reflect sufficiently homogeneous cognition capabilities and preferences to reduce information overload and increase disclosure utility (Ben-Shahar and Porat 2021).

If data gathered in the sandbox show that a big group of consumers is especially exposed to the risk of overdue payment, then that could constitute a cluster (and a disclosure rule highlighting the consequences of payment delay, instead of a standardized all-inclusive warning list may be tested).

Only if testing sessions are repeated enough evidence is gathered of individual reaction that allows for clusterization. As known, the more data is gathered on the reaction and interaction of individuals, the easier is for the algorithm to identify clusters, based on its predictive capabilities.

Diversification of rules by clusters allows rulemakers to strike a balance between the use of predictive capabilities of algorithms, while at the same time conceiving of disclosure regulation that is compliant with the proportionality principle. In Ben-Shahar and Porat's words, a mandated disclosure regime that grants different people different warnings to account for the different risks they face gives all people better protection against uninformed and misguided choices than uniform disclosures do. (at 156).

Also, clusterization allows for targeted disclosure to be respectful of privacy and data protection rights, while preserving of innovation and the market dynamics (Di Porto and Maggiolino at 21). But even if personalized rules are permissible under a particular jurisdiction's privacy law, the state may economize by identifying clusters of people who share sufficiently similar characteristics and draft one disclosure rule for them instead of many disclosure rules for each of them.

3.2.2. *Getting to best ever disclosures (BED) through regulatory sandboxes*

1. Governance Design Issues

It is on the rulemaker to propitiate a regulatory sandbox, pooling together experimental groups, which would include, the final consumers, individuals representing digital firms (inclusive of platforms and SMEs,[49] which of course vary depending on the topic of algorithmic disclosures), and technical experts. Special attention should be paid to equal representation of stakeholders in each sector-specific sandbox.[50] As said, the goal of the group is to train the selected algorithm (the HOD) for designing different layouts of the best disclosures (Di Porto 2018).

Repeated sessions of tests and feedback would lead to elaborate, with the agreement of all participants, the final sets of targeted disclosures. The latter, by then would become, very emphatically, the Best Available Disclosures or BEDs, to be deployed at large scale (see below, Sect. 3.2.3).

Indeed, insofar as algorithmic HOD are fed-in with behavioral data on the reactions of real people (in an anonymized and clustered format), they could become differentiated, targeted, and timely, thus meeting the different informational needs of recipients (Di Porto 2018 at 509; Busch 2019 at 312). So, for instance, to tackle the problem of online click-through contracts (people do not read standard form contracts before agreeing), one could provide different layouts to different groups of consumers, depending on their reading preferences. These layouts would target clusters of similar consumers and would be derived from behavioral data that was generated only in the sandbox as a separate, isolated environment, but not from data of every individual.

To be more concrete, each disclosure in the HOD algorithm would target each group showing similar characteristics. For instance, three groups, depending on their capabilities, may be detected:

- the modest (to whom a super-simplified format may be preferable),
- the sophisticate (to be targeted through extensive disclosures), and
- the intermediate (a mix of the previous ones).

Testing may prove successful if exposure to the three layouts results in increased reading, understanding and, especially, meaningful choice (e.g.

[49] It is especially important to select these stakeholders in a way that the interests of the business users are well represented before those of the platforms and enough receptive of those of final consumers.

[50] See Expert Group on Regulatory Obstacles to Financial Innovation (ROFIEG), Thirty Recommendations on Regulation, Innovation and Finance, 13 December 2019, at 70.

they start refusing third party tracking cookies).

Every choice the participants make will be tracked during the test, and this data will then feed the algorithm (on the technicalities of such feeding see below), providing it with information on how to produce the best disclosures, meaning those that fail the least to be read, understood, and give due course of action. At each session, new data will be recorded regarding how groups of individuals (firms, the regulator, and consumers) react to the provided information. Also, choices from firms regarding disclosure clauses should be tracked and feed in the algorithm.

So for instance, pre-contractual information regarding the right of withdrawal from distant contracts must be provided to consumers according to new Arts. 6 and 8 of the Consumers Rights Directive.[51] In particular, Art. 8 deals with the information provided on 'mobile devices', stating that notice on the right of withdrawal should be:

'provided by the trader to the consumer in an *appropriate way*'.

In a regulatory sandbox, various messages to provide such information would be tested before consumers and digital firms, meaning that both will respond to the different layouts. All such reactions would be coded, and feedback registered.

Testing is also relevant to implement rapid amendments to the algorithmic disclosures, both the texts and the graphic layouts (i.e. where the information is located in a mobile phone screen, or when is displayed on a mobile device) should reactions of consumers not occur.[52]

To help further this, the regulator should enjoy real-time monitoring powers. Indeed, the pre-testing phase also allows detecting with some precision what are the informational needs and understanding capabilities of the users. In this sense, algorithmic disclosures would produce useful information, by dynamically adapting their content and format to what the cluster recipients need at the time they need.

[51] See Arts. 6 and 8 of CRD, as amended by the New Deal for Consumers Directive 2019/2161.

[52] As noted by the WP29, algorithms are subject to bias and 'can result in assessments based on imprecise projections'. See WP29 (supra, note 138) p. 27. Therefore, it is crucial to 'carry out frequent assessments on the data sets…to check for any bias, and develop ways to address any prejudicial elements, including overreliance on correlations'. Those checks and audits require 'regular reviews of the accuracy Footnote 52 (continued) and relevance of automated decision-making, including profiling … not only at the design stage but also continuously' *Ibid.*, p. 28.

2. Technical issues: using knowledge graph/ontology

Diverse computational techniques could be used to develop algorithmic disclosures.

Like with HOD, also to get to the BED we suggest using a knowledge graph: this way, the textual libraries from the HOD can be enriched with behavioral data coming from the sandbox (hence, we start the process with three libraries).[53] In the knowledge graph, both the text and behavioral data will be integrated employing users' experience. To make a parallel, this operation resembles (but differs) the way Google search engine operates (through domains and supra-domains). When Google users are shown a picture and asked to 'confirm' that what they see is X and not Y, by clicking 'I confirm' they reinforce a node of the graph. Similarly, human stakeholders in the sandbox provide behavioral data that confirm the layout and text of a proposed clause, thus reinforcing nodes, and gradually strengthening the links in our BED knowledge graph.

For instance, per each group of consumers (modest, intermediate, sophisticate), the stakeholders will have to confirm the layout of a clause of privacy disclosure. The confirmation data of each group will feed the knowledge graph. If they see different layouts of cookie banners, confirmation will tell which one performed best in increasing the ability to avoid click-though (Table 4).

Behavioral data coming from the regulatory sandbox are also relevant to confirm or contradict the links and reality described by the graph. The human presence, as said, is essential to monitor if errors occur in the building of the knowledge graph: technicians supervising in the sandbox may intervene to eventually deactivate any error that may affect the algorithm (Yang and Li 2018, at 3267). That explains why we need technicians to participate in the sandbox, besides regulators, firms, and consumers.

Technically speaking, for the knowledge graph to be implemented, we need to connect all the data: the linked texts of the HOD and the behavioral ones coming from the sandbox. All of these data and information shall remain in the knowledge graph.

[53] One key reason why knowledge graphs seem fit to do so 'is that they can provide a common (even if not neutral) language to express' the information of the different libraries. Also, they provide 'at the same time a tool for conceptual retrieval and a model of content which maintains strict references to the text', thus being very much in line with what is required in the context of this proposal. Benjamins (2005), at 116.

Table 4 Using ontology to test HOD in sandbox and get to the Best Ever Disclosures (BED)

Group A of Consumers vs HOD	→ Layout A	→ Confirmation	→ data	BED
	→ Layout B	→ Confirmation	→ data	
	→ Layout C	→ Confirmation	→ data	

To that end, we should use an ontology. The ontology serves to link all the pieces with concepts of the domain, supra-domain, and vertical domain. for instance, imagine we aim to link the term 'fintech' (domain) to the normative goal (supradomain) to a sector-specific term, like 'transparency in financial fintech' (vertical domain).

Because most of the time, rules do not speak in such a detail, we need to use a meta-level to provide further instructions (benjamins 2005, at 39). For instance, very often rules in the financial domain do not require retailers of financial products to disclaim full detailed composition of their products, but would instead require for general transparency. Therefore, we would need to provide a meta-level whereby to instruct the algorithm this way: 'when using the word "rules", link it to the concept "transparency", then link it to "disclaimer".

To sum up, the knowledge graph technology is used to refine the bed by performing the following tasks:

- memorizing the linked texts prepared in the HOD,
- annotating them (through an ontology);
- building a grid of theoretical-legal concepts, specific to a subject and goal, and to a sector, like in privacy.

In conceptualizing the sandbox, we should elaborate on the concepts typical of a specific sector (like privacy or online consumer contracts). To do so, we need to create relationships with a natural language sandbox, which serves to allow humans to participate in the sandbox, to either confirm or reject them. On this basis, we will provide them to the final consumers and the firms (i.e. the stakeholders). By saying that they are 'satisfied' (or 'confirm' the clauses/layouts), they will feed into the sandbox.

This should be repeated in several formats and for several times (sessions) until we get to the point where all participants are mostly satisfied and least dissatisfied. We should repeat this with the clauses of each disclosure per each of the 3 or n layouts we want to target the cluster consumers. In this way, we get to the BED we can implement at large scale.

3.2.3. *Getting to best ever disclosures (BED) through regulatory sandboxes*

1. Automatic implementation of BED at large scale

After the sandbox testing, disclosure should be available, that are targeted at different groups and self-implementable at large scale: these are the BED. The expected output is that the BED algorithm can produce different rules with different messages to convey to each group of consumers (a); on the industry side, BED's specifications will be used for implementation (b); thus, firms' trade secrets will be safe (c).

(a) *Allocation of consumers in the diverse cluster*s

Once the BED algorithm producing automatic disclosure rules is launched on the market (implemented at large scale), users are first allocated a default intermediate group (b). However, they remain free to switch from one group to the other by choosing the preferred disclosure option.

Interactions with the algorithm will produce more data, that will be tracked and help further refining it. Choices made by the consumers between the three (or n) rule layouts and the switches among them, after due pseudonymization, may feedback into the BED algorithm and ameliorate it. On the contrary, individual choices made due to the BED (hence, their effects on a large scale) would not possibly be registered nor further analyzed due to privacy constraints,[54] unless a law expressly authorizes that.[55]

(b) *BED's specifications in lieu of industry-led implementation*

BED algorithmic disclosures are automatically implementable. However, for BED disclosure to be launched on the markets, the industry must make an effort to technically implement its specifications, which are made publicly available. Being the latter sector-specific, and thoroughly discussed among stakeholders in the sandbox, a lot of time and costs for

[54] At the EU level, individual consent to data treatment would be required under the GDPR if the rulemaker wanted to test whether clustered disclosures were effective after implementation on a large scale (unlike the design phase). Even there, mass data treatment would possibly contrast with the principle of minimization of treatment (in this case, by public authorities).

[55] This is the case in the EU thus far: under EU law, consent is not the only legal ground legitimizing automated processing of personal data: Art. 22(2) lit. b) GDPR allows EU or Members states to adopt laws authorizing it, under the condition that same laws 'lay down suitable measures to safeguard the [individual]'s rights and freedoms and legitimate interests.' Hence, a statutory law may be adopted authorizing algorithmic production of disclosures.

producing disclosures will be saved to the industry.

Making specifications open to individuals and firms, is also a means to allow the regulator to monitor the efficacy of algorithmic disclosure. Furthermore, it allows for accountability of the disclosed information and the algorithmic decision.

(c) (continued) *Without disclosing any trade secrets*

Despite a broad consensus on an increased need for transparency when algorithmic decisions are involved, 'it is far from obvious what form such transparency should take' (Yang and Li at 3266). While the most straightforward response to this heightened transparency requirement would probably be the disclosure of the source codes used by firms, this approach is not feasible as the latter are unintelligible to most lay persons and highly secretive.

When adopting a 'regulatory sandbox' solution, however, there would be no need for the platform to disclose any of its own algorithms (which might easily remain secret) to other stakeholders participating in the trials.[56] That is because the kinds of algorithms that are being used to get to the BED are publicly available.[57] The consumers, platforms and SMEs contribute with their behavioral data to feed the BED algorithm: for instance, in case of disclosures of standard form contracts, the experimental sandbox phase would consist of the stakeholders testing different formats of ToCs. Thus, they would be enabled to enhance their disclosures without having to publicize any of their algorithms or similarly sensitive information.

2. Post-implementation modification of the BED

As far as amendments to algorithmic disclosures are concerned, these could be done in the regulatory sandbox, and consequently implemented at large scale in an automated way. This is still another step, different from both the creation of the HOD algorithm, its testing with behavioral data to become BED and the latter implementation at large scale. Suppose we have already a BED algorithm working on the market that produces targeted disclosures for short-term rental service terms. Imagine that a new

[56] Notoriously, algorithms are covered by IPRs (and are usually qualified as trade secrets). Legally speaking, under EU law, firms are not entitled to any general right to be informed about the overall system used to make automatic decisions, nor can they demand the full disclosure of the algorithm: see Recital 27 and Art 5(6) Regulation EU 1150/2019.
[57] See *supra* Sect. I.C (discussing how the HOD is built).

EU Regulation is adopted (e.g. Art 12(1) of the Digital Services Act)[58] amending the CRD and mandating digital providers to inform users about potential "*restrictions*"[59] to their services contained in the terms and conditions in an "*easily accessible format*" and written in "*clear and unambiguous*" language. Such new piece of law would require a refinement of BED, that we suggest doing in the sandbox, instead of starting the whole process from scratch.

This way, all modifications to BED algorithmic disclosures, that participants to the sandbox accept—and the regulator certifies—could become directly implementable by the digital firms on large scale, given that they have been 'pre-tested' in the sandbox. That would comply with the best practice identified by the already mentioned Guidelines of the WP29, and would allow such modifications to feedback into the BED algorithm to ameliorate it and, consequently, the disclosures.

Technically speaking, BED algorithmic disclosures update constantly depending on three factors: changes in the law/regulation or the jurisprudence (in which case the text libraries and nodes in HOD ontology update); change from the sandbox (i.e. update in the behavioral library, and consequently links to the texts through validation/confirmation in BED); change from real-world behavior after implementation on large scale.

To make a step further, one could also think of using sandboxes to "suggest" and "approve" rule modifications. This path would be another innovative, venue: all proposed modifications could be discussed, tested, and approved directly in the regulatory sandbox. That would clearly substitute the usual democratic process. So for instance, the regulator might propitiate and stakeholders in the sandbox agree to change the wording of a rule: they may agree to modify Art. 6a(1)(a), CRD and make information on the ranking parameters of search queries available.

'by means of n icons on the x-side of the presented offer',

instead of

'in a specific section of the online interface that is directly and easily accessible from the page where the offers are presented.'

as it currently is.

[58] European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM(2020)825), 15 December 2020.

[59] Meaning "any restrictions that they impose in relation to the use of their service in respect of information provided by the recipients of the service, in their terms and conditions": DSA, Art. 12(1).

Once the change validated in the sandbox, the BED library is modified accordingly and can thus be directly implemented.

From a legal perspective, post-implementation modifications would not only be self-implementable, but could also be given a special effect: for instance, because they have been pre-tested and validated in the sandbox, they could produce a direct effect (or be enforceable) among the parties, or in some instances provide for safe harbors. For example, an amendment to the disclosure of a certain service's Terms of Contract in a given sector, which is agreed upon in the sandbox, and implemented in the algorithmic disclosure, could become immediately effective. Also, some of its clauses might escape liability.

3.2.4. *Discussion of BED*

1. Choice of algorithm provider

One possible limitation of our BED solution is selecting the algorithm provider. It seems problematic to have private parties providing the algorithm for rule-making purposes, since, as purported by Casey and Niblett, they would inevitably reflect their own interests in the definition of the objectives to pursue (Casey and Niblett, at 357). Also, there may be strong economic incentives for private parties for not disclosing information about how their rulemaking algorithm was created or why some results were generated. For instance, they might want to 'heighten barriers to competition, or favor one side because of repeat-player issues'..[60]

One possibility to overcome rent-seeking and riming rules by firm stakeholders could be for the state to open the provision of BED to competition, similarly to auctions hold for the provision of public goods (Levmore and Fagan 2021). Alternatively, the state could consider undergoing some type of approval process, similar to safety certification. However, that might prove costly.

2. Liability of digital operators

Why would the digital firms want to participate in the BED instead of producing their own disclosures? In the end, they have greater technical skills, knowledge and data about consumers to stay away of BED.

In addition, anything that happens in the sandbox implies some disclosure of trade strategies to the regulator, competitors, and consumers. Information is an asset, and even in the little margins left by the disclosure

[60] Ibidem.

duties, firms might not want to share the way to convey it to their clients.

Also, the BED solution only holds for firms that operate through algorithms and big data technologies, while it leaves aside those not working in the digital sphere (think e.g. to SMEs who lack resources to invest in these technologies).

Moreover, there are industries (like the pharmaceutical) where the BED solution would not possibly be applicable, as full, lengthy, and complete disclosures are needed and not suppressible. Therefore, targeted and summarized information could not work.

While the last objection is insurmountable, one way to eventually commit digital operators to take part in the pre-testing and continue their support in the implementation of BED at large scale is that regulators establish a safe harbor.[61] The safe harbor would work for companies that commit in advance to the terms and clauses of the disclosures agreed upon by the participants in the sandbox (and of course in all subsequent periodical updates). Afterward, if a company fails to qualify for the safe harbor (because, for instance, it does not duly implement the technical specifications provided for in the BED), it may incur additional legal liability in case of litigation, provided that the plaintiff can prove that the disclosure fails the BED standard.

Eventually, one might consider the BED as a "minimum requirement for disclosure compliance" (e.g. at the Federal or EU levels), so that Member states would remain free to set stricter requirements and thus technical specifications to add to the BED. That way, national (member) states would be able to also take account of their own jurisprudence more widely and incorporate it into the algorithm to the level deemed appropriate.

On a more general reputational ground, engaging in the BED project might be convenient for the industry as firms might demonstrate to engage in pro-consumer actions, while at the same time reducing their costs of compliance to disclosure regulation requirements.

3. Are recipients better off?

One possible drawback of algorithmic BED is that they may end up 'offering finite choices to users effectively forc[ing] them to guess the category under which their information falls.' (Citron 2008, at 1300)

[61] This approach has recently been incorporated into the Australian 'Treasury Laws Amendment (2018 Measures No. 2) Bill 2019'. The bill (esp. Section 926B) facilitates the exemptions for firms participating in a regulatory sandbox to test financial and credit products from certain regulations for the time of testing and under certain conditions.

Also, it may well be that consumers are irresponsive, for reasons we are not able to assess, to the algorithmic targeted disclosures. For instance, as not all consumers are prone to intensive online marketing campaigns or dark patterns, it may well be that a noticeable portion of consumers is not becoming aware or that different pieces of information are needed for them in their decision-making process.

If we agree that this might be the case, we acknowledge that there is no evidence unless we try to seek some. And the BED project is especially aimed at providing the consumers with different types of information (instead of just one) to minimize their cognitive effort while maximizing her individual autonomy. As said, to prove the effectiveness of the provided information to also commit to a choice that maximizes her utility, consumers' online behavior ought to be tracked.

4. Which rulemaker?

On the rulemaker side, one limitation is about who—meant as which authority—should be given responsibility for designing and monitoring the applicability of BED disclosures. In our model, being disclosures sector- and topic-specific, the regulator participating in the sandbox would be, each time, the one responsible for the issue at stake. So for instance, if disclosures in the realm of distant contracts for energy provision are being discussed, then the energy regulator (together with the data protection agency) should take the lead of the testing. In a similar vein, the recently created Utah Fintech Sandbox will be administered by the Utah Department of Commerce.[62]

However, if that solution might accommodate national disclosures, where domestic rulemakers might be given legal responsibility for leading the project, one might wonder who should take the lead at the (US) Federal or EU levels. For instance, in Europe, one might wonder whether the Commission enjoys enough political support to do so, eventually with the support of the Jrc. That would also mean, because disclosures are usually written in different languages, that the translation language service of the EU should be included in the project.

[62] A similar program in Arizona, on the other hand, will be supervised by the Arizona Attorney General. KAYE AC (2019) Utah's new regulatory sandbox. Consumer Finance Monitor. Available at: https://www.consumerfinancemonitor.com/2019/06/11/utahs-new-regulatory-sandbox/.

4. *Conclusions*

Modern rulemaking has for centuries been a purely human activity. But algorithms are there to support in ways the legal scholarship has started exploring. This Article has drawn a roadmap to employ NLP and ML tools to help save disclosure regulation failure, its stated goal being to reframe how to create better disclosure rules. To do so, it has addressed three types of challenges: regulatory (why does disclosure regulation fail in the online privacy and consumer transaction contexts?); technical (what algorithms could best tackle both textual and behavioral failures of disclosures at the two, enactment and implementation, phases?); and legal (can algorithms legally produce self-implementing disclosure norms?).

To these, this article has provided solutions elucidating on how to build an algorithm for the linking of existing openly accessible datasets of *de iure* and *de facto* disclosures and then selecting those that fail the least. Further, it has addressed the question of how to attenuate legitimacy problems stemming from lack of democratic representativeness of the algorithm, by integrating elements of collaborative and procedural democracy (using a regulatory sandbox) into a knowledge graph. That, in turn, with the final goal of creating Algorithmic Disclosures, which are self-implementing rules.

In the future work, we intend to analyze how disclosure duties are created at the EU level, to check where possible source of failure might stand. To do so, we plan to use NLP and ML tools to analyze the feedback documents submitted by the stakeholders to the EU consultation process on new disclosure duties contained in the proposed Digital Services Act and Digital Markets Act 2020. This way we seek to identify possible semantic differences in the use and understanding of words that pertain to disclosure duties. If such differences exist, then they may provide fresh evidence of why disclosures fail.

Acknowledgements Funding was provided by Lady Davis Fellowship Trust, Hebrew University of Jerusalem.

References

T. Agnoloni, L. Bacci, M. van Opijnen, *BO-ECLI parser engine: the extensible european solution for the automatic extraction of legal links*, in A. Wyner, G. Casini (eds) *Legal knowledge and information*

systems, proceedings of the 2nd workshop on automated detection, extraction and analysis of semantic information in legal texts, June 16, 2017, London, UK, 2017, pp 113–118. <https://ebooks.iospress.nl/publication/480522017>.

G.A. Akerlof, *The market for "Lemons": quality uncertainty and the market mechanism*, in *Q. J. Econ*., vol. 84, 1970, pp. 488–500

W. Alschner, D. Skougarevskiy, *Consistency and legal innovation in the BIT Universe*, in *Stanford Public Law Working Paper No*. 2595288, 2015, p 2.

K. D. Ashley, D. Kevin, *Artificial intelligence and legal analytics: new tools for law practice in the digital age*. Cambridge University Press, Cambridge, 2017.

I. Ayres, A. Schwartz, *The no-reading problem in consumer contract law*, in *Stan. L. Rev*., vol. 66, 2014, pp. 545–610.

Y. Bakos et al, *Does anyone read the fine print? Consumer attention to standard-form contracts*, in *Legal Stud*, vol. 43(1), 2014, pp. 1–35

O. Bar-Gill, *Consumer transactions*, in E. Zamir, D. Teichman (eds), *The Oxford handbook of behavioral economics and the law*, Oxford University Press, Oxford, 2014, pp 465–490.

R. Bartlett, J. Nyarko, V. Plaut, *Do you ever read the fine print? The potential and limitations of text analysis for consumer contracts*, 2019. Unpublished <https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CELS2019&paper_id=271>.

R. Benjamins, (ed), *Law and the semantic web: legal ontologies, methodologies, legal information retrieval, and applications*, in *Lecture notes in artificial intelligence*, 1st ed., Berlin, 2015.

O. Ben-Shahar, A. Chilton, *Simplification of privacy disclosures: an experimental test*, in *J. Legal Stud*, vol. 45(S2), 2016, pp. S41–S67.

O. Ben-Shahar, A. Porat, *Personalized law*. Oxford University Press, Oxford, 2021.

O. Ben-Shahar, C. E. Schneider, *More than you wanted to know: the failure of mandated disclosure*, Princeton University Press, Princeton, 2014.

G. Boella, L. Di Caro, V. Leone, *Semi-automatic knowledge population in a legal document management system*, in *Art. Intel. L*., vol. 27(2), 2019, p. 228.

G. Boella et al., *Semantic relation extraction from legislative text using generalized syntactic dependencies and support vector machines*, in L. Morgenstern et al (eds), *Theory, practice, and applications of rules on the web*, 2013, pp. 218–225.

G. Boella et al., *Linking legal open data: breaking the accessibility and language barrier in European legislation and case law*, in *Proceedings of the 15th international conference on artificial intelligence and law. Association for Computing Machinery*, 2015, pp. 171–175.
M. Botel, A. Granowsky, *A formula for measuring syntactic complexity: a directional effort*., in *Elementary Engl*., vol. 49(4), 1972, pp. 513–516.
V. C. Brannon, *Assessing commercial disclosure requirements under the first amendment. CRS Report No. R45700*. Congressional Research Service, Washington, D.C., 2019, <https://fas.org/sgp/crs/misc/R45700.pdf>.
H. Brignull, *Dark patterns: inside the interfaces designed to trick you*. The Verge, 2013.
C. Busch, *Implementing personalized law. Personalized disclosures in consumer law and data privacy law*, in *U. Chi. L. Rev*., vol. 86, 2019, pp. 309–331.
R. Calo, *Digital market manipulation*, in *Geo. Wash. L. Rev*., vol. 82(4), 2014, p. 995.
A. J. Casey, A. Niblett, *Framework for the new personalization of law*, in *U. Chicago L. Rev*, vol. 86(2), 2019, p. 359.
D. K. Citron, *Technological due process*, in *Washington U. L. Rev*, vol. 85(6), 2018, pp. 1249–1313.
J. C. Coffee, *Market failure and the economic case for a mandatory disclosure system*, in *Virginia L. Rev.*, vol. 70(4),1984, pp. 717–753.
C. Coglianese, D. Lehr, *Regulating by robot: administrative decision making in the machine-learning era*, in *Geo. L. J*., vol. 105(5), 2016, pp. 1147–1224.
G. Contissa et al., *CLAUDETTE meets GDPR. Automating the evaluation of privacy policies using artificial intelligence*, 2018a. <https://www.beuc.eu/publications/beuc-x-2018-066_claudette_ meets_gdpr_report.pdf>; <http://uterms.software/documentation/>.
G. Contissa, K. Docter, F. Lagioia, M. Lippi, H-W Micklitz, P. Palka, G. Sartor, P. Torroni, *Automated processing of privacy policies under the EU General Data Protection Regulation*, in M. Palmirani (ed) *Legal knowledge and information systems. JURIX 2018: the thirty-first annual conference*, 2018b, pp. 51–60.
E. Costante, Y. Sun, M. Petkovi´c, J. den Hartog, *A machine learning solution to assess privacy policy completeness*, in *ACM workshop on privacy in the electronic society*, 2012, pp 91–96.
K. Crawford, J. Schultz, *Big data and due process: toward a framework to redress predictive privacy harms*, in *Boston Coll. L. Rev*., vol. 55(1), 2014, pp. 93–128.

C. Devins, T. Felin, S. Kauffman, R. Koppl, *The law and big data*, in *Cornell J. L. Pub. Pol'y* , vol. 27, 2017, pp. 357–413.

F. Di Porto, *In praise of an empowerment disclosure regulatory approach to algorithms*, in *IIC Int. Rev. Intellect. Property Compet. L.*, vol. 49(5), 2018, pp. 507–511.

F. Di Porto, M. Maggiolino, *Algorithmic information disclosure by regulators and competition authorities*, in *Glob Jurist*. 2019, <https://doi.org/10.1515/gj-2018-0048.

F. Di Porto, M. Zuppetta, *Co-regulating algorithmic disclosure for digital platforms*, in *Pol Soc'y* 2020, <https://doi.org/10.1080/14494035.2020.1809052>.

B. Fabian, T. Ermakova, T. Lentz, *Large-scale readability analysis of privacy policies*, in *Proceedings of the international conference on web intelligence. Association for Computing Machinery*, Leipzig, Germany, 2017, p 21.

F. Fagan, *Big data legal scholarship: toward a research program and practitioner's guide*, in *Virginia J. L. Technol.*, vol. 20(1), 2016, pp. 1–81

A. Fung, M. Graham, D. Weil, *Full disclosure: the perils and promise of transparency*. Cambridge University Press, Cambridge, 2007.

J. Gluck, F. Schaub, A. Friedman, H. Habib, N. Sadeh, L. F. Cranor, Y. Agarwal, *How short is too short? Implications of length and framing on the effectiveness of privacy notices*. Paper presented at the twelfth Symposium on Usable Privacy and Security (SOUPS 2016), 2016.

G. Governatori, M. Hashmi, H-P Lam, S. Villata, M. Palmirani, *Semantic business process regulatory compliance checking using LegalRuleML*, in E. Blomqvist, P. Ciancarini, F. Poggi, F. Vitali (eds) *Knowledge engineering and knowledge management*. Springer, Berlin, 2016, p. 749.

S. J. Grossman, J. E. Stiglitz, *On the impossibility of informationally efficient markets*, in *Am. Econ. Rev.*, vol. 70(3), 1980, pp. 393–408.

H. Harkous, K. Fawaz, R. Lebret, F. Schaub, K.G. Shin, K. Aberer, *Polisis: Automated analysis and presentation of privacy policies using deep learning*, in *Proceedings of the 27th USENIX Security Symposium*, 15–17 August 2018, Baltimore.

A. Koene et al., *A governance framework for algorithmic accountability and transparency*, 2019, PE 624.262. European Parliamentary Research Service.

R. Lepina, G. Contissa, K. Drazewski, F. Lagioia, M. Lippi, H-W Micklitz, P. Pałka, G. Sartor, P. Torroni, *GDPR privacy policies in CLAUDETTE: challenges of omission, context and Multilingualism*, in *Proceedings of the third workshop on automated semantic analysis of information in legal text*, ASAIL, 2019, pp 1–7.

S. Levmore, *Probabilistic disclosures for corporate and other law*, in *Theor. Inquiries L.*, vol. 22(1), 2021, pp. 263–284.
S. Levmore, F. Fagan, Competing algorithms for law: sentencing, admissions, and employment. Univ Chicago Law Rev., vol. 2021, p. 367.
R. Liepina, G. Contissa, K. Drazewski, F. Lagioia, M. Lippi, H-W Micklitz, P. Pałka, G. Sartor, P. Torroni, *GDPR privacy policies in CLAUDETTE: challenges of omission, context and Multilingualism*, in *Proceedings of the third workshop on automated semantic analysis of information in legal text* (ASAIL 2019).
M. Lippi et al., *CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service*, 2018, arXiv preprint arXiv:1805.01217.
F. Liu et al., *Modeling language vagueness in privacy policies using deep neural networks*, in *Association for the advancement of artificial intelligence fall symposium series*, 2016.
M. A. Livermore, D. N. Rockmore (eds), *Law as data*, SFI, 2019.
J. Luguri, L. Strahilevitz, *Shining a light on dark patterns*, in *J. Legal Anal*, vol. 13, 67, 2021.
F. Marotta-Wurgler, *Even more than you wanted to know about the failures of disclosure*, in *Jerusalem Rev. Legal Stud.*, vol. 11(1), 2015, pp. 63–74.
W. Mattli (ed), *Global algorithmic capital markets: high-frequency trading, dark pools, and regulatory challenges*, Oxford University Press, Oxford, 2018.
M. Medvedeva, M. Vols, M. Wieling, *Using machine learning to predict decisions of the European Court of Human Rights*, in *Artif. Intell. L.*, vol. 28, 2019, pp. 237–266. <https://doi.org/10.1007/s10506- 019-09255-y>.
E. Montiel-Ponsoda, V. Rodríguez-Doncel, *Lynx: building the legal knowledge graph for smart compliance services in multilingual Europe*, in G. Rehm, V. Rodríguez-Doncel, J. Moreno-Schneider (eds) *Proceedings of the 1st workshop on LREC* (Language Resources and Technologies for the Legal Knowledge Graph) workshop, 12 May 2018, pp 19–22. <https://delicias.dia.fi.upm.es/members/vrodriguez/pdf/2018.legalkg.pdf>.
K. Mysore Sathyendra et al., *Identifying the provision of choices in privacy policy text*, in *Proceedings of the 2017 conference on empirical methods in natural language processing*. Association for Computational Linguistics Copenhagen, Denmark, 2017, pp. 2774–2779.
R. Nanda, G. Siragusa, L. Di Caro, G. Boella, L. Grossio, M. Gerbaudo, F. Costamagna, *Unsupervised and supervised text similarity systems for automated identification of national implementing measures of European directives*, in Artif. Intell. L., vol. 27, 2019, pp.199–225.

M. Palmirani, M. Martoni, A. Rossi, C. Bartolini, L. Robaldo, *PrOnto: privacy ontology for legal reasoning*, in *EGOVIS2018, 7th International Conference, EGOVIS 2018*, Regensburg, Germany, September 3–5, 2018, Proceedings. LNCS, vol 11032. Springer, pp. 139–152.

M. Palmirani, G. Governatori, *Modelling legal knowledge for GDPR compliance checking*, in M. Palmirani (ed), *Legal knowledge and information systems*, 2018, p. 101.

Y. Panagis, U. Sadl, F. Tarissan, *Giving every case its (legal) due. The contribution of citation networks and text similarity techniques to legal studies of European Union law*. Paper presented at the 30th international conference on legal knowledge and information systems, JURIX 2017, Luxembourg, December 2017.

P. G. Picht, G. T. Loderer, *Framing algorithms—competition law and (other) regulatory tools*. MPI Research Paper No. 18-24, 2018.

V. C. Plaut, R. P. Bartlett, *Blind consent? A social psychological investigation of non-readership of click-through agreements*, in *L. Human Behav.*, vol. 36(4), 2012, pp. 293–311.

N. Sannier, M. Adedjouma, M. Sabetzadeh, L. Briand, *An automated framework for detection and resolution of cross references in legal texts*, in Requir. Eng., vol. 22(2), 2017, pp. 215–237.

D. Sarne et al., *Unsupervised topic extraction from privacy policies*, in *Companion Proceedings of the 2019 World Wide Web Conference on—WWW '19*, vol 563. ACM Press, San Francisco, p. 564.

G. Sartor, P. Casanovas, M. Biasiotti, M. Fernández-Barrera (eds), *Approaches to legal ontologies. Theories, domains, methodologies*. Springer, Berlin, 2011.

R. Shedlosky-Shoemaker, A. C. Sturm, M. Saleem, K. M. Kelly, *Tools for assessing readability and quality of health-related web sites*, in *J. Genet Couns*, vol. 18(1), 2009, pp. 49–59.

B. Shrader, *What is the difference between an ontology and a knowledge graph?*, 2020, <https://enterpriseknowledge.com/whats-the-difference-between-an-ontology-and-a-knowledge-graph/>.

Y. Shvartzhneider, N. Apthorpe, N. Feamster, H. Nissenbaum, *Analyzing privacy policies using contextual integrity annotations*, 2018. arXiv preprint arXiv:1809.02236.

Y. Shvartzshnaider, Z. Pavlinovic, A. Balashankar, T. Wies, L. Subramanian, H. Nissenbaum, P. Mittal, *VACCINE: using contextual integrity for data leakage detection*, in *The World Wide Web Conference on—WWW '19*. ACM Press, San Francisco, 2019.

A-L Sibony, G. Helleringer, (2015*) EU consumer protection and behavioural sciences: revolution or reform?*, in A. Alemanno, A-L Sibony (eds) *Nudge and the law. A European perspective*. Hart Publ., 2015, pp. 209–233.

Stigler Center at Chicago Booth, *Report by the committee for the study of digital platforms—privacy and data protection subcommittee*, 2019.

B. Szmrecsanyi, *On operationalizing syntactic complexity*, in *JADT 2004: 7es Journées internationales d'Analyse statistique des Données Textuelles*, 2004, pp 1031–1038.

R. Thaler, *Nudge, not sludge*, in *Science*, vol. 361, 1, 2018.

C-Y Tsang, *From industry sandbox to supervisory control box: rethinking the role of regulators in the era of Fintech*, in *Proceedings of the Comparative Corporate Governance Conference*, Singapore, January 24, 2019, p 359.

M. Waddington, *Research note. Rules as code*, in *Law Context*, vol. 37(1), 2020, pp. 1–8. <https://doi.org/10.26826/law-in-context.v37i1.134>.

S. Wilson, F. Schaub, A.A. Dara, F. Liu, S. Cherivirala, P. Giovanni Leon, M. Schaarup Andersen, S. Zimmeck, K. M. Sathyendra, N. C. Russell, T. Norton, E. Hovy, J. Reidenberg, N. Sadeh, *The creation and analysis of a website privacy policy corpus*, in *Proceedings of the 54th annual meeting of the Association for Computational Linguistics* (vol 1: long papers). Association for Computational Linguistics, Berlin, Germany, 2016.

D. Yang, M. Li, *Evolutionary approaches and the construction of technology-driven regulations*, in *Emerg Markets Finance Trade*, vol. 54(14), 2018, pp. 3266.

E. Zamir, D. Teichman, *Behavioral law and economics*. Oxford University Press, Oxford, 2018.

S. Zuboff, *The age of surveillance capitalism: the fight for the future at the new frontier of power*. Public Affairs, 2019.

Annalisa Signorelli

La prevedibilità della e nella decisione giudiziaria

Abstract: The paper analyses the implications of the use of Artificial Intelligence technologies in the civil process, according to two different meaning of "predictability" referred to the judge's decision: predictability "of" rulings, which allows the identification - using data research analysis tools - of the past case-law having greater relevance, and predictability "in" rulings, which instead concerns the judge's decision making processes.
The paper explores the conditions and limits of predictive justice theory applied to the Italian civil process, with a general overview on European perspectives too.

1. *Introduzione*

Il termine "giustizia predittiva" viene oggi riferito a quell'insieme degli strumenti di supporto alla funzione legale e giurisdizionale capaci di analizzare in tempi brevi una grande quantità di informazioni con l'obiettivo di prevedere l'esito, o i possibili esiti, di un giudizio[1].

In generale, la possibilità di "prevedere" le soluzioni dei casi concreti sottoposti all'attenzione del giudice si disvela particolarmente utile, infatti, non solo ai fini del calcolo delle probabilità di successo di una causa, ma anche nella prospettiva più ampia della calcolabilità e della misurabilità del diritto[2].

Negli ultimi anni un grande passo in avanti è stato compiuto rispetto alle primigenie forme di utilizzo delle tecnologie informatiche: in una

[*] L'articolo è stato originariamente pubblicato in R. Giordano et al, in *Il Diritto nell'era digitale. Persona, Mercato, Amministrazione, Giustizia*, Giuffrè 2022, p. 997 ss.

[1] V. L. Viola, voce *Giustizia Predittiva*, in «*Enc. Giur. Treccani online*», 2018. V. in argomento anche, S. De La Oliva, *Giustizia predittiva, interpretazione matematica delle norme, sentenze robotiche e la vecchia storia del* «justizklavier», Rivista Trimestrale di Diritto e Procedura Civile, 2019, 838 ss.

[2] Secondo parte della dottrina (P. Rossi, *Razionalismo occidentale e calcolabilità giuridica*, in *Calcolabilità giuridica* (a cura di) A. Carleo, Bologna 2017, 32), presupposti della "calcolabilità" in senso giuridico sono: l'esistenza di un insieme di norme tra loro coordinate; la rappresentazione di queste norme da parte di soggetti economici impegnati nel mercato; l'esistenza di un apparato coercitivo, tribunali o altro, a cui rivolgersi in caso di necessità; la capacità di intervento efficace da parte di questo apparato al fine di garantire la conformità del comportamento alle norme statuite.

prima fase, il connubio tra la tecnologia e il processo civile operava in funzione di ausilio delle professioni legali ovvero della giurisdizione, al fine di razionalizzare il funzionamento del processo civile (si pensi al processo civile telematico), di predisporre database giurisprudenziali o di creare strumenti di documentazione e organizzazione dei servizi a supporto degli operatori della giustizia[3]. Si tratta di impieghi dei *tools* tecnologici che sono stati finora sperimentati e hanno disvelato i potenziali vantaggi in termini di efficientamento del sistema giustizia, di maggiore rapidità e di agevole praticità.

Dall'idea del digitale come strumento prodromico alla razionalizzazione e all'organizzazione della giustizia si è passati ad indagare le possibilità di utilizzo dell'informatica e dell'intelligenza artificiale nel settore della decisione giudiziaria. Lo sviluppo di nuove tecnologie, sempre più avanzate, ha condotto a sperimentare gli utilizzi degli algoritmi che impiegano contenuti delle decisioni sia con finalità analitica-induttiva, mediante l'impiego e l'elaborazione di dati riguardanti decisioni già assunte, sia con finalità prospettico-predittiva, riuscendo a sviluppare prognosi di decisione[4].

Si discute pertanto di "giustizia predittiva"[5] quale forma di esercizio della funzione giurisdizionale mediante l'impiego di strumenti algoritmici di elaborazione dei dati e delle informazioni rilevanti nel giudizio, che presenti i vantaggi di maggiore rapidità ed efficienza rispetto all'*agere* esclusivamente umano[6].

Il fenomeno di cui stiamo trattando concerne dunque la realizzazione di sistemi automatizzati di decisione giudiziale con applicazione di metodi che vedono interessata l'azione di algoritmi con grado di autonomia dell'intelligenza artificiale molto elevato e con emulazione dell'intelligenza

[3] G. Mammone, *Considerazioni introduttive sulla decisione robotica*, in *Decisione robotica* (a cura di) A. Carleo, Bologna 2019, 30.

[4] L. De Renzis, *Primi passi nel mondo della giustizia «high-tech»: la decisione in un corpo a corpo virtuale fra tecnologia e umanità*, in *Decisione robotica* (a cura di) A. Carleo, Bologna 2019, 148.

[5] In Italia lo sviluppo di progetti sulla giustizia predittiva è ancora allo stato primordiale, ma la presa d'atto della rilevanza del tema è già avvenuta: v. A. Bonafine – A. Panzarola, *Brevi considerazioni sul D.D.L. per la riforma del processo civile approvato dal Consiglio dei Ministri il 5 dicembre 2019*, *Giustiziacivile.com*, 2019, 7.

[6] C. Castelli - D. Piana, *Giustizia predittiva. La qualità della giustizia in due tempi*, in *Questione Giustizia*, maggio 2018.

umana[7], che siano in grado di assicurare un diritto calcolabile[8], ossia, per citare Natalino Irti, «*fondato su fattispecie normative, giudizio di sussunzione e metodologia ermeneutica*»[9].

Se la calcolabilità del diritto costituisce oggi il punto di partenza, nonché la giustificazione sul piano assiologico dell'auspicato utilizzo dell'AI in relazione alla decisione giudiziaria, occorre però fare una considerazione preliminare sulla possibilità stessa di calcolo e sulla funzione che – nell'ambito considerato – riveste la teoria della probabilità.

Tra le diverse definizioni[10], il concetto di probabilità applicata al diritto è stato sempre più declinato in senso soggettivo, *i.e.* la probabilità non esiste come cosa in sé (oggettiva) perché è l'opinione di un dato soggetto, in un dato istante e con un dato insieme di informazioni riguardo al verificarsi di un determinato evento (di cui il soggetto non conosce l'esito)[11]. L'assunto di fondo di tale concezione è, dunque, quello di tenere distinto il carattere soggettivo della probabilità e il carattere oggettivo degli elementi cui essa si riferisce; e, secondo taluni, tale assunto rimane inalterato anche laddove il processo decisionale sia operato non già (solo) da un agente umano, ma anche da un agente intelligente.

Tuttavia, sebbene gli *input* (i.e. le informazioni fornite all'algoritmo per lavorare) siano elaborate da "filtri" soggettivi, la macchina è un filtro oggettivo e produce delle decisioni che, quanto all'aspetto procedurale, sono oggettive: ciò in quanto essa non lascia indeterminatezza nel processo decisionale, opera con un algoritmo su una linea logica «formalmente normativa».

---

[7] M.R. Covelli, *Dall'informatizzazione della giustizia alla «decisione robotica»? Il giudice del merito*, in *Decisione robotica* (a cura di) A. Carleo, Bologna 2019, 131, a proposito dello sviluppo robotico nell'attività giudiziaria. Si discute dunque della possibilità di elaborazione automatizzata dei dati in funzione predittiva in ambito giurisdizionale, ossia all'uso di strumenti basati su dati preesistenti e ricavati da banche dati ai fini della decisione del caso concreto in base a criteri statistico-matematici.

[8] N. Irti, «Calcolabilità» e crisi della fattispecie, in *Rivista di diritto civile*, 2014; Id, *La crisi della fattispecie*, in *Rivista trimestrale di diritto e procedura civile*, 2015.

[9] N. Irti, *Un diritto incalcolabile*, in *Rivista di diritto civile*, 2014; Id, *La crisi della fattispecie*, in *Rivista trimestrale di diritto e procedura civile*, 2015.

[10] Dalla definizione classica della probabilità di un evento come il rapporto fra il numero degli esiti favorevoli, che fanno sì che l'evento si verifichi, e il numero degli esiti possibili; alla concezione frequentista secondo la quale la probabilità è la frequenza relativa (quando il numero delle prove è sufficientemente alto).

[11] V. in tal senso De Felice M., *Calcolabilità e probabilità. Per discutere di «incontrollabile soggettivismo della decisione»*, in Calcolabilità giuridica (a cura di) A. Carleo, Bologna, 2017, 37, che richiama in incipit Bruno de Finetti in "Teoria della probabilità" il quale afferma icasticamente che «*La probabilità non esiste*».

Nella consapevolezza che – da un punto di vista strettamente matematico – «*la previsione non è predizione*»[12], nel senso che il soggetto deve tenere conto di tutte le circostanze oggettive note, di eventuali "simmetrie" (come nella definizione classica di probabilità) purché siano frequenze ben definite, ben interpretate e adeguate allo scopo, si vorranno analizzare le possibili implicazioni degli strumenti di IT ai fini della "calcolabilità oggettiva" delle decisioni del giudice.

Ma è davvero possibile costruire un modello di calcolabilità oggettiva, che sia in grado di «*evitare la deriva della giurisdizione verso l'instabilità del diritto "liquido", in un giusto equilibrio tra dimensione creativa e plurale del diritto giurisprudenziale e i principi di uniformità e prevedibilità della decisione*»[13]? E tale costruzione corrisponde alle premesse assiologiche di partenza e risponde all'attuale assetto istituzionale della giurisdizione civile?

Come è stato acutamente osservato, l'idea del giudice-robot è nata e si è sviluppata «*come la rappresentazione icastica di una soluzione che si vorrebbe davvero capace di realizzare, addirittura in via ottimale, l'obiettivo di coonestare, finalmente in modo penetrante, il concetto di efficienza nelle sue plurime declinazioni, nello svolgimento dell'attività giudiziaria*»[14].

Tuttavia, sebbene *in votis* destinata ad un efficientamento del sistema giudiziario e ad una generale garanzia di certezza del diritto, occorre considerare che nelle sue applicazioni pratiche (se non addirittura già nelle concettualizzazioni teoriche) la giustizia predittiva può recare seco anche possibili conseguenze svantaggiose o attrarre esternalità negative, *prima facie* non evidenti o obnubilate dalle paventate opportunità e dalla cultura del determinismo tecnologico[15].

Un primo ordine di criticità, di ordine generale, investe il rispetto dei principi costituzionali del giusto processo: l'utilizzo degli algoritmi nelle

[12] De Felice M., *Calcolabilità e probabilità. Per discutere di «incontrollabile soggettivismo della decisione»,* cit..

[13] Così Canzio G., *Relazione sull'amministrazione della giustizia nell'anno 2015*.

[14] E. Vincenti, *Il «problema» del giudice robot*, in *Decisione robotica* (a cura di) A. Carleo, Bologna 2019, 111.

[15] L'espressione è ascrivibile al sociologo ed economista statunitense Thorstein Veblen, e descrive la concezione secondo la quale la tecnologia (l'a. richiama, a titolo esemplificativo, i particolari sviluppi tecnici, le tecnologie della comunicazione o i media) è l'unica causa (intesa come antecedente logico-causale) dei cambiamenti nella società, e pertanto è considerata come la condizione fondamentale che soggiace ai modelli di organizzazione sociale. La tecnologia rappresenta la variabile indipendente, cioè un fattore dotato di una forza che, in maniera autonoma, è in grado di guidare l'azione umana e di mutare la società. Cfr. T. Veblen, *Theory of the Leisure Class: An Economic Study of Institutions*, 1899.

decisioni giudiziarie pone anzitutto i problemi di assicurare l'imparzialità e la terzietà del giudice, oltre la trasparenza e la pubblicità della funzione giurisdizionale, quali valori fondamentali riconosciuti dalla nostra Carta costituzionale.

Sul punto, le recenti conclusioni del 21 ottobre 2020 della Presidenza del Consiglio dell'Unione Europea hanno chiesto di affrontare l'opacità, la complessità, la parzialità, un certo grado di imprevedibilità e il comportamento parzialmente autonomo di alcuni sistemi di IA, per garantire la loro compatibilità con i diritti fondamentali e per facilitare l'applicazione delle norme giuridiche[16].

Nella Proposta di Regolamento del Parlamento Europeo e del Consiglio del 21 aprile 2021, che stabilisce norme armonizzate in materia di intelligenza artificiale, si riconosce espressamente[17] che tra gli obiettivi del nuovo quadro regolatorio sull'intelligenza artificiale dovrebbero esserci sia il rispetto e la più efficace applicazione del diritto vigente in materia di diritti fondamentali e i valori dell'Unione da parte dei sistemi di IA, sia la garanzia della certezza del diritto per facilitare gli investimenti e l'innovazione nell'IA.

Un secondo ordine di criticità riguarda invece l'impatto dell'utilizzo degli algoritmi di intelligenza artificiale sulle decisioni giudiziarie, alla luce delle peculiarità del processo civile come disciplinato nel nostro ordinamento.

Il contributo vuole analizzare le possibili implicazioni dell'impiego delle tecnologie algoritmiche nel processo civile, alla luce degli studi della giurimetria[18], secondo due diverse declinazioni del concetto di prevedibilità.

[16] Council of the European Union, Presidency conclusions - The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change, 11481/20, 2020, disponibile al seguente link: https://www.consilium.europa.eu/media/46496/st11481-en20.pdf.

[17] *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts*, pag. 3, disponibile al seguente link: https://ec.europa.eu/transparency/regdoc/rep/1/2021/EN/COM-2021-206-F1-EN-MAIN-PART-1.PDF . Come emerge dalla Comunicazione della Commissione Europea sulla promozione di un approccio europeo all'intelligenza artificiale del 21.04.2021, pag. 4, la richiamata proposta della Commissione per un quadro normativo sull'IA rappresenta uno snodo fondamentale nel percorso verso la protezione della sicurezza e dei diritti fondamentali e quindi la garanzia di fiducia nello sviluppo e nell'adozione dell'IA. L'atto della Commissione è stato preceduto dalle *Guidelines for Trustworthy AI* (2019) e dal *White Paper on AI* (2020) al quale è seguita una consultazione pubblica internazionale.

[18] La giurimetria può essere definita come la scienza che studia l'applicazione di metodi matematici per la risoluzione di problemi giuridici; in particolare, si è occupata della

La prevedibilità può infatti essere intesa in due distinte accezioni: prevedibilità *delle* decisioni giudiziarie, che opera *ab externo* e che consente di individuare – mediante l'utilizzo di strumenti di *data research analysis* – i precedenti giurisprudenziali aventi maggior rilievo, e prevedibilità *nelle* decisioni giudiziarie, che invece opera internamente ed inerisce ai processi di *decision making* del giudice.

L'interrogativo di fondo che ci accompagnerà nella presente trattazione investe dunque il profilo della permeabilità dell'attuale sistema processual-civilistico alle istanze innovatrici dell'Intelligenza Artificiale.

2. *La prevedibilità della decisione giudiziaria*

Nella prima accezione delineata, la giustizia predittiva è funzionale a ricavare la probabilità dell'esito di una causa in un certo tribunale o corte, alla luce dei precedenti e degli orientamenti giurisprudenziali invalsi. Come anticipato, la prevedibilità della decisione giudiziaria opera, dunque *ab externo*: ci si riferisce a quei fenomeni di calcolabilità delle pronunce del giudice.

Una premessa di ordine generale impone di considerare le diversità tra gli ordinamenti di *common law*, caratterizzati dalla vincolatività del precedente giurisprudenziale (stare decisis) e gli ordinamenti di *civil law*, in cui il precedente ha un valore meramente persuasivo[19].

Tale distinzione implica che il discorso sulla calcolabilità della decisione giudiziaria mediante l'impiego di strumenti algoritmici avrà un peso specifico diverso nell'uno e nell'altro sistema ordinamentale, comportando conseguenze diverse a seconda del grado di intensità del precedente stesso.

Tuttavia, il discorso che si intende condurre è analogo nei presupposti di partenza: per poter parlare di prevedibilità delle sentenze è infatti necessario che[20]: a) il giudice sia obbligato (in via più o meno intensa) a conformarsi

"misurazione" delle decisioni giudiziarie e della costruzione di modelli per la sua prevedibilità: v. V. L. Viola, voce *Giustizia Predittiva*, in *Enc. Giur. Treccani online*, 2018.

[19] Occorre dunque tenere presente la diversità di approcci filosofici e culturali alle decisioni giudiziarie, per cui, in alcuni paesi europei, compresa la Francia, vi è una cultura di precedenti e una conoscenza dettagliata da parte dei giudici delle banche dati di fatto di tutte le decisioni di primo e secondo istanza (banca dati Ariane) nel campo della giustizia amministrativa, mentre altri paesi o sistemi favoriscono l'indipendenza intellettuale di ogni tribunale, insieme al desiderio di affrontare ogni situazione caso per caso.

[20] M. Nuzzo, *Il problema della prevedibilità delle decisioni: calcolo giuridico secondo i*

ad una decisione adottata in una precedente sentenza pronunciata da un giudice dello stesso grado (vincolatività orizzontale) ovvero da un giudice di grado superiore (vincolatività verticale); b) l'efficacia vincolante della sentenza sia rapportata alla sola *ratio decidendi*, i.e. le argomentazioni giuridiche poste a fondamento e a giustificazione della decisione del giudice, mentre gli *obiter dicta* avranno mero valore persuasivo[21]; c) il vincolo operi nei limiti in cui vi sia coincidenza tra gli elementi del caso concreto portato all'attenzione del giudice chiamato a decidere e gli elementi dei casi decisi in precedenza.

Sebbene il nostro ordinamento giuridico sia di *civil law*, è noto che le riforme del legislatore processuale degli ultimi anni si siano indirizzate verso una riscoperta del valore del precedente, soprattutto di legittimità, vuoi con finalità deflattive (si veda, ad esempio, l'art. 360-*bis* c.p.c.) vuoi con finalità di assicurare l'uniforme interpretazione e applicazione della legge (si vedano l'art. 374, c. 3, c.p.c. e l'art. 118 disp. att. c.p.c.)[22].

Le recenti riforme del giudizio di cassazione (D.lgs. n. 40/2006, L. n. 69/2009, e D.L. n. 83/2012, convertito in L. n. 134/2012), hanno finito col delineare un nuovo ruolo nomofilattico della Corte di Cassazione[23],

precedenti, in Calcolabilità giuridica (a cura di) A. Carleo, Bologna 2017, 145-146.

[21] Tuttavia, vi è chi (Viola, voce *Giustizia Predittiva* cit.) ritiene che si possa far riferimento anche ad un *obiter dictum*, sull'assunto che quando si cerca il precedente giurisprudenziale sia più corretto riferirsi alla c.d. dottrina della giurisprudenza (G. De Nova, *Sull'interpretazione del precedente giudiziario*, in *Contratto e Impresa*, 1986, 782).

[22] In particolare, le tre disposizioni processuali (art. 360-*bis* e 374 c.p.c., art. 118 disp. att. c.p.c.) oggetto degli interventi legislativi rappresentano gli addentellati positivi del rinnovato valore del precedente giudiziario nel nostro ordinamento, poiché operano direttamente a tre diversi livelli soggettivi di interazione. Nei rapporti interni all'organo giudiziario di legittimità, viene in rilievo il novellato art. 374 c.p.c., che prevede l'obbligo delle sezioni semplici della Corte di Cassazione di rimettere la questione alle Sezioni Unite civili se intendono discostarsi dai precedenti orientamenti di queste ultime. Nei rapporti tra giudice di merito e giudice di legittimità, l'art. 118 disp. att. c.p.c., nel prevedere la possibilità per i giudici di merito di inserire nella motivazione della sentenza il richiamo alla precedente giurisprudenza conforme, finisce per consentire al giudice di merito di discostarsi dall'orientamento giurisprudenziale di legittimità solo in presenza di "forti e apprezzabili ragioni giustificative" (Cass. civ., 9 gennaio 2015, n. 174). Infine, nei rapporti tra giudice e cittadino (*rectius*: difensore legale del cittadino) l'art. 360-*bis* c.p.c. impone una formulazione dei motivi di ricorso per cassazione che scongiuri il rischio della censura di inammissibilità, prevedendo che il ricorso sia inammissibile se il provvedimento impugnato sia conforme alla giurisprudenza della Corte e l'esame dei motivi non offra elementi per un mutamento di indirizzo. La disposizione in esame finisce per trasformare la prassi giurisprudenziale della Suprema Corte da mero elemento persuasivo a vero e proprio requisito di accesso alla giustizia di legittimità.

[23] Anzitutto, si è dato atto del definitivo superamento *ex iure positivo* della prima stagione

introducendo norme specificamente dirette all'obiettivo della tendenziale uniformità della giurisprudenza (in ossequio all'art. 65 dell'Ordinamento giudiziario)[24] e lasciando spazio alle teorizzazioni sulla contrapposizione tra *ius costitutionis* e *ius litigatoris*[25].

A tali modifiche si aggiunge anche il nuovo art. 348-*bis* c.p.c. che prevede l'inammissibilità dell'appello quando non ha una ragionevole probabilità di essere accolto, il che si verifica di solito quando la sentenza impugnata richiama principi consolidati della giurisprudenza della Corte di Cassazione[26].

Il precipitato logico della nuova disciplina è il rafforzamento – per via legislativa – del valore del precedente giurisprudenziale, con conseguenze rilevanti a livello istituzionale, sistematico e ordinamentale[27], nella prospet-

della Corte di Cassazione (v. sul punto C. Punzi, *La Cassazione da custode dei custodi a novella fonte di diritto?*, in *Historia et ius – Rivista di storia giuridica dell'età medievale e moderna*, 1/2012, 1), quale organo deputato a custodire e a far osservare rigorosamente la legge.Si rammenti che da tale concezione era emersa la nota ricostruzione della Cassazione civile operata da Piero Calamandrei - in chiave tutta pubblicistica - con il riconoscimento in capo al giudice di legittimità del ruolo di vero e proprio organo costituzionale, così accentuando il rischio della perdita definitiva della distinzione fondamentale tra le fonti di produzione della legge e le fonti di attuazione della stessa.

[24] Cfr. L. Salvaneschi, *L'iniziativa nomofilattica del Procuratore generale presso la Corte di Cassazione nell'interesse della legge*, «Rivista di Diritto Processuale», 1/2019, 65. Sul ruolo della nomofilachia a seguito delle più recenti riforme processuali, v. L. Passanante, *Il precedente impossibile. Contributo allo studio del diritto giurisprudenziale nel processo civile*, Torino 2018. Sul rapporto tra precedenti e nomofilachia, v. AA.VV., *Il vincolo giudiziale del passato. I precedenti* (a cura di) A. Carleo, Bologna 2018 (ed *ivi*, in particolare, i saggi di: N. Irti, *Sulla relazione logica di conformità (precedente e susseguente*), 17 ss.; G. Canzio, *Nomofilachia e diritto giurisprudenziale*, 27 ss.; R. Rordorf, *Il precedente nella giurisprudenza*, 89 ss.; P. Curzio, *Il giudice e il precedente*, 239 ss.; F. Patroni Griffi, *«Consuetudini e usi giudiziari» e diritto giurisprudenziale*, 255 ss.).

[25] V. in argomento B. Sassani, *La deriva della cassazione e il silenzio dei chierici*, «Rivista di Diritto Processuale», 1/2019, 45. Tra l'altro, non v'è chi non veda in tale contrapposizione l'effetto paradossale di ipergarantire i diritti delle parti: per tutti, B. Sassani, *Giudizio sommario di cassazione e illusione nomofilattica*, «Rivista di Diritto Processuale», 1/2017, 37.

[26] Secondo taluna giurisprudenza (Trib. Milano, 16/09/2016; App. Roma 23/01/2013), il giudizio di ragionevole probabilità di accoglimento dell'appello a norma dell'art. 348 *bis* c.p.c. non si risolve né in una valutazione sommaria parificabile a quella identificata con il *fumus boni iuris*, né in una valutazione a cognizione parziale come quelle relativa ai procedimenti a contraddittorio eventuale. Deve infatti ritenersi che l'appello non ha ragionevoli probabilità di accoglimento quando è *prima facie* infondato.

[27] Sia consentito il rinvio a A. Signorelli, *L'attivismo giudiziario tra diritti fondamentali e sicurezza giuridica: dal giudice* bouche de la loi *al giudice* law-maker, in *Cahiers Jean Monnet*, 2020, 475.

tiva di assicurare la certezza del diritto e l'uniformità delle decisioni.

In quest'ottica, l'applicazione dei sistemi di Intelligenza Artificiale può avvenire in due diverse direzioni. Da un lato, ai fini dell'analisi (ed eventualmente, del contenimento) dei fenomeni di *overruling* sostanziale e processuale e, dunque, per scopi *lato sensu* limitativi del cd. attivismo giudiziario. Tale funzione risponderebbe ad esigenze di giustizia "ragionevolmente eguale", secondo il disposto dell'art. 3 della Costituzione[28].

Dall'altro lato, le nuove forme di tecnologia potrebbero essere utilizzate nel contesto degli artt. 348-*bis* e 360-*bis* c.p.c., *i.e.* allo scopo di assicurare l'operatività dei filtri al giudizio delle corti. Sebbene tale funzione sia attualmente positivizzata con riferimento all'*agere* umano, e dunque la compatibilità con l'art. 24 Cost. sia già stata vagliata positivamente dal legislatore ordinario, la trasposizione in chiave tecnologica potrebbe porre significativi dubbi in punto di accesso alla giustizia, perché «dall'esito del filtro dipenderà la possibilità di ottenere o meno giustizia»[29].

Le nuove frontiere della prevedibilità sono rappresentate oggi dall'utilizzo del NLP (Natural Language Processing) e dei metodi computazionali[30], al fine di studiare sia il ruolo del precedente di una corte sia di studiare il ragionamento giuridico e la motivazione delle decisioni giudiziarie.

La maggior parte degli studi giuridici computazionali fino ad oggi è stata dedicata allo studio delle decisioni giudiziarie, spesso attraverso la *network analysis* con o senza l'aggiunta del NLP[31].

I metodi computazionali possono rivelare se le decisioni giudiziarie agiscono come fonti del diritto (nei paesi di common law) e se servono come precedenti di fatto (nei paesi di *civil law*), in quanto è stato dimostrato che i tribunali si basano su casi precedenti nella loro giurisprudenza con

[28] Come sottolinea L. Viola, voce *Giustizia Predittiva* cit., «*l'art. 3 Cost. evidenzia, più di altri articoli, la visione di un diritto oggettivo e certo che deve permeare l'intero ordinamento: è imposto di trattare in modo uguale situazioni giuridiche uguali, che vuol dire assicurare il medesimo risultato (stesso trattamento) a parità di variabili (medesima situazione)*».

[29] L. De Renzis, *Primi passi nel mondo della giustizia «high-tech»* cit., 150.

[30] Sull'assunto che se è vero che la legge ben può essere informatizzata e costruita tramite algoritmi (R. Borruso, *L'informatica del diritto,* Milano 2004, 316), è anche vero che gli algoritmi possono essere utilizzati anche per analizzare il *corpus* testuale delle disposizioni normative e delle sentenze che ne fanno applicazione.

[31] V. W. Alschner, *The Computational Analysis of International Law*, in *Research Methods in International Law: A Handbook* (a cura di) R. Deplano – N. Tsagourias, Ottawa Faculty of Law Working Paper No. 2019-33, 29 luglio 2019 https://ssrn.com/abstract=3428762. L'a. descrive l'utilizzo dei metodo computazioni ai fini dello studio delle fonti del diritto internazionale e delle corti e dei tribunali internazionali.

una frequenza crescente[32].

Gli studiosi usano anche misure di rete (*network measures*) per quantificare l'importanza di un caso in una rete di citazioni[33], per tracciarne l'uso crescente o decrescente per valutare se una determinata decisione offre ancora una valida interpretazione, per individuare i fattori che trasformano un caso in un precedente per una decisione futura, per rivelare collegamenti finora poco noti tra aree della giurisprudenza collegate attraverso le citazioni e per fornire nuovi metodi per tracciare l'evoluzione della giurisprudenza man mano che le decisioni precedenti vengono citate in nuovi contesti.

I metodi computazionali possono inoltre testare e raffinare le teorie esistenti sulla natura del precedente. Per esempio, cosa trasforma una decisione giudiziaria in un precedente autorevole?

Nel 1956 John Henry Merryman ha teorizzato che è in parte la citazione stessa a trasformare quella decisione in una fonte di autorità[34], e ciò ingenera un processo di autorevolezza "a cascata": i lettori di una decisione giudiziaria presumeranno che una fonte citata sia autorevole e saranno più propensi a citarla a loro volta[35].

Per cui, non è sufficiente inserire nel *dataset* della macchina una pluralità di precedenti giurisprudenziali[36], ma occorrerebbe contestualizzare il "peso" di quel precedente alla luce di una pluralità di fattori: anzitutto di carattere storico e culturale, ma anche in rapporto ad altri precedenti contrastanti nei casi in cui su una determinata questione non vi sia stata una pronuncia nomofilattica a Sezioni Unite. Il correttivo rispetto a questo effetto di

[32] N. Ridi, *"Mirages of an Intellectual Dreamland"? Ratio, Obiter and the Textualization of International Precedent*, in *Journal of International Dispute Settlement*, 2019, https://academic.oup.com/jids/advance-article/doi/10.1093/jnlids/idz005/5418549.

[33] J. H. Fowler et alt., *Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court*, 15 Political Analysis 324, 2007; J. Pauwelyn, *Minority Rules: Precedent and Participation Before the WTO Appellate Body*, in *Establishing Judicial Authority in International Economic Law* (a cura di) J. Jemoelniak – L. Nielsen – H. Palmer Olsen, Cambridge University Press, 2016.

[34] J. H. Merryman, *The Authority of Authority: What the California Supreme Court Cited in 1950*, in *Stanford Law Review*, vol. 6, 1954, 613.

[35] Questo processo, conosciuto nella *network analysis* come attaccamento preferenziale o come "il fenomeno del più ricco diventa più ricco" diventa cumulativo: più frequentemente un precedente giudiziario è citato nelle decisioni giudiziarie, più frequentemente sarà citato in quelle successive. V. W. Alschner, *The Computational Analysis of International Law* cit..

[36] M. Luciani, *La decisione giudiziaria robotica*, in *Decisione robotica* (a cura di) A. Carleo, Bologna 2019, 85-86. L'a. stigmatizza infatti il falso mito che l'inserimento nel repertorio del robot di tutti i precedenti giurisprudenziali potrebbe condurre ad un diritto oggettivo ed incontrovertibile.

"autorevolezza automatica" può essere rappresentato dalla combinazione tra l'analisi delle citazioni basata sui paragrafi con il NPL al fine di mostrare perché e come un caso viene citato.

Infine, i metodi di ricerca computazionale possono anche essere sfruttati per migliorare la comprensione dei ragionamenti logico-giuridici seguiti dalle corti nella decisione del caso concreto. In particolare, tali strumenti computazionali sono stati impiegati sia per indagare le motivazioni istituzionali dei singoli tribunali sia, in ottica comparativa, per verificare come tribunali elaborano le loro argomentazioni e citano i precedenti.

Significativi risultati sono stati raggiunti dagli studi relativi all'impiego del Natural Language Processing e del Machine Learning per costruire modelli predittivi che possono essere utilizzati per svelare i *pattern* che guidano le decisioni giudiziarie[37]. In particolare, l'analisi empirica condotta dagli studiosi dell'UCL ha consentito di elaborare modelli predittivi utili, sia per gli avvocati che per i giudici, come strumento di assistenza per identificare rapidamente i casi ed estrarre i modelli che portano a determinate decisioni[38].

3. *La prevedibilità nella decisione giudiziaria*

Nella seconda accezione esaminata, la prevedibilità nella decisione giudiziaria riguarda la possibilità di calcolare i meccanismi di *decision making* del giudice.

Se, come affermava il Maestro Calamandrei, il diritto processuale civile non è che una tecnica o «metodo di ragionamento» per ottenere una «sentenza giusta» e la scienza processuale una «metodologia», ovverosia una

[37] N. Aletras, D. Tsarapatsanis, D. Preoţiuc-Pietro, V. Lampos, *Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective*, in *PeerJ Computer Science* 2:e93, 2016, disposnibile al seguente indirizzo: https://doi.org/10.7717/peerj-cs.93.

[38] Si tratta del primo studio sistematico sulla previsione dell'esito dei casi giudicati dalla Corte europea dei diritti dell'uomo basato esclusivamente sul contenuto testuale. Su 584 cause della Corte europea dei diritti dell'uomo il programma è giunto alle stesse conclusioni dei giudici umani nel 79% dei casi. Il meccanismo di funzionamento si fonda su una classificazione binaria in cui l'*input* è il contenuto testuale estratto da un caso giudiziario e l'*output* di destinazione è la sentenza effettiva che stabilisce se c'è stata una violazione di un articolo della convenzione dei diritti umani. Le informazioni testuali sono rappresentate usando sequenze di parole contigue, cioè N-grammi, e argomenti; lo studio ha rivelato che i fatti formali di un caso sono il fattore predittivo più importante

«tecnica del ben ragionare in giudizio»[39], i sistemi di AI dovrebbero poter essere in grado di intercettare i "meccanismi del ragionamento giuridico" e riprodurli.

Ma, per la verità, il processo decisorio del giudice si compone di una moltitudine di passaggi logici e di valutazioni interinali variabili ed eterogenee, per cui non si può aprioristicamente dare una risposta affermativa o negativa al quesito circa la riproducibilità del ragionamento decisorio.

Si procederà dunque ad ipotizzare la prevedibilità dell'esito della singola controversia partendo dalle questioni di rito e dalle questioni di merito.

Le prime appaiono *prima facie* più facilmente adattabili a processi decisionali automatizzati, posto che il soggetto giudicante è chiamato ad effettuare una valutazione circa la sussistenza o meno dei presupposti processuali da verificarsi sì nel caso concreto, ma sulla base di parametri oggettivabili e, dunque, algoritmizzabili. I presupposti processuali sono infatti un *numerus clausus* e non sono disponibili dalle parti.

Ma anche qui occorre distinguere, poiché – come noto – la risoluzione di alcune delle questioni processuali implica la necessità che il giudice "si affacci" al merito per comprendere, seppur preliminarmente, le condizioni per l'azione in giudizio.

Infatti, secondo i noti insegnamenti[40], i presupposti processuali hanno una loro fattispecie astratta, descritta dalla norma, entro cui sussumere la fattispecie concreta; quest'ultima può essere realizzata o da un fatto interno ovvero da un fatto esterno al processo. Nel caso dei presupposti processuali integrati da fatti endo-processuali, laddove cioè la fattispecie del presupposto è integrata non tanto dall'effettiva esistenza del fatto storico bensì dalla sua affermazione in giudizio, il data-set dell'algoritmo dovrebbe già contenere (perché il "dato" è già stato immesso) l'informazione fattuale sufficiente a ritenere sussistente o meno quel presupposto.

Diversamente, nel caso dei presupposti processuali integrati da fatti extra-processuali, occorre procedere ad una verifica circa l'effettiva esistenza storica del fatto stesso.

[39] Come richiamato da A. Panzarola, *Una lezione attuale di garantismo processuale: le conferenze messicane di Piero Calamandrei*, in *Rivista di diritto processuale*, Vol. 74, Nº 1, 2019, 165. L'a. aggiunge che se si ha a che fare con una "tecnica del ben ragionare in giudizio", «ne deriva che il processo stesso, innervato dal nutrimento del contraddittorio, non è una mera sequenza di atti o semplice successione di forme ed è invece uno strumento gnoseologico volto a ricomporre un fatto del passato e a propiziare l'applicazione del diritto.

[40] F.P. Luiso, *Diritto processuale civile*, Vol I Principi generali, Milano, 2011, 118.

In questi casi, l'accertamento del fatto extraprocessuale nella sua materiale esistenza richiederebbe uno "sforzo conoscitivo" del mondo esterno che difficilmente potrebbe essere compiuto autonomamente dall'algoritmo, dovendosi invece procedere ad una verifica – mediante, ad esempio, l'assunzione di sommarie informazioni, secondo quanto previsto dall'art. 38 c.p.c. per la competenza – per mano umana.

Parimenti, nel caso delle pronunce relative alla litispendenza, alla continenza e alla connessione, i singoli *dataset* degli strumenti algoritmici potrebbero non essere sufficienti, se non opportunamente interoperabili simultaneamente tra loro al fine di consentire la verifica dei presupposti di operatività degli istituti in questione (identità o continenza, o tipologie di connessione fra cause).

Le questioni di merito, invece, sono per definizione "mobili", ovverosia eccessivamente mutevoli e non facilmente predeterminabili se non sulla base di una casistica ampia e variegata, e sono quelle che potrebbero maggiormente scontare il verificarsi di *bias* cognitivi e valutativi dell'organo giudicante.

Con riferimento alla *quaestio facti*, può anzitutto rilevarsi che esattamente come i fatti (rilevanti) devono essere portati a conoscenza del giudice (divieto di scienza privata), e quindi entrano a far parte della cognizione del giudice umano in quanto allo stesso rappresentati, parimenti nel caso di decisione algoritmizzata i fatti rilevanti per la decisione sono "immessi" nella macchina ed elaborati, con tutti i limiti cognitivi che ne possono ugualmente derivare[41].

Con riferimento alla *quaestio iuris*, invece, la piena operatività di una decisione algoritmica dipende dalla tipologia di disposizione normativa, i.e. dal grado di determinatezza della fattispecie astratta contemplata dalla legga, posto che «*affinché la legge possa essere applicata da un algoritmo, occorre che sia formulata in modo chiaro ed inequivoco*»[42].

Sul punto può osservarsi, in linea generale, che il *decision making* del giudice civile è stato sempre basato sull'adozione di un metodo sillogistico-deduttivo[43], i.e. di sussunzione del caso concreto nelle fattispecie astratte

[41] M. Luciani, *La decisione giudiziaria robotica* cit., 81, per il quale dunque il "calcolo" esatto della macchina dipende pur sempre da un fallibile processo cognitivo umano. Il rilievo è condiviso anche da E. Battelli, *Giustizia predittiva, decisione robotica e ruolo del giudice*, in *Giustizia Civile*, fasc.2, 1 febbraio 2020, 280 ss. Sull'esigenza di completezza dei fatti allegati e/o dedotti dalle parti v. Vincenti, *Il «problema» del giudice robot* cit., 119.

[42] A. Di Porto, *Calcolo giuridico secondo la legge nell'età della giurisdizione. Il ritorno del testo normativo*, in *Calcolabilità giuridica* (a cura di) A. Carleo, Bologna 2017, 130.

[43] F. Patroni Griffi, *Tecniche di decisione* cit., 177 ss.

e conseguente applicazione della *regula iuris* così individuata, il che presuppone che il processo decisionale faccia riferimento a modelli normativi astratti compiutamente definiti.

Se così è, la prevedibilità del diritto giurisprudenziale mediante l'utilizzo di algoritmi sembrerebbe *prima facie* di facile realizzabilità, poiché è la struttura stessa dell'algoritmo ad avere una base sillogistica: date alcune premesse in fatto e il diritto (*input*) l'algoritmo è in grado di produrre un risultato, una conclusione (*output*) che è conseguenza diretta e immediata delle premesse. Questo modello presuppone l'impiego di un algoritmo semplice, cui vengono fornite determinate istruzioni (premesse in fatto e in diritto) e che vengono da questo elaborate senza attingere ad informazioni o dati esterni, diversi da quelli forniti.

È un modello che, tuttavia, può ben funzionare solo per giudizi a bassa complessità, per tali intendendosi: a) quei giudizi su casi concreti di facile e pronta soluzione; *b)* quei giudizi che non implicano la necessità di cd. eterointegrazione fattuale e/o valoriale.

Con riferimento all'ipotesi *sub a)*, si vuol fare riferimento a quei casi, sottoposti all'attenzione del giudicante, che ben si attagliano ad un metodo decisionale automatizzato in quanto connotati da una scarna istruzione probatoria e/o la cui decisione implica *sic et simpliciter* l'applicazione di una regola di diritto al caso concreto.

Con riferimento all'ipotesi *sub b)*, invece, ci si riferisce a quei casi in cui l'algoritmo può funzionare senza necessità di apprendere nuovi dati e informazioni dalla realtà esterna ovvero quando può procedere all'applicazione di una regola giuridica posta da una norma di legge dal contenuto precettivo di immediata applicazione, ovverosia quelle fattispecie compiutamente determinate. Ne restano cioè esclusi quei processi decisionali che richiedono, invece, una integrazione del compendio fattuale e/o un giudizio di ordine valoriale o per clausole generali.

Infatti, come osservato[44], il diritto civile sostanziale conosce i cd. modelli normativi "aperti", in cui la determinatezza della fattispecie va riempita *aliunde* ad opera dell'interprete e del giudice; inoltre il legislatore sostanziale ha predisposto una serie di clausole codicistiche generali, che riferiscono a concetti elastici, quali la buona fede, la correttezza, la volontà effettiva delle parti.

Ciò implica che l'attività del giudicante si sostanzia in un giudizio di applicazione delle clausole generali al caso concreto e decida di questo solo sulla base dei canoni di interpretazione giurisprudenziale. In questo senso,

[44] V. ancora F. Patroni Griffi, *Tecniche di decisione* cit., 178.

il metodo decisionale non sarà quello del sillogismo, bensì la cd. tecnica delle clausole generali che non presuppone una fattispecie determinata: anzi, «*a ben guardare, nemmeno di una fattispecie*»[45]. Ecco dunque che il giudice deve farsi "inventore"[46] di una fattispecie "giurisprudenziale", che potrà essere suscettibile di previsione mediante utilizzo di algoritmi: ma questo attiene alla prevedibilità *della* decisione giudiziaria, anche di quella che concerne una fattispecie "giurisprudenziale".

La questione qui attiene invece alla prevedibilità *nella* decisione giudiziaria, e quindi la domanda che si pone all'attenzione è la seguente: è possibile prevedere – mediante algoritmi – una decisione del giudice "inventore" di una fattispecie giurisprudenziale? È possibile cioè riprodurre il metodo decisionale del giudice persona fisica in termini computazionali, laddove quel metodo non sia strettamente sillogistico ma si traduca nell'eterointegrazione fattuale della fattispecie posta (incompiutamente) dalla norma? E ancora: è possibile tradurre in linguaggio computazionale un giudizio valoriale del giudice?

È dunque un problema *di metodo di decisione* del giudice[47], ma è ancor prima un problema di *cognizione* del giudice.

Sotto questo profilo può osservarsi che alcuni sistemi di IA hanno la capacità di autoapprendimento, ovvero la capacità di assumere ed elaborare dati che vadano oltre al *dataset* di base, e che derivano dalla programmata elaborazione di operazioni poste in essere frequentemente. Quindi lo strumento algoritmico è in grado di generare un *output* che è frutto dell'elaborazione di informazioni e di pregresse esperienze; ma questa metodologia è applicabile alle decisioni del giudice?[48]

Occorrerebbe prima chiedersi *come* decide il giudice. Secondo le indicazioni di Natalino Irti[49], è possibile distinguere tra diverse forme di decisione: decisione secondo la legge, decisione secondo i precedenti, decisione secondo il fatto, decisione secondo i valori, secondo una scala decrescente di calcolabilità. L'ordinamento costituzionale «ha scelto il

[45] V. ancora F. Patroni Griffi, *Tecniche di decisione* cit., 179.

[46] Nel senso latino di *invenire*, i.e. trovare il diritto: cfr. P. Grossi, *Sull'odierna incertezza del diritto*, *Giustizia civile*, 2014, 18.

[47] A. Carcaterra, *Machinae autonome e decisione robotica*, in *Decisione robotica* (a cura di) A. Carleo, Bologna 2019, 51 si domanda dunque se la macchina potrebbe decidere in senso simile a quanto fa il giudice, sulla base dell'analisi di alcuni fatti che possano essere comunicati alla macchina stessa o che quest'ultima raccoglie autonomamente e sulla base delle norme giuridiche che regolano un determinato settore.

[48] Mammone, *Considerazioni introduttive* cit., 25.

[49] N. Irti, *Per un dialogo sulla calcolabilità giuridica*, in *Calcolabilità giuridica* (a cura di) A. Carleo, Bologna 2017, 20 ss, e spec. 26.

decidere secondo la legge», poiché si tratta di una decisione che assicura il massimo grado di "calcolabilità" in senso weberiano[50]. Per cui può ipotizzarsi che la decisione secondo legge (nei modelli normativi "chiusi") sia algoritmizzabile, mentre la decisione su valori diviene più problematica, e dipende da come li si interpreta[51]. A prescindere da ciò, alcuni autori ritengono che comunque la decisione per valori «*conduce ad una incalcolabilità giuridica*», a meno che i valori non siano "misurabili"[52].

3.1 *Profili di criticità della decisione algoritmica: ambito di operatività limitato, apparato rimediale, motivazione e responsabilità del giudicante*

È condividibile l'opinione di chi[53] ritiene che l'uso legittimo di un algoritmo a fini della decisione robotica soggiaccia ad un insieme di regole fondamentali volte ad assicurare la comprensibilità dei processi decisionali, il diritto delle parti soggette al processo algoritmizzato di richiedere spiegazioni o esperire rimedi, la responsabilità di una persona fisica o per gli effetti derivanti dall'uso di un sistema algoritmico.

Tale considerazione disvela che, oltre agli esaminati *impasse* di carattere logico-matematico e giudico, la teorizzazione di una decisione totalmente algoritmizzata in rapporto all'attuale sistema di diritto processuale civile sconta tre ulteriori ordini di limiti.

Il primo, già in parte evidenziato, concerne l'ambito oggettivo di operatività degli algoritmi in funzione (totalmente o parzialmente) sostitutiva del giudice umano, che – per le ragioni suesposte – rischierebbe

[50] È stato acutamente osservato in dottrina che la prima fase della decisione secondo legge e la decisione secondo precedenti sono entrambe «*fondate sulla identificazione della fattispecie*»; pertanto tali forme di decisione sarebbero agevolmente riproducibili in termini algoritmici, mediante un sistema di analisi dei dati passati che sia in grado di operare una classificazione per fattispecie. E ciò in quanto «*l'esame dei dati passati per prendere una decisione è pertanto fatto comune tanto alla decisione robotica quanto a quella giuridica*»: cfr. A. Carcaterra, *Machinae autonome e decisione robotica*, in *Decisione robotica* (a cura di) A. Carleo, Bologna 2019, 53.

[51] Si rinvia anche qui a N. Irti, *Per un dialogo sulla calcolabilità giuridica* cit., 23-24. L'a. distingue due possibili accezioni di "valori": come principi generali dell'ordinamento giuridico, ricavabili mediante un processo di generalizzazione dalle norme positive; ovvero come criteri-metapositivi.

[52] A. Carcaterra, *Machinae autonome e decisione robotica* cit., 57. L'a. ritiene che la misurazione dei valori possa avvenire mediante la cd. metrica dei valori

[53] D. De Kerckhove, *La decisione datacratica*, in *Decisione robotica* (a cura di) A. Carleo, Bologna 2019, 105. L'a. richiama i risultati raggiunti sul punto da un gruppo di Bertelsmann Stiftung.

di essere fortemente limitato alle sole controversie seriali[54], alle cause molto semplici, da istruirsi documentalmente[55], alle cause standardizzabili (*i.e.* cause che si connotano per una maggiore presenza di fattori di automaticità e per una giurisprudenza oramai consolidata (ad esempio, le cause di infortunistica stradale). Resterebbero invece escluse le cause complesse da un punto di vista istruttorio e quelle caratterizzate da variabilità ed eterogeneità del quadro normativo di riferimento e/o da valutazioni rimesse alla discrezionalità del giudice[56] (quali, ad esempio, quelle che prevedono l'applicazione delle clausole di buona fede, dei concetti giuridici indeterminati o delle clausole generali).

In secondo luogo, alcune criticità sorgono anche con riferimento ai possibili rimedi esperibili avverso una decisione automatizzata. Gli interrogativi che si pongono all'attenzione del giurista sono due: l'uno, più specifico, attinente al perimetro di operatività delle censure in sede di giudizio di impugnazione, l'altro, più generale, attinente all'onere della motivazione sancito dalla nostra Carta costituzionale (art. 111 Cost.).

Con riferimento al primo, può osservarsi che difficilmente la sentenza sarebbe attaccabile sotto il profilo degli *errores in procedendo*, posto che l'errore informatico nel processo decisionale algoritmico è di non facile evidenza e riscontro (e dunque non agevolmente dimostrabile, poiché resta nella sequenza di istruzioni interne all'algoritmo di intelligenza artificiale)[57]. Ma, a ben vedere, sembrerebbero escluse anche le censure relative agli *errores in iudicando de iure*, posto che se la decisione del caso concreto è rimessa ad una combinazione algoritmica di fatti "rilevanti" e di norme nelle quali sussumere i primi, e se gli elementi della premessa maggiore e minore sono interamente immessi nel sistema informatico,

[54] Sebbene sul punto non vi sia concordia di opinioni. In particolare, si accolgono i rilievi di chi (L. De Renzis, *Primi passi nel mondo della giustizia «high-tech»* cit., 149-150), ha opportunamente evidenziato che la serialità della causa non è sintomo di automatica semplicità della stessa, e che anzi vi siano delle controversie seriali che si presentano complesse da un punto di vista tecnico-giuridico, che hanno anche una rilevanza nell'ambito economico-giuridico e che riguardano settori strategici per la società civile.

[55] M.R. Covelli, *Dall'informatizzazione della giustizia alla «decisione robotica»?* cit., 133, che richiama esempi in materia di diritto previdenziale (ATP).

[56] L. Breggia, *Prevedibilità, prevedibilità e umanità nella soluzione dei conflitti*, «Rivista Trimestrale di Diritto e Procedura Civile», fasc.1, 1 marzo 2019, 395 ss., per la quale «*là dove siano in gioco valutazioni discrezionali, deve essere la persona ad agire, non la macchina che necessariamente riduce la discrezionalità ad un calcolo probabilistico*».

[57] M. Luciani, *La decisione giudiziaria robotica* cit., 83. L'a. solleva diversi dubbi sul punto, rilevando che sovente l'errore della macchina è "invisibile".

quest'ultimo deciderà «*tota lege perspecta*»[58].

Sotto il secondo profilo, connesso al primo (poiché «il diritto di impugnazione si sostanzia proprio nella critica alla motivazione»[59]) è stato osservato che in conseguenza della segretezza o inintelligibilità della logica sottesa al processo decisionale[60], con la decisione algoritmica la motivazione, a bene vedere, finirebbe per assolvere essenzialmente solo alla funzione extraprocessuale di permettere il controllo sociale sull'operato del giudice-robot, e non anche a quella endo-processuale di consentire la verifica dell'*iter* logico seguito dal giudice[61].

Si pone dunque un problema di verificabilità del ragionamento logico seguito dall'algoritmo, al fine della sua censurabilità in sede di impugnazione. Pur ammettendo che la parte interessata ad impugnare la decisione algoritmica riesca a scalfire l'iter logico-matematico seguito dall'algoritmo decidente, sorge l'ulteriore problema dell'effettiva rispondenza del giudizio di impugnazione così introdotto alle esigenze di giustizia: in altri termini, anche l'eventuale impugnazione, così algoritmizzata, resterebbe priva di quel carattere di "ripensamento" del percorso decisionale[62].

In ultima analisi, preme osservare come alla decisione algoritmica si accompagna poi la più complessa questione della responsabilità del decidente e della riparabilità degli errori giudiziari [63], da esaminarsi primariamente sotto tre diversi punti-chiave:

1 è socialmente accettabile una responsabilità connessa alla decisione algoritmica? Secondo alcuni autori[64], il punto della responsabilità potrebbe essere affrontato considerando che se il giudizio operato dal robot è frutto di una metrica di valori condivisa e accettata dalla collettività, allora le conseguenze di tale decisione algoritmizzata sono inevitabili e devono essere considerate come «il minimo danno

[58] M. Luciani, *La decisione giudiziaria robotica* cit., 89-90.

[59] E. Vincenti, *Il «problema» del giudice robot*, in *Decisione robotica* (a cura di) A. Carleo, Bologna 2019, 121.

[60] V. G. Resta, *Algoritmi, diritto, democrazia*, in *Giustiziacivile.com*, 11 aprile 2019, 4. Sul problema del cd. black box v. anche F. Pasquale, *The black box society: The secret algorithms that control money and information*, Cambridge-London, 2015; D. Pedreschi - F. Giannotti et al., *Open the Black Box. Data-driven Explanation of Black Box Decision Systems*, «ArXiv», 1, 2018, 1-2.

[61] M. Luciani, *La decisione giudiziaria robotica* cit., 89-90.

[62] L. De Renzis, *Primi passi nel mondo della giustizia «high-tech»* cit., 148.

[63] A. Carcaterra, *Machinae autonome e decisione robotica* cit., 51.

[64] A. Carcaterra, *Machinae autonome e decisione robotica* cit., 61.

accettabile» dato che l'utilizzo di un determinato tool tecnologico è stato socialmente e giuridicamente accettato *a priori*.

2 qual è centro di imputazione della responsabilità e sulla base di quali criteri si fonda la riparabilità dell'errore giudiziario della macchina? La questione è alquanto complessa e meritevole di più approfondita trattazione; ci si limita in questa sede a rilevare che la stessa si rimette all'inquadramento della dinamica uomo-macchina in termini oggettivi (sotto lo statuto della responsabilità da prodotto difettoso di cui alla Direttiva 85/374/CEE) ovvero soggettivi (secondo la configurazione di un rapporto di agency[65]

3 a quali condizioni sussisterebbe la responsabilità civile del giudice-robot? La risposta al quesito richiederebbe un vaglio delle ipotesi di responsabilità civile del magistrato contemplate dalla Legge Vassalli (L. 117/1988), come modificata dalla L. 18/2015, sotto la lente della "verificabilità" in concreto, apparendo comunque difficoltosa la ricostruzione dei casi di dolo o colpa grave[66].

4. *Prospettive di regolazione della giustizia predittiva*

Le riflessioni sull'interazione tra algoritmi e decisioni giudiziarie sono già state avviate da tempo, con riferimento non solo al settore civile, sia nei sistemi ordinamentali di *civil law* sia in quelli di *common law*[67].

[65] G. Teubner, *Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten*, in *Ancilla Iuris*, 2018.

[66] Tra i casi di colpa grave, la legge in questione menziona la "violazione del diritto dell'Unione Europea", per tale intendendosi anche il mancato assolvimento dell'obbligo – laddove previsto, alla luce anche della giurisprudenza eurounitaria – di rinvio pregiudiziale alla CGUE nel caso di supposto contrasto tra la norma di diritto interno e la norma di diritto euro-unitario. È evidente dunque che l'algoritmo si troverebbe qui a dover effettuare una valutazione preliminare circa la sussistenza o meno di siffatto contrasto con il diritto dell'Unione; e che l'esito positivo o negativo di tale valutazione sarebbe idoneo ad incidere sull'*an* della responsabilità civile del magistrato. Se sussistono *in apicibus* dubbi, da un punto di vista tecnico-matematico, circa la riproducibilità tecnica di un siffatto ragionamento dubitativo in termini algoritmici, ne discenderebbe che l'errore del sistema di IA non sarebbe mai imputabile a titolo do "colpa grave".

[67] In particolare, in Francia significativi passi in avanti sono stati effettuati con l'adozione

L'Unione Europea ha recentemente elaborato diverse strategie e approcci regolatori in relazione alla cd. "cybergiustizia", al fine di fornire una visione chiara e completa di come le nuove tecnologie aiuteranno il sistema giudiziario a svolgere i suoi compiti implementando soluzioni che sostiene il lavoro di qualsiasi professionista coinvolto e fornisce o facilita la trasmissione di informazioni utili al giudizio di un caso portato a un tribunale.

Nel dicembre 2018 il CEPEJ ha adottato la Carta Etica Europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari[68], in cui si rappresenta che l'uso degli strumenti di IA, *in votis* destinati a migliorare l'efficienza e la qualità della giustizia, deve comunque assicurare il rispetto dei diritti fondamentali della persona. Per cui se l'elaborazione delle decisioni giudiziarie mediante strumenti di intelligenza artificiale è in astratto possibile in materia civile, commerciale e amministrativa (al fine di contribuire a migliorare la prevedibilità dell'applicazione della legge e la coerenza delle decisioni giudiziarie), ciò deve avvenire nel doveroso rispetto dei cinque principi ivi elencati[69].

della Loi pour une République numérique del 7 ottobre 2016 che ha introdotto il principio della generale pubblicazione gratuita di tutte le decisioni giudiziarie, precisando unicamente che «*cette mise à disposition du public est précédée d'une analyse du risque de ré-identification des personnes», anche se già da tempo erano state sviluppate soluzioni di giustizia predittiva mediante l'impiego di algoritmi che analizzano le decisioni dei giudici delle corti con lo scopo di anticipare l'esito del giudizio (è il caso della piattaforma digitale Predictice.com* che consente il raffronto tra cause di risarcimento del danno precedenti e quella considerata, sfruttando il linguaggio "standard" impiegato nelle decisioni giudiziali, in modo tale da poter calcolare le probabilità di successo di un procedimento giudiziario e di ottimizzare la strategia processuale degli avvocati.). Negli Stati Uniti gli algoritmi di *predictive justice* sono stati impiegati prevalentemente nel settore penale (*Loomis v. Wisconsin*, 881 N.W.2d 749 (Wis. 2016)e lo Stato del New Jersey ha sostituito le udienze per la concessione della libertà su cauzione con delle valutazioni di rischio ottenute attraverso algoritmi, di talché chiunque può essere rilasciato, anche senza pagare una somma di denaro, se risponde a certi criteri. Per garantire decisioni scientifiche e imparziali, i giudici usano dei punteggi generati dalle macchine. (V. E. Livni, *Nei tribunali del New Jersey è un algoritmo a decidere chi esce su cauzione,* in *Internazionale.it*, 2017).

[68] La "European Ethical Charter on the use of AI in the judicial systems and their environment" è stata adottata il 3-4 dicembre 2018), e rappresenta il primo tentativo di porre le basi per una regolamentazione a livello europeo dell'utilizzo dei sistemi di AI nei sistemi giudiziari.

[69] Si tratta del principio che impone di garantire che la progettazione e l'attuazione di strumenti e servizi di intelligenza artificiale siano compatibili con i diritti fondamentali; del principio di non discriminazione, del principio di qualità e sicurezza, che richiede per quanto riguarda il trattamento delle decisioni e dei dati giudiziari, di utilizzare fonti certificate e dati immateriali con modelli elaborati in modo multidisciplinare, in un ambiente tecnologico sicuro; del principio di trasparenza, imparzialità ed equità

Si precisa inoltre che laddove lo strumento algoritmico sia stato progettato per finalità decisionali, è essenziale che il trattamento venga effettuato con trasparenza, imparzialità ed equità, da una valutazione esterna e indipendente di esperti. Per cui, i giudici devono essere sempre messi in condizione di poter rivedere e riesaminare le "decisioni algoritmiche", in relazione alle peculiarità del caso concreto.

Da ultimo, la Proposta di Regolamento della Commissione Europea per la disciplina dei sistemi di Intelligenza Artificiale[70], adottata lo scorso 21 aprile, offre un primo quadro legislativo generale a livello europeo per queste nuove tecnologie.

Per ciò che rileva ai fini del nostro discorso, le applicazioni di Intelligenza Artificiale che riguardano l'amministrazione della giustizia vengono qualificate come ad alto rischio "stand alone"[71]. In particolare, secondo l'Allegato III della citata Proposta, si tratta dei sistemi di intelligenza artificiale destinati ad «*assistere un'autorità giudiziaria nella ricerca e interpretare i fatti e la legge e applicare la legge a un insieme di fatti concreti*»

Da ciò deriva l'obbligo, per i progettisti e gli utilizzatori di sistemi di AI ai fini dell'adozione di una decisione giudiziaria, di osservare le rigide prescrizioni dettate dal legislatore europeo.

Infatti, la Proposta di Regolamento della Commissione prescrive una serie di obblighi di *accountability* sia per i *provider* di applicazioni di

che postula l'accessibilità e la comprensibilità dei metodi di trattamento dei dati; e del principio "sotto il controllo dell'utente", volto ad un approccio prescrittivo e a garantire che gli utenti siano attori informati e nel controllo delle scelte fatte.

[70] *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts* COM(2021) 206 final, disponibile al seguente link: https://ec.europa.eu/transparency/regdoc/rep/1/2021/EN/COM-2021-206-F1-EN-MAIN-PART-1.PDF .

[71] L'espressione "stand alone" si riferisce all'autonomia dell'applicazione (e dunque del software) rispetto un hardware fisico: le app non sono, cioè, componenti di prodotti fisici. In base al Considerando n. 40 della Proposta di Regolamento «*Alcuni sistemi di IA destinati all'amministrazione della giustizia e dei processi democratici devono essere classificati come ad alto rischio, considerando il loro impatto potenzialmente significativo su democrazia, Stato di diritto, libertà individuali e diritto a un ricorso effettivo e ad un giusto processo. In particolare, per affrontare i rischi di potenziali pregiudizi, errori e opacità, è opportuno qualificare come sistemi di IA ad alto rischio quelli destinati all'assistenza nella autorità giudiziaria nella ricerca e interpretazione dei fatti e della legge e nell'applicazione della legge ad un insieme concreto di fatti». Tuttavia, occorre precisare che lo stesso Considerando esclude dalla definizione di "alto rischio" i sistemi di IA destinati ad "attività amministrative puramente accessorie che non incidono sull'effettiva amministrazione della giustizia in casi individuali, come l'anonimizzazione o pseudonimizzazione di decisioni giudiziarie, documenti o dati, comunicazione tra personale, compiti amministrativi o allocazione di risorse*".

AI sia per gli utenti: tali regole riguardano pertanto tutto il ciclo di vita dell'applicazione di AI, a partire dalla configurazione dei *data set*, elemento strategico in ogni settore ma soprattutto in quello della giustizia predittiva.

Specificatamente, il primo requisito indispensabile è rappresentato dall'alta qualità dei dati, funzionale ad evitare l'insorgere di *bias* e discriminazioni *by design*. Si richiede infatti che i dati siano pertinenti, rappresentativi, completi e corretti con riguardo allo scopo specifico a cui tende l'applicazione di AI in particolare.

In secondo luogo, il data set deve essere formulato con appropriate proprietà statistiche, per salvaguardare la diversità di condizione e contesto[72].

Inoltre, le *legal tech* operanti in questo settore sono tenute a tutta una serie di adempimenti riguardanti la documentazione e la registrazione, la trasparenza e fornitura di informazioni agli utenti, l'obbligo di supervisione umana, la robustezza, la precisione e la sicurezza, con riguardo ad ogni componente del sistema, compresa la infrastruttura.

Con riferimento agli obblighi degli utenti, è significativo che la proposta di Regolamento abbia inteso disciplinare – seppur in via generale – l'utilizzo di applicazioni di AI ad alto rischio nel *decision making* dei giudici. L'addentellato positivo di riferimento è rappresentato dall'art. 29, in base al quale gli strumenti di giustizia predittiva dovranno essere utilizzati in conformità con le istruzioni d'uso a corredo dei sistemi e nel rispetto del Diritto dell'Unione e del diritto nazionale. Resta ferma la discrezionalità dell'utente nell'organizzare le proprie risorse e attività al fine di attuare le misure di supervisione umana indicate dal fornitore. Inoltre, si prevede che nel caso in cui gli uffici giudiziari partecipino alla messa a punto dei data-set giudiziari, saranno chiamati a garantire che i dati di *input* siano pertinenti rispetto a quanto previsto scopo del sistema di IA ad alto rischio.

Infine, sono previsti obblighi di informazione al fornitore se vi è motivo di ritenere che l'uso conforme con le istruzioni per l'uso possa generare un rischio o se sono stati identificati incidenti gravi o malfunzionamenti.

[72] Si fa riferimento, ad esempio, alle aree geografiche, al contesto comportamentale e funzionale, alla specificità di certi dati personali, con la possibilità di attuare il monitoraggio, la individuazione e la correzione di eventuali *bias*.

5. *Conclusioni*

Nell'era della pervasività della digitale numerosi tentativi di automatizzazione coinvolgono anche le attività "ontologicamente umane", quali quelle che costituiscono esercizio della funzione di *ius dicere*.

Tali tentativi, condotti sotto il vessillo della certezza e dell'oggettività del diritto, dell'efficienza e del rigore metodologico dei processi decisionali, hanno condotto alle teorizzazioni sulla prevedibilità della e nella decisione giudiziaria, consentendo di vagliare la permeabilità del sistema processuale alle istanze innovatrici della tecnologia.

Si è dimostrato che, allo stato dell'arte, l'idea di una giustizia totalmente algoritmizzata sconta una pluralità di criticità e di limiti, di ordine tecnico-matematico, giuridico e soprattutto di effettività della tutela giurisdizionale.

Ma ciò non vuol dire rifuggire dall'idea di realizzare concrete applicazioni degli algoritmi alle decisioni giudiziarie (nella duplice accezione di prevedibilità supra delineata), nel condivisibile convincimento che la giustizia predittiva non è necessariamente un male, ma lo è certamente un diritto incalcolabile: l'impiego di un metodo matematico per l'interpretazione e applicazione della legge[73], per la previsione delle risposte del sistema giustizia e per la risoluzione delle controversie, assicura una maggiore stabilità degli orientamenti giurisprudenziali, rafforzando il valore del precedente, e soprattutto consente di superare i limiti intrinseci del *prospective overruling* sia sostanziale sia processuale che – ad oggi – ha un ambito di applicazione limitato.

Poiché alla certezza del diritto deve accompagnarsi la certezza dell'interpretazione e dell'applicazione del diritto stesso, l'elaborazione di algoritmi funzionali (quantomeno) all'interpretazione della legge potrebbe fornire un concreto apporto all'attività dei giudici, per ripristinare il "vincolo della fattispecie"; per dosare il ruolo e l'effettività del "vincolo del precedente"[74], e per contribuire ad una gestione efficiente e sostenibile del sistema giustizia.

D'altronde, l'art. 101 Cost. e l'art. 65 ord. giud. esprimono l'impersonale oggettività del diritto e la funzionalità tecnica della sua applicazione,

[73] V. L. Viola, *Interpretazione della legge con modelli matematici. Processo, a.d.r., giustizia predittiva*, vol. I, Milano, 2017, p. 82, il quale ritiene che «proprio per il tramite di formule matematiche, l'interpretazione giudiziale possa essere prevista in conformità all'esigenza di certezza del diritto, intesa appunto non solo come prevedibilità della disposizione di legge applicabile, ma anche prevedibilità dell'esito giudiziale».

[74] M. Nuzzo, *Il problema della prevedibilità delle decisioni: calcolo giuridico secondo i precedenti*, cit.

vietando pre-giudizi e pre-comprensioni[75]; per cui se il diritto è oggettivo, nel senso di avere una base di regole predeterminate e vincolanti, allora deve essere possibile prevederne l'applicazione.

Tuttavia, occorre evitare di confondere i valori dell'indipendenza e dell'autonomia di giudizio con impossibili e improprie funzioni neutralizzanti, tese cioè a fornire l'illusione dell'infallibilità e dell'oggettività dell'applicazione del diritto[76], almeno per quattro ragioni.

In primis, perché l'algoritmo non può costituire un filtro per il diritto impersonale, perché dietro all'algoritmo c'è sempre l'uomo[77] e perché l'attività di interpretazione e applicazione della norma è strutturalmente connotata da un "insopprimibile soggettivismo"[78].

Inoltre un'ottica generale di progresso dell'ordinamento giuridico-sociale, andrebbe sempre evitata la cristallizzazione delle decisioni supportate da algoritmi preservando l'evoluzione della giurisprudenza[79].

Si consideri poi che il grado di affidabilità e di attendibilità dell'algoritmo dipende dalla qualità dei dati utilizzati e dalle procedure risolutive scelte[80].

Infine, perché al centro di qualsiasi ordinamento giuridico vi è la persona umana, e bisogna assicurare sempre lo spazio per un vero e proprio diritto processuale dell'uomo e per l'uomo, col fermo proposito di avvicinarsi all'ideale della «*miglior giustizia attraverso maggior libertà*»[81].

Il quadro normativo attuale, anche non specificatamente riferibile al

[75] L. Breggia, *Prevedibilità, prevedibilità e umanità nella soluzione dei conflitti,* cit., 395 ss., la quale icasticamente ritiene che «*la macchina categorizza, necessariamente astrae, a rischio di travisamenti e pregiudizi*»

[76] V. nota precedente.

[77] Vincenti, *Il «problema» del giudice robot* cit., 118 ricorda che per i filosofi contemporanei «*la tecnologia non è neutra, ma fa politica*».

[78] N. Irti – E. Severino, *Dialogo su diritto e tecnica*, Roma-Bari, 2001. La componente soggettiva e della decisione giudiziaria può financo condurre ad una decisione maggiormente imparziale: v. A. Callegari, *Il giudice tra emozioni, biases ed empatia*, Padova, 2017.

[79] M.R. Covelli, *Dall'informatizzazione della giustizia alla «decisione robotica»?* cit., 133. V. anche L. De Renzis, *Primi passi nel mondo della giustizia «high-tech»* cit., 148, la quale elenca, tra i vari "rischi" della decisione robotica: la predominanza del criterio interpretativo letterale; l'annichilimento del ricorso all'analogia; la (paradossale) introduzione del "pre-giudizio", poiché l'algoritmo decide sulla base dei soli dati che vengono immessi dal programmatore; il definitivo superamento del momento valutativo e assiologico che è proprio dell'*agere* del decisore umano.

[80] E. Gabellini, *La «comodità nel giudicare»: la decisione robotica*, in *Rivista Trimestrale di Diritto e Procedura Civile*, fasc.4, 1 dicembre 2019, 1305 ss., per la quale l'effettività dello strumento robotico dipenda dal "patrimonio giuridico" che si decide di inserire all'interno del *software.*

[81] A. Panzarola, *Una lezione attuale di garantismo processuale* cit., 165.

processo civile[82], sembra escludere una decisione algoritmica *tout court*, ovverosia basata esclusivamente sul ragionamento logico "robotico", e anche le prime istanze regolatorie a livello euro-unitario sembrano orientarsi nel senso di una partecipazione dei sistemi di IA meramente ausiliaria alla disciplina dell'attività giurisdizionale.

[82] Si segnalano, in particolare: l'art. 8 del d.lgs. n. 51 del 18 maggio 2018 di attuazione della Direttiva UE 680/2016 in materia penale sul "processo decisionale automatizzato relativo alle persone fisiche"; l'art. 22 del GDPR in materia di trattamento automatizzato di dati personali, in forza del quale nessun atto o provvedimento giudiziario o amministrativo che implichi la valutazione di un comportamento umano può essere fondato esclusivamente su un trattamento automatizzato di dati personali volto a definire il profilo o la personalità del soggetto interessato; la Risoluzione del Parlamento Europeo del 16 febbraio 2016 recante raccomandazioni alla Commissione Europea concernenti norme di diritto civile sulla robotica, laddove si afferma l'imprescindibilità del giusto processo, della trasparenza del processo decisionale robotico e la necessità del controllo e della verifica da parte dell'uomo.

SECTION II

Antitrust, A.I. and Big Data

Fabiana Di Porto, Tatjana Grote,
Gabriele Volpi, Riccardo Invernizzi

Talking at Cross Purposes? A computational analysis of the debate on informational duties in the digital services and the digital markets acts

Abstract: Since the opaqueness of algorithms used on online platforms opens the door to discriminatory and anti-competitive behaviour, increasing transparency has become a key objective of lawmakers. Leveraging the analytical power of Natural Language Processing, this paper investigates whether key terms related to transparency in digital markets were used in the same way by different stakeholders in the consultation on the EU Commission's DSA and DMA proposals. We find significant differences in the employment of terms like 'simple' or 'meaningful' in the position papers that informed the drafting of the proposals. These findings challenge the common assumption that phrases like 'precise information' are used the same way by those implementing transparency obligations and might partially explain why they frequently remain ineffective.

1. *Introduction*

When EU Executive Vice-President Margarethe Vestager presented the latest Commission proposals on digital platforms, the Digital Markets Act (DMA) and the Digital Services Act (DSA)[1], she compared them to the invention of the traffic light, which was created in response to the rapidly

* This article was first published in *Journal of Regulation and Technology*, 2021, pp. 87-106. We are thankful to Prof. Julian Nyarko from Stanford University for providing some of the data we used in the empirical part and to a number of other people for insightful discussions: Dr Roberto Marseglia, Professors Daniela Piana and Giuseppe Italiano and the anonymous referees. All mistakes remain our own.

[1] European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM(2020)825), 15 December 2020 [hereinafter Digital Services Act, DSA]; European Commission, Proposal for a Regulation of the European Parliament and the Council on contestable and fair markets in the digital sector (Digital Markets Act) (COM(2020) 842), 15 December 2020 [hereinafter Digital Markets Act, DMA].

increasing importance of the car. She concluded that 'just like back then, … now we have such an increase in the online traffic that we need to make rules that put order in the chaos'[2].

This twin-proposal suggests many new rules for digital intermediary services and online platforms[3]. With the DSA and DMA, the Commission closes a period during which stakeholders (and doctrine)[4] have been harshly discussing new ex ante rules for digital markets, both from a consumer protection and a competition law perspective[5].

Although the two proposals differ in scope and focus[6], both reveal

[2] European Commission, Statement by Executive Vice-President Vestager on the Commission proposal on new rules for digital platforms, 15 December 2020, <https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_20_2450>, (accessed 15 February 2021).

[3] There is no perfect alignment in the definition of platform services in the DSA and DMA. In the DSA, the widest concept is that of online 'intermediary service', which covers all services within the scope of Art. 1(3), including 'online platforms' (providing hosting services) under the meaning of Art. 2(1)h DSA. In the DMA, the widest category is that of 'core online platform'. Art. 2(2) 'online intermediation services' are one service type among the many 'core platform services' (together with e.g., cloud services, social networks, video-sharing platforms). Some 'core online platforms', then, may be designated as 'gatekeepers' (DMA, Art. 3) if they (a) have a significant impact on the internal market, (b) serve as a gateway between business and end-users, (c) enjoy an entrenched and durable position. The requisites are presumed to exist: (i) if the 'core platform service' was provided in at least 3 MS and given thresholds of average market capitalization are overcome; (ii) the core platform has more than 45 million monthly active end-users plus 10.000 business users; (iii) the thresholds in point (ii) were met in each of the last three financial years. (Art. 3, DMA). Hence, for the sake of parallel applicability of the DSA and DMA transparency rules, not every (core) very large platform is a gatekeeper, but it is likely that every gatekeeper will also be a very large (core) online platform (see Art 3(2)b DMA).

[4] P. Ibáñez Colomo, *Whatever Happened to the 'More Economics-Based Approach'?*, in *Journal of European Competition Law & Practice*, vol. 11, 9, 2020, pp. 473–74, (discussing the shift from the so called 'more economic approach' to the growing demand for ex ante intervention against big digital platforms in the European legal community).

[5] For challenges related to competition law, see e.g., A. Ezrachi & M. Stucke, *Virtual competition: the promise and perils of the algorithm-driven economy*, Harvard University Press, 2016, and P. Marsden & R. Podszun, *Restoring Balance to Digital Competition – Sensible Rules, Effective Enforcement*, Konrad-Adenauer-Stiftung, 2020, pp. 1-87. On consumer protection and its relation to data protection and competition law, see W. Kerber, *Digital markets, data, and privacy: competition law, consumer law and data protection*, in *Journal of Intellectual Property Law & Practice*, vol. 11(11), 2016, pp. 856-866.

[6] Both the DMA and DSA take a resolute stance, through ex ante regulation, against the big platforms. However, the DSA aims primarily to 'ensur[e] a safe and accountable environment' by applying asymmetric ex ante rules to online digital platforms, according to two parameters: the company's role (i. intermediary services, ii. hosting services, iii. online platforms), and size (a. large online platforms and b. very large platforms i.e.,

that one key instrument the Commission relies upon in 'ordering' chaotic traffic in digital markets is informational duties (inclusive of both transparency and disclosure obligations)[7].

This is surprising and unsurprising at the same time. According to the standard narrative, informational duties play a central role in the realm of consumer protection[8] and serve to rebalance unequal bargaining power in trade relationships[9]. And digital markets would be no exception[10].

On the other hand, the very utility of informational duties has been systematically questioned[11]. Overall, such duties seem to have more

those reaching more than 45 million consumers, which will have to comply with special rules). The DSA imposes obligations on transparency, illegal content, and accountability requirements. Therefore, it addresses negative externalities and asymmetric information. On the other hand, the DMA's goal is to 'ensur[e] fair and open digital markets' by applying asymmetric rules against large online platforms designated as 'gatekeepers', which are addressed with a list of does and don'ts. Taken together, they can be read as an ex ante toolbox, made of a mix of competition and consumer protection rules. While the DSA amends the e-commerce directive (2000/31/EC), the DMA centers around concerns and seeks to complement EU competition rules (mostly Art 101, 102 TFEU). Finally, the DSA applies to all 'intermediary services' (Art 1), while the scope of the latter is limited to 'core platform services' offered by 'gatekeepers' as defined in Art 3 DMA.

[7] We use disclosure, transparency and informational duties interchangeably as what is relevant to the analysis is the way the terms related to the provision of information are used by the stakeholders. However, we acknowledge that there are duties owed to users and those to public authorities; and that information may well be provided for purposes of public or private disclosure, or for reasons of investigations. A taxonomy of transparency and disclosure duties is nonetheless provided for in Table 1 in the Appendix, to which reference is made in the legal analysis of Section 2.3 below.

[8] European Parliament resolution of 20 October 2020 with recommendations to the Commission on the Digital Services Act: Improving the functioning of the Single Market (2020/2018(INL)), 20 October 2020, 12 (no. 31, 32).

[9] See e.g., E. A. Posner, *ProCD v. Zeidenberg and Cognitive Overload in Contractual Bargaining*, in *U. of Chicago L. Rev.*, vol. 77(4), 2010, pp. 1181-1194.

[10] Algorithm Watch, *Governing Platforms – Final Recommendations*, 2020, available at https://algorithmwatch.org/wp-content/uploads/2020/10/Governing-Platforms_DSA-Recommendations.pdf (accessed on Feb. 17, 2021), 1.

[11] See e.g., O. Ben-Shahar & C. E. Schneider, *Coping with the Failure of Mandated Disclosure*, in *Jerusalem Rev. of Legal Studies*, vol. 11(1), 2015, pp. 83–93; F. Marotta-Wurgler, *Even More Than You Wanted to Know About the Failures of Disclosure*, in *Jerusalem Rev. of Legal Studies*, vol. 11(1), 2015, pp. 63–74. E. Zamir & D. Teichman, *Behavioral Law and Economics*, Oxford University Press, 2018, pp. 171-177; F. Di Porto, & M. Maggiolino, *Algorithmic Information Disclosure by Regulators and Competition Authorities*, in *Global Jurist*, vol. 19(2), 2019, p. 11; E. Bardach & R. A. Kagan, *Going by the book: The problem of regulatory unreasonableness*, Temple University Press, 1982, pp. 249-256; A. Prat, *The Wrong Kind of Transparency*, in *American Economic Rev*. vol.

of a symbolic (*rectius*, political) value rather than true utility[12]. In the digital realm, many argue that extra-long disclaimers and hard-to-read terms of contract would be useless, or sometimes run counter consumers empowerment[13]. A similar argument is made for platform-to-business relations, where information duties are often considered insufficient to mitigate unequal bargaining power[14].

This paper aims to investigate *why*, despite the long-lasting scholarly debate about their limited effectiveness, and overwhelming evidence supporting it, the DSA and DMA rely heavily on disclosure[15]. More specifically, we investigate what are the possible *sources* of ineffectiveness.

There have been many attempts to do that, the behavioral literature on disclosure being the most relevant in two regards. On one side, it has provided empirical evidence of the impact of informational arrangements[16] adopted by big digital platforms by measuring how much they affect the behavior of consumers. On the other, it has accounted for the effectiveness of disclosure duties by measuring how many consumers like or dislike them[17]. However, these studies take the legal duty as a given, an external

95(3), 2015, p. 862.

[12] Di Porto & Maggiolino, *supra* note 1, p. 14.

[13] S. K. Ripken, T*he Dangers and Drawbacks of the Disclosure Antidote: Toward a More Substantive Approach to Securities Regulation*, in *Baylor L. Rev.*, vol. 58(1), 2006, p. 160.

[14] Marsden & Podszun , *supra* note 5, p. 18; F. Di Porto & M. Zuppetta, *Co-Regulating Algorithmic Disclosure for Digital Platforms*, in *Pol'y and Society*, vol. 0(0), 2020, p. 3-4; C. Busch, *Crowdsourcing, Consumer Confidence: How to Regulate Online Rating and Review Systems in the Collaborative Economy*, in C. Economy & A. De Franceschi (Eds.), *European Contract Law and The Digital Single Market: The Implications of The Digital Revolution*, 2016, p. 223.

[15] See M. Sentfleben & C. Angelopoulos, *The Odyssey of the Prohibition on General Monitoring Obligation on the Way to the Digital Services Act: Between Article 15 of the e-Commerce Directive and Article 17 of the Directive on Copyright in the Digital Single Market*, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3717022 (accessed on April 23, 2021) and G. Frosio, *Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime*, 2020, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3747756, discussing transparency duties in the DSA. For an analysis of disclosure remedies in the DMA, see P. Ibáñez Colomo, *The Draft Digital Markets Act: A Legal and Institutional Analysis*, 2021, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3790276 (accessed on April 23, 2021).

[16] See e.g., J. Luguri & L. Strahilevitz, *Shining a Light on Dark Patterns*, 2021, https://doi.org/10.2139/ssrn.3431205 (accessed 26/06/2021) (discussing the impact of dark patterns, including informational ones).

[17] See e.g., O. Katz & E. Zamir, *Do People Like Mandatory Rules? The Choice Between Disclosures, Defaults, and Mandatory Rules in Supplier-Customer Relationships*, in *JELS*, vol. 18(2), 2021, pp. 421-60 (who compare the desirability of disclosures duties, from

variable. On the contrary, we contend that much can be said about their origin and the process through which this duty is formed.

Therefore, we propose to leverage the power of computational tools, among which Natural Language Processing (NLP) and Machine Learning (ML) techniques: by linguistically analyzing the debate that preceded the adoption of these duties, our empirical study suggests searching for possible sources of failure in the feedback documents to the consultation, that were input to these rules.

Our contribution innovates in several regards. First, our methodology is not effects-based, in the sense that to assess the efficacy of transparency duties, it does not look at the impact on nor the perceptions of those who receive the information, being this input context-specific. We rather analyze the wording that conflated the debate around the provisions establishing informational duties of the DSA and DMA. Especially, we ask whether the meaning and use of terms that were discussed and finally became parts of information duties were fully shared among the stakeholders or not. For instance, terms like 'clear' or 'unambiguous' (referred to in Art. 24 DSA and extensively discussed before its adoption) are understood the same way by online platforms using personalized ads (addressed by the duty to disclose information) and the consumers (addressee of the information piece)? If this is not, could that be a source of disclosure ineffectiveness?

To assess if this is the case, we look at the stakeholder's submissions to the Commission's public consultation over three Inception Impact Assessment documents (IAs) that were input to the DSA and DMA proposals, namely: the so-called 'New Competition Tool',[18] the 'Ex ante regulatory instrument for large online platforms'[19] (hereafter also: ex ante tools), and the (then) 'Digital Services Act'[20].

the perspective of the consumer, as compared to mandatory rules and default rules).

[18] *New Competition Tool, Inception impact assessment*, Ares (2020) 2877634, 4 June 2020, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12416-New-competition-tool (accessed on Mar. 31, 2021).

[19] The Ex ante regulatory instrument for large online platforms with significant network effects acting as gate-keepers in the European Union's internal market, *Inception impact assessment*, Ares (2020) 2877647, 4 June 2020, https://ec.europa.eu/info/law/better-regulation/have-yoursay/initiatives/12418-Digital-Services-Act-package-ex-ante-regulatory-instrument-of-very-large-online-platforms-acting-as-gatekeepers

[20] The (then) Digital Services Act, Deepening the Internal Market and clarifying responsibilities for digital services, *Inception impact assessment*, Ares(2020)2877686, 4 June 2020, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12417-Digital-Services-Act-deepening-the-internal-market-and-clarifying-respon-

Second, we add computational analysis to standard manual reading of submissions that is done by the Commission without the help of algorithms[21]. The total of 2.862 replies to questionnaires and feedback documents contain the comments of all stakeholders regarding the proposals put forward by the Commission in its inception IAs. They, therefore, constitute an exceptional source of knowledge about who supported and opposed these duties among them, and especially, how individuals and organizations understand and use relevant terms of transparency. While manually processing the replies might still allow identifying the need for transparency duties, there are two shortcomings of this approach. First, any manual 'analysis' of the feedback documents comes with quite substantial labor cost, something that 'distant reading' can do more efficiently[22]. Second, no human reader can quantify the extent to which the same terms are used in the same way by different stakeholders. For instance, while both a large online platform and a consumer or smaller business might speak of a need for more 'precise' information, the underlying understanding and consequent use of this term could differ. In the context of transparency obligations, this is problematic since these duties might remain ineffective if a disclosure statement is only 'readable' in the eyes of the platform drafting it, but not in the eyes of the individual consumer or the micro organization reading it.

One way to cope with such limitations is to computationally analyze the feedback submitted to the Commission through the means of a mixed supervised and unsupervised ML technique, that would *complement* standard processing by public officials in the Directorates General (DG). Specifically, we propose doing so by using Word Embedding Alignment[23],

sibilities-for-digital-services_en.

[21] R. Senninger, *Analyzing the EU Commission's Regulatory Scrutiny Board through quantitative text analysis*, in *Regulation & Governance*, 2020, p. 1; C. M. Radaelli, *Regulating Rule-making via Impact Assessment*, in *Governance*, Vol. 23(1), 2010, pp. 89–108; C. A. Dunlop & C. M. Radaelli, *Impact Assessment in the European Union: Lessons from a Research Project*, in *European Journal of Risk Regulation*, vol. 6(1), 2015, pp. 27–34.

[22] J. Grimmer & B.M. Stewart, *Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts*, in *Political Analysis*, vol. 21(3), 2013, pp. 267–297.

[23] See e.g., D. Alvarez-Melis & T. S. Jaakkola, *Gromov-Wasserstein Alignment of Word Embedding Spaces*, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 1881–1890. Association for Computational Linguistics; N. Yehezkel Lubin, J. Goldberger, & Y. Goldberg, *Aligning Vector-spaces with Noisy Supervised Lexicons*, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, pp. 460–465.

a state-of-the-art model for translation[24], which can be adapted to our task, i.e. monolingual translation from a language to itself to evaluate the difference in the use of the same word in different corpora[25]. As a plus, word embedding modelling is highly compatible with unsupervised learning, a feature[26] that is very useful since, as explained before, in this context we should avoid the participation of human coding during the training process as much as possible.

This way, we aim to answer two central questions: (1) Do different groups of contributors share the same understanding (measured as semantical differences between terms) and use of the central terms and issues surrounding transparency and disclosure duties in the DSA and DMA? (2) Can we identify different clusters of opinions towards key concepts and can they be a possible source of disclosure failure? Our success in finding an answer to these questions with the help of said tools will be reflected with a view to a third overarching question: (3) can computational techniques help to partially automate the collection and analysis of opinions that are inputs to a rulemaking process? If this is the case, then we should recognize their potential in supporting the creation of better information disclosure rules, as is the proclaimed goal of the DSA and DMA consultation procedure, that is disclosure rules that are less prone to failure.

The article is structured as follows. The following section outlines the informational challenges posed by digital markets and the role of transparency duties set forth in the DSA and DMA proposals in mitigating their negative effects on consumers and businesses (Section 2). We then present our computational text analysis of the consultation documents and results, showing that not only are similar opinions expressed by groups that usually belong to different clusters (i.e., medium and big organizations); but also that groups of stakeholders use central

[24] A. Abdelsalam, O. Bojar & S. El-Beltagy, *Bilingual Embeddings and Word Alignments for Translation Quality Estimation. Proceedings of the First Conference on Machine Translation*, Vol. 2, Shared Task Papers, 2016, pp. 764–771.

[25] J. Nyarko & S. Sanga, *A Statistical Test for Legal Interpretation: Theory and Applications*, 2020, https://juliannyarko.com/wp-content/uploads/other/nyarko_sanga_legal_interpretation.pdf. (showing how word embedding modelling can fit very well our task).

[26] T. Wada & T. Iwata, *Unsupervised Cross-lingual Word Embedding by Multilingual Neural Language Models*, 2018, arXiv:1809.02306 [cs]; A. Conneau, G. Lample, M. Ranzato, L. Denoyer & H. Jégou, *Word Translation Without Parallel Data*, 2018, arXiv:1710.04087 [cs].

terms in different ways (Section 3). We lastly conclude by sketching how a similar procedure could help to draft smarter disclosure regulations in a larger context.

2. *Informational Malpratice in the Digital Era*

For many commentators, the prominent role of transparency obligations in the DSA and DMA did not come as a surprise.[27] Disclosure duties of all kinds have long been conceived as a key policy instrument to tackle the manifold challenges arising from digital markets. This section will give a snapshot of these challenges focusing and explaining the role of transparency in theory and in the DSA and DMA.

2.1. *Talking at Cross Purposes. The Debate on the Need to Update Informational Duties through the DSA and DMA*

Consumers benefit in many ways from the impressive development of digital markets.[28] However, certain characteristics of digital markets come with new challenges and risks. Concerning consumer protection, the sale of illicit goods in online marketplaces and unfair contractual clauses are key concerns.[29] But opaque online environments, as the Crémer report rightly emphasized, may also be 'a competition policy issue'.[30]

[27] See e.g., Global Network Initiative, *Thinking Through Transparency and Accountability Commitments Under The Digital Services Act*, 20 July 2020, https://medium.com/global-network-initiative-collection/thinking-through-transparency-and-accountability-commitments-under-the-digital-services-act-e4dce3cee909 (accessed on Jan. 22, 2021); S. Stolton, *Make Big Tech accountable, Austria says in Digital Services Act recommendations*, in *Euractiv*, 30 November 2020, https://www.euractiv.com/section/digital/news/make-big-tech-accountable-austria-says-in-digital-services-act-recommendations/ (accessed on Jan. 22, 2021).

[28] See Recital 1 DSA. To name just a few of these benefits: digital marketplaces facilitate cross-border trade and amplify product choices, social media allows cheap, easy, and quick communication, digital start-ups spur innovation and offer new services.

[29] Concerning contractual clauses, an empirical analysis has identified potentially unfair contractual clauses in roughly 10% of a sample of 50 online consumer contracts. M. Lippi, P. Pałka, G. Contissa, F. Lagioia, H. Micklitz, G. Sartor & P. Torroni, *CLAUDETTE: An automated detector of potentially unfair clauses in online terms of service*, in *Artificial Intelligence and Law*, vol. 27(2), 2019, pp. 117–139.

[30] J. Crémer Y. de Montjoye & H. Schweitzer, *Competition policy for the digital era, European Commission Report*, 2019, https://data.europa.eu/doi/10.2763/407537

The relationship between transparency on the one side, and competition law and consumer protection, on the other, is bidirectional. A lack of competition might force business users to accept a level of transparency they do not feel comfortable with, in absence of an alternative supplier of the online service they are consuming.[31] This is an important realization since digital markets show certain characteristics which are likely to favor highly concentrated markets.[32]

Taken together, these factors work in favor of large online platforms, which might accumulate some kind of 'gatekeeping' power and impose the level of transparency they deem appropriate on the market they dominate. Of course, they technically still underly certain transparency obligations, for instance, those included in the GDPR.[33]

However, the GDPR does not cover all relevant phenomena and users[34].

Furthermore, platforms' understanding of specific requirements like e.g., 'clear and easy' language, might effectively determine the usefulness of disclosures for consumers, the small and medium enterprises. When consumers are not able to switch to a different provider giving information in a way that better fits their needs and capacities, a lack of competition could thus result in a lack of transparency.

(accessed on Feb.14, 2021) [hereinafter Crémer Report], 63.

[31] This problem is well-framed as follows: 'a lack of options to switch to qualitatively similar other search engines or social networks might lead users to accept also very high prices (in form of collected data) and privacy policies that do not match their specific privacy preferences'. Kerber, *supra* note 5, p. 867.

[32] Crémer report, *supra* note 30, pp. 2-3; M. Gal & N. Petit, *Radical Restorative Remedies for Digital Markets*, in *Berkeley Technology L. Journal*, vol. 37(1), 2021, pp. 5-6; OECD, *Roundtable on Algorithms and Collusion - Executive Summary* (DAF/COMP/M(2017)1/ANN3/FINAL), 26 September 2018, 5; F. Scott Morton, P. Bouvier, A. Ezrachi, A. Jullien, R. Katz, G. Kimmelman, D. Melamed & J. Morgenstern, *Committee for the Study of Digital Platforms, Market Structure and Antitrust Subcommittee*, Stigler Center for the Study of the Economy and the State [hereinafter Stigler report], 2019, p. 14. P. G. Picht & G. T. Loderer, *Framing Algorithms: Competition Law and (Other) Regulatory Tools*, in *World Competition*, vol. 42(3), 2019, p. 406.

[33] Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 27 April 2016, O.J. L 119/1 [hereinafter GDPR].

[34] For instance, the GDPR is not really relevant for business users, for it covers the personal data of individuals only (Art 2(1) in connection with Art 4(1) GDPR). It does not touch on the circumstances under which data (or content) deliberately shared by an individual can be removed by a platform. Neither does it regulate how data shared by a business user of an intermediary service should be displayed and what the user ought to know about this, which is central from a competition perspective.

The other way around, there are also situations in which a lack of transparency can endanger competition due to allowing for certain anti-competitive practices. In its investigation report on competition in digital markets, the US Congress subcommittee on Antitrust, Commercial Law and Administrative Law has summarized this as follows: 'Without transparency or effective choice, dominant firms may impose terms of service with weak privacy protections that are designed to restrict consumer choice, creating a race to the bottom'[35]. Clearly, that depends on the fact that in digital markets products are mainly zero-priced, and 'privacy and quality of service can be differentiating factors'[36]; hence, granting transparency or effective choice can help ensure competition.

Such a problem may arise in case platforms manipulate the order in which offers from business customers are presented[37]. Only if the parameters used to rank products are transparent, it will be possible to know whether an online platform is distorting competition by preferencing certain offers[38], leaving consumers in the dark about the 'trade-offs they are facing', and hence inhibiting competition in a significant manner. In particular, self-preferencing by the big tech has been long debated as a cause of competition law infringement[39].

[35] U.S. House Committee on the Judiciary (2020). Investigation of Competition in Digital Markets. Washington, D.C.: Government Printing Office. The Subcommittee report also mentions manipulative design interfaces, so called dark patterns, nudging consumers into certain choices. Ibid, 53.

[36] Ibid, 54.

[37] Some authors argue that where consumer choices are being influenced, there is a special need for transparency duties: "A core element of such duties could be the obligation to thoroughly explain the workings of an algorithm, not on a technical level but regarding its impact on the customer, especially where it is designed to replace customer choice". Picht & Loderer, *supra* note 32, p. 416.

[38] Contra, L. Signoret, *Code of competitive conduct: a new way to supplement EU competition law in addressing abuses of market power by digital giants*, in *European Competition Journal*, vol. 16(2-3), 2020, 221, at 244 (contending that where platforms gain market power by being more efficient or winning consumers based on free choice by providing better offers, this would not constitute a violation of competition law).

[39] Self-preferencing was at the heart of the Microsoft saga (see J.P. Jennings, *Comparing the US and EU Microsoft Antitrust Prosecutions: How Level Is the Playing Field*, in *Erasmus Law and Economics Review*, vol. 2, 2006, pp. 71–86.) and was also heavily discussed by the doctrine at the time of the Google Shopping case. In fact, the Google Shopping case established that self-preferential placements are, indeed, not compatible with competition law. Google Search (Shopping) Case C(2017) 4444, 27 June 2017, paras 9, 10 of summary decision. See e.g., P. Ackman, *The Theory of Abuse in Google Search: A Positive and Normative Assessment Under EU Competition Law*, in *Illinois Journal of Law, Technology & Policy*, vol. 2017, (2), pp. 301–372.

2.2. *Legal Grounds for Updating Informational Duties*

In the debate on how to react to some of these challenges, the e-Commerce Directive (ECD) has been central[40]. It is the piece of legislation the DSA updates and amends as 20 years of technological developments necessarily opened up some transparency-related lacunas.

First, platforms have quite simply become significantly larger and more important[41]. And with the reach of platforms, the amount of user-generated content has increased exponentially[42]. Hence, it is the increase in volume and magnitude of markets that justify a different approach. Second, existing rules were adopted when content moderation by automated means was not yet a widespread practice, if available at all[43]. Third, the increased relevance of recommender systems, digital nudging, personalized advertising also did not exist and was therefore not addressed by the ECD[44].

Against the background of these developments, commentators and lawmakers have advocated in favor of significantly expanding the information duty framework of Arts 5, 6, and 10 ECD, with the aim of 'putting meaningful transparency at the heart' of new EU rules on digital services[45].

With regards to the DMA, general shortcomings of EU competition rules when dealing with opaque online practices have been highlighted[46],

[40] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on Electronic Commerce), 17 July 2000, O.J. L 178/1 [hereinafter ECD]; The ECD is considered by some as "the cornerstone of the Digital Single Market", European Parliament (n 8) 17.

[41] Given that they reach a massive number of users, illegal or otherwise problematic content and practices will now impact considerably more citizens. S. B. Micova & A. De Streel, *Digital Services Act – Deepening the Internal Market and Clarifying Responsibilities for Digital Services, Centre on Regulation in Europe Report*, 2 December 2020, https://cerre.eu/publications/digital-services-act-responsibility-platforms/ (accessed on Feb. 16, 2021) [hereinafter CERRE DSA Report], p. 10.

[42] Alarmingly, this development has been associated with a rise in hate speech and disinformation. European Parliament, *supra* note 8, p. 3.

[43] Micova & De Streel, *supra* note 41, p. 10.

[44] European Parliament, *supra* note 8, on page 12, mentions 'advertising, digital nudging and preferential treatment; paid advertisements or paid placement in a ranking of search results' as novel challenges to be addressed. Algorithm Watch, *supra* note 10, p. 1; European Parliament, *supra* note 8, p. 5.

[45] Algorithm Watch, *supra* note 10, p. 1.

[46] The Crémer report points out several criticalities: (1) not all gatekeepers enjoy a dominant position in the sense of Art. 102 TFEU; (2) the relevant market might be

showing that law, albeit helpful, would most likely not suffice to achieve a satisfactory level of transparency[47].

In light of these interconnected challenges for consumer protection and competition, the strong focus of the European Commission on informational duties as an easily enforceable means to increase transparency and mitigate information asymmetries seems reasonable in principle[48].

However, over time, critics of information duties have continuously added evidence to the list of phenomena hampering the effectiveness of disclosures, which now includes e.g., information overload[49], confirmation bias[50], decision-making aversion[51], the no-reading problem[52], and dislike[53].

Despite this criticism, the Commission reports that 'many' in the consultation process have been calling for more informational duties. In the DMA, these 'many' correspond to civil society and media publishers,

substantially harder to define than in non-digital cases; (3) not every problematic practice has a demonstrable effect on the relevant market. The authors conclude that greater emphasis should be put on the theory of harm, instead. Crémer report, *supra* note 31, pp. 3-4. Moreover, digital markets are often moving at a rapid pace, which is not necessarily a characteristic they share with competition law. Hence, there are concerns whether competition law could be applied with the necessary speed to address urgent competition needs. A. de Streel, *Digital Markets Act – Marking Economic Regulation of Platforms Fit for the Digital Age*, Centre on Regulation in Europe Report, 24 November 2020 [hereinafter CERRE DMA report], p. 59; Recital 5 DMA

[47] Information duties have also increasingly been acknowledged as competition remedies by courts, partly shifting from traditional cease and desist orders towards transparency duties see S. W. Waller, *Access and Information Remedies in High-Tech Antitrust*, in *Journal of Competition Law and Economics*, vol. 8(3), 2012, 575, p. 576.

[48] J. C. Coffee, *Market Failure and the Economic Case for a Mandatory Disclosure System*, in *Virginia Law Review*, vol. 70(4), 1984, pp. 717–753; S.J. Grossman & J.E. Stiglitz, *Information and Competitive Price Systems*, in *The American Economic Review*, vol. 66(2), 1976, pp. 246–253; *S. J. Grossman & J. E. Stiglitz*, *On the Impossibility of Informationally Efficient Markets*, in *The American Economic Review*, vol. 70(3), 1980, pp. 393–408; P. G. Mahoney, *Mandatory Disclosure as a Solution to Agency Problems*, in *The University of Chicago Law Review*, vol. 62(3), 1995, pp. 1047–1112.

[49] H. A. Simon, *A Behavioral Model of Rational Choice*, in *The Quarterly Journal of Economics*, vol. 69(1), 1995, pp. 99–118.

[50] A. Tversky & D. Kahneman, *Judgment under Uncertainty: Heuristics and Biases*, in *Science*, 1, vol. 185, (4157), 1974, p. 1124–1131.

[51] O. Ben-Shahar & C.E. Schneider, *The Failure Of Mandated Disclosure*, in *University of Pennsylvania Law Review*, vol. 159, p. 727, 2011, IIdd (2015) (nt 11).

[52] For an empirical investigation of this issue, see Y. Bakos, F. Marotta-Wurgler & D. R. Trossen, *Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts*, in *The Journal of Legal Studies*, vol. 43(1), 2014, pp. 1–35.

[53] Katz & Zamir, *supra* note 17.

who 'called for an adequate degree of transparency in the market as well as the respect of consumers' autonomy and choice'[54]. In the DSA, the quest for 'algorithmic accountability and transparency audits, especially with regard to how information is prioritized and targeted' online comes from 'a wide category of stakeholders', and is particularly voiced by 'civil society and academics'[55].

Apart from these brief notes, one cannot find more reference to the position of stakeholder groups with regards to transparency duties in the inception IAs. It is therefore relevant to see whether this synthesis duly captured the existing variegated positions. Before moving to our empirical analysis, we will briefly illustrate the actual transparency duties contained in the DSA and DMA proposals. These constitute the formalization of the debate we illustrated above, and we will use it as a blueprint for our empirical research.

2.3. *The Actual Informational Duties in the DSA and the DMA*

The European Commission's vision of what transparency rules might look like, as recently elucidated in the consultation on the DMA and DSA, will be briefly presented in the following. Some of these duties are new, while others are state-of-the-art for many operators. Indeed, especially those enlisted in the DSA are simply restated from the 2019 Platform-to-Business Regulation[56] and the amended Consumer Rights Directive[57].

2.3.1. *DSA: Arts. 12(1), 13, 23-25, 29 and 33*

As summarized in Table 1 in the Appendix, the DSA proposal includes

[54] DMA, at 8 (summarizing the results of stakeholder consultations and impact assessments).

[55] DSA at 9. See also Algorithm Watch, *supra* note 10, p. 1; CERRE DSA report (n 41) 39; European Parliament, *supra* note 8, p. 5; European Commission, White Paper on Artificial Intelligence - A European approach to excellence and trust, COM/2020/65 final, 19.2.2020, 15.

[56] Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services OJ L 186, 11.7.2019, p. 57–79.

[57] Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernization of Union consumer protection rules OJ L 328, 18.12.2019, p. 7–28.

a variety of transparency and disclosure obligations (together: informational duties) for providers of intermediary services[58].

Art 12(1) would entail a general obligation to inform users about potential restrictions to their services contained in the terms and conditions. This information would need to be publicly available, provided in an *easily accessible format*, and written in *clear and unambiguous language*.

Whereas agreeing to the terms and conditions of a platform can be a one-time action, Art 13 DSA would oblige platforms to publish yearly reports about their content moderation practices. These reports would need to be drafted in a *clear and comprehensible language* and include certain specific information[59].

While these obligations would apply to all providers of intermediary services, online platforms would additionally have to provide information about the out-of-court dispute settlements, content suspensions, and the use of automatic tools for content moderation (Art 23 DSA). Concerning the latter, the platform would be obliged to elucidate the '*precise* purposes, indicators of the accuracy of the automated means in fulfilling those purposes and any safeguards applied'. Consequently, it seems fair to expect that the understanding of terms like 'precise' 'clear' 'unambiguous' would be crucial factors in determining the scope and form of the information provided to users[60].

For online platforms displaying advertisements, Art 24 DSA would establish further informational duties. Advertisements and their publishers would have to be identifiable in a 'clear' and 'unambiguous manner'. Furthermore, platforms would have to share 'meaningful information about the main parameters used to determine the recipient to whom the

[58] Above, note 7. In Table 1 (Appendix), we specify whether the norm imposes a transparency or disclosure obligation. Here we use the two as synonyms.

[59] i.e., the number of removal orders received from Member States, categorized by the type of illegal content and the average time required to remove such content; the amount of notice submitted pursuant to Art 14, any action taken thereupon, average time needed for this action, own-initiative, content moderation measures affecting availability, visibility and accessibility of information, and the number of complaints received by the internal complaint system (Art 17 DSA).

[60] For a discussion of the 'clearly, comprehensibly, and unambiguously' requirement in Art 10 e-Commerce Directive, see A. Lodder & A. Murray, *EU Regulation of E-Commerce, Edward Elgar Publishing*, 2017, p. 26. While case law on the matter is rather sparse, the ECJ clarified that information that can only be accessed by a number of clicks is still provided in a clear and comprehensible manner. Bundesverband der Verbraucherzentralen und Verbraucherverbände - Verbraucherzentrale Bundesverband eV v Amazon EU Sàrl, Case C649/17, 10 July 2019, para. 52.

advertisement is displayed' with the platform user. In addition to the obligations laid down in Art 24 DSA, very large online platforms within the meaning of Art 25 DSA[61], would further need to offer application programming interfaces (APIs) to access information on the advertisements they display (Art 30(1), (2) DSA).

Apart from advertisement algorithms, rankings and recommender systems have been identified above as another platform architecture component requiring increased transparency[62]. For very large online platforms this challenge is addressed by Art 29 DSA: in their terms and conditions, very large online platforms would have to flag the use of recommender systems and explain in a 'clear, accessible, and easily comprehensible manner' how these systems work (i.e., which parameters they use and how they can be modified or influenced)[63].

Again, the question of how simple, precise and understandable disclosures are understood seems central regarding the *de facto* effect of these transparency duties.

Lastly, Art 33 sets out comprehensive transparency obligations for very large online platforms[64]. These more pronounced transparency obligations for very large online platforms reflect the differentiated approach the Commission took for the design of the DSA, explicitly mentioned in Recital 39 of the proposal[65].

2.3.2. *DMA: Arts. 5(g) and 6(1)g*

The bottom part of Table 1 clearly shows that transparency duties in the DMA are more scarce than in the DSA and mostly relate to

[61] Per the thresholds chosen by the Commission for the designation of very large online platforms under Art 25(2) DSA and the relation with the different notion of gatekeeper in the DMA see note 3 and 6 above.

[62] Recital 62 DSA.

[63] Moreover, the service recipient would have to be provided with an easily accessible functionality allowing her to select her preferred option for the recommender system the platform is using (Art 27(2) DSA).

[64] Not only do they have to publish reports every six months (instead of yearly), they also have to include a risk assessment (pursuant to Art 26 DSA), risk mitigation measures (pursuant to Art 27 DSA), audit reports (pursuant to Art 28(3) DSA), and audit implementation reports (pursuant to Art 28(4) DSA).

[65] For a thorough discussion of how differentiating rules better ensure the proportionality of regulatory intervention, see F. Di Porto & N. Rangone, *Behavioural Sciences in Practice: Lessons for EU Policymakers*, in A. Alemanno and A. Sibony (eds) *Nudge and the Law*, 2014, pp. 20-59. With reference to transparency duties, Di Porto and Maggiolino, *supra* note 12, pp. 12-22. See also CERRE DSA report, *supra* note 41, p. 11.

rankings and advertising services[66]. They are nonetheless a breakthrough in competition law, because they are ex ante policies envisaged to prevent severe hindrance to market forces from occurring. That justifies the choice to analyze them here.

The main provisions of interest are Arts 5(g) and 6(1)g DMA, especially if read in combination with Recitals 42 and 53. Art 5(g) DMA would oblige gatekeepers, with respect to their core platform services (within the meaning of Art 3(7) DMA), to 'provide advertisers and publishers …, upon their request, with information concerning the price paid by the advertiser and publisher, as well as the amount or remuneration paid to the publisher'[67].

Furthermore, advertisers and publishers can request, and obtain free of charge access to performance measuring tools and the information that is needed to perform their own verification to assess how satisfied they are with the advertisement product they are paying for (Art 6(1)g DMA).

While these obligations are rather specific, Art 10 DMA would open the door to add further transparency duties in the future if a market investigation pursuant to Art 17 DMA identified a need to do so for the sake of safeguarding fair competition.

To sum up, this section has shown that despite the many criticisms, transparency duties loom large in the DSA and DMA proposals. By analyzing in greater detail the actual disclosure duties of the two acts, we provided evidence of the way the Commission seeks to attain a high level of consumer protection and fair competition for digital services.

The analysis shows a stark contrast between what most commentators critique regarding the utility to enact more transparency duties and what the proposals purport. That suggests exploring other and new research routes to understand how these duties were implemented in the DSA and DMA proposals.

[66] Note that we are focusing on general informational duties, not those which only apply if there is an investigation underway (see Art 19 DMA).

[67] This is a self-enforcing obligation for gatekeepers vis-à-vis advertisers and publishers to which they provide advertising services. Gatekeepers should inform about the price paid their counterparts as well as the remuneration paid to the publisher for the publishing of an ad and for the advertising services provider by the same gatekeeper. Such transparency duty, as clarified in Recital 42, is needed for the parties to better understand the real value of the service provided.

3. *A Computational Analysis of the DSA and DMA Consultation Process*

In this section, we ask whether informational duties are what stakeholders asked for in the consultation process and whether their actual wording in the DSA and DMA reflects the way each group uses the relevant terms. This is a relevant step, as it is important that those who implement disclosure duties (typically digital firms, be they small, medium or large) and the beneficiaries of information (individuals, but also micro-organizations) agree on the meaning of the duties (e.g., 'clear', 'accessible', or 'unambiguous language').

To do so, we leverage the power of ML and computational text analysis techniques. In the following, we present our empirical analysis of the replies and position papers submitted by stakeholders to the EU consultation process for three inception IAs. We first give a high-level description of our methodology (for a more detailed description, see Appendix)[68], before presenting our results.

3.1. *Our Methodology*

We collected and analyzed a total of 2,862 replies to the questionnaires and 1,862 of the respective feedback documents attached to the replies[69]. In total, we built a dataset of 3,032,418 words. To do so, we automatically downloaded all the relevant files from the Commission's website[70]. Unlike

[68] The methods we used and describe hereafter largely overlap with those described in F. Di Porto et al., *I see something you don't see. A computational analysis of the DSA and the DMA*, appeared in *Stanford Computational Antitrust*, vol. (1)6, 2021. However, there we focused our analysis on terms related to competition in digital markets and used the theoretical legal framework typical of antitrust law. In this paper, we deploy algorithms on informational duties proposed by the DSA and DMA and use theories of regulation to interpret the results of our computational analysis.

[69] Note that the replies were used partially: we only employed those drafted in English and related with disclosure terms (we manually coded these: see Appendix for further details).

[70] All the documents we used can be found under the following links. As per the DSA proposal: European Commission, Digital Services Act – deepening the internal market and clarifying responsibilities for digital services, 11 January 2021, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12417-Digital-Services-Act-deepening-the-Internal-Market-and-clarifying-responsibilities-for-digital-services (accessed on Jan. 28, 2021) As per what became the DMA proposal: European Commission, Digital Services Act package – ex ante regulatory instrument of very large online platforms acting as gatekeepers, 11 January 2021, https://ec.europa.eu/info/law/better-regulation/

the replies (in excel), most attached submissions came in PDF format, so we first converted them into text and then constructed three large clusters.

3.1.1. *Groups Identification*

To identify groups of stakeholders, we relied on the Commission's categorization scheme for the organization 'size' of the feedback contributors, which groups feedback comments from (1) individuals, micro (10 employees), (2) small (50 employees), (3) medium (250 employees), and (4) large (250 or more) organizations[71]. We then aggregated the different sub-categories (3) and (4) to form three larger categories:

A. individuals and micro firms/organizations;
B. small firms/organizations; and
C. medium and big firms/organizations.

As explained in the previous paragraph, the initial clusters were based on European Commission's 'size' division. From that clustering, we aggregated medium and big firms, as suggested by: (1) the cluster size, and (2) a Kolmogorov-Smirnov test performed on the questionnaires accompanying the consultation (further explained in the Appendix).

Neither the size of companies nor the questionnaire answers we chose to perform the K-S test on were re-used for the Word Embedding Modeling (see below, A.2), hence avoiding double-dipping.

Our decision on *how* to do this aggregation was based on a qualitative and quantitative analysis of the questionnaire accompanying the feedback documents[72].

This allowed us to find out which groups of consultation participants are the most similar and should be clustered together. Note that a Kolmogorov-Smirnov test we performed on the categorical (i.e., multiple-choice) questions in the questionnaire showed that 'medium and large'

have-your-say/initiatives/12418-Digital-Services-Act-package-ex-ante-regulatory-instrument-of-very-large-online-platforms-acting-as-gatekeepers; and European Commission, Single Market – new complementary tool to strengthen competition enforcement, 11 January 2021, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12416-New-competition-tool.

[71] The Commission distinguishes the feedback also by 'types' of contributors. E.g respondents to the DSA were: the general public (66%), companies/businesses organizations (7.4%), business associations (6%), and NGOs (5.6%) authorities (2.2%), academic/research institutions (1.2%), trade unions (0.9%), and consumer/environmental organizations (0.4%) (see DSA at 8).

[72] See European Commission (n 57) for the questionnaire. A detailed description of how we analyzed the questionnaire can be found in the Appendix.

entities should be grouped together as they can be assumed to be one cluster[73]. This is per se a relevant finding, because although different in size, and despite the fact that in most economic surveys they are considered separately, medium and large entities are a cluster for the purpose of text analysis. That is justified by both qualitative and quantitative factors.

First, our algorithm assessed replies provided by firms and organizations together, while in economic surveys just firms are grouped in one cluster. It is therefore possible that the presence of organizations attenuated the distance in the use of terms.

Second, that is extremely relevant because even if medium and large entities decide through different mechanisms (e.g., taking a decision may involve only one manager in medium organizations, while requiring dozens in big ones), what we assess is the way they understand and use terms related to transparency duties. Hence, the size of organizations is not a relevant parameter, as it is semantics. Third, by analyzing the text of organizations' opinions, as formalized in the feedback documents and replies, and later encapsulated in the DMA and DSA informational rules, we are able to capture how medium and large entities make use of terms related to transparency.

3.1.2. *World Embedding Modelling: Training the Algorithm*

After having identified the most sensible way to cluster the consultation documents, we built three corpora:

- 744 documents with 35,949 unique words for corpus A (*Individuals and micro enterprises and organizations*),
- 393 documents with 32,100 unique words for corpus B *(small companies/organizations)*,
- and 689 documents with 39,815 unique words for corpus C *(medium and large companies/organizations)*.

We always compared two corpora, hence we analyzed three corpus pairs (A-B, B-C, A-C).

By constructing three different corpora, we were able to train a neural network on the documents of each cluster, hence having three networks that capture the intricacies of each corpus. Based on the number of times

[73] This choice can not only be backed by our data, but also by some scholarly findings, e.g., R. Kemp & C. Lutz, *Perceived barriers to entry: Are there any differences between small, medium-sized and large companies*, in *International Journal of Entrepreneurship and Small Business*, vol. 3(5), 2006, pp. 538–553. For a more detailed description as to why we cumulated medium and large entities, instead of clustering medium with small ones, see the Appendix.

words occur next to each other, this network allowed us to calculate a vector for each word in each corpus, a so-called Word Embedding Model (more specifically, we used Gensim's CBOW word2vec model)[74]. These models are remarkable in the sense that they can capture the semantic meaning of words in a set of numbers. For instance, in a well-trained model, the distance between the vector of the words 'Paris' and 'France' will be roughly the same as between 'Rome' and 'Italy'. Hence, the relative positions of vectors in the model approximately represent the meaning of certain terms. This means that while a simple algorithm would require researchers to formulate explicit rules to approximate the semantic meanings of words, ML (or the neural network, to be precise) learns the implicit rules directly from the data we feed it. This does not only increase the performance of the algorithm but also prevents an undue influence of the researchers' conscious or subconscious assumptions[75].

3.1.3. *Making sense of semantic distance*

However, it needs to be noted that models trained on different corpora are not directly comparable. Since the vectors making up the models are based on the frequency of words occurring next to each other, they depend on the corpus the model was trained on. Hence, even the position of words that most definitely have the same meaning for all groups (e.g., 'and') will have very different vectors, which we would normally interpret as a semantic difference. In this case, however, the distance between the two vectors will not be the result of a different use of a word, but simply the particularities of the corpuses the model was trained on. Consequently, to make the models we trained on the different corpuses comparable, we used unsupervised vector space alignment. This allowed us to bring the vectors trained on two different corpuses together in one model space, where they would be comparable. Put differently, in the aligned model space, strongly differing vectors represent actual differences in the use of a word, instead of being a result of a different training basis.

However, we still needed to ascertain that these differences were not merely incidental, but actually of a certain significance. To do so, we

[74] T. Mikolov, K. Chen, G. Corrado & J. Dean, *Efficient Estimation of Word Representations in Vector Space*, 2013, ArXiv:1301.3781 [Cs]. R. Řehůřek, *Word2vec embeddings*, 2019, <https://radimrehurek.com/gensim/models/word2vec.html> (accessed on June 22, 2021).

[75] For instance, a researcher might assume that a word needs to be used in the same sentence at least x times for the two to be related and design her algorithm accordingly. For our algorithm, we do not need these kinds of assumptions or rules as the algorithm learns directly from the data.

employed a statistical test. This test relies on the assumption that the distance between the vectors for the same word from two different corpora can be split into three components: a semantic difference (i.e., a difference in meaning), a non-semantic difference (e.g., syntactical differences), and a random difference. We then set two assumptions: first, we assume that the semantic difference between corpora for a certain set of words (the control vocabulary) is zero. This means that we assume all stakeholder groups use words like 'and' or 'one' in the same way. Based on this, we were able to construct an empirical distribution of the non-semantic difference and the random difference, assuming that there is no semantic difference. This distribution is our second assumption.

Knowing how our vectors should look like if there was no semantic difference between the clusters, we were then able to check for each word if the distance between its vectors from two different corpora is compatible with this hypothesis of a uniform use. If it is not, we can conclude with a certain level of confidence that there is a statistically significant difference in its semantic meaning between the different corpora.

With these tools at hand, we analyzed the stakeholder submissions to the DSA and DMA consultation process. Given that the stakeholders whose opinions we analyze are to a large extent those who will either draft or receive the abundant transparency statements envisioned in the proposals[76],76 their uses and view of terms related to informational duties should be of great interest both for legislators and scholars debating the factual role of informational obligations.

The questionnaires raise several points, not all of which immediately related to informational duties. For instance, the NCT questionnaire also discusses competition problems (such as agreements, self-preferencing, or collusion); while the DSA one includes questions on liability of intermediaries.

Because we are interested in the use of certain terms only, we created an initial list of 119 terms, based on the glossaries of the consultation questionnaires which explain terms that might be new to some consultation participants. However, after the first analysis, we realized that our list of terms might be too narrow for two reasons.

First, the wording of the Inception Impact Assessments (IIAs) which were discussed in the consultations differs from the final draft DSA and

[76] This includes the general public, authorities and consumer/environmental organizations (as addressees), and companies/businesses organizations, business associations, and trade unions (as drafters); but will exclude NGOs, and individual academics and research institutions.

DMA. The change in vocabulary is especially marked in the DMA[77], where classic concepts of competition law (such as market, dominance, efficiency gains) are mostly abandoned, and new ones are defined[78]. Since we used corpora from comments to the three IIAs documents to run our analysis and needed it to reflect this change, we proceeded with hand-coding. Therefore, we combined words from two sources: (i) all glossaries[79] attached to previous legislation (all EU Directives and Regulations) that were recalled by the DSA and DMA proposals (for a total of 119 words); and (ii) terms related to transparency (e.g. 'disclos*', 'transparency', 'inform*' and the like) that were manually selected from the questionnaires (102 words). As a result, we ended up with a list of 194 words (102 from the DSA's questionnaire and 92 from the DMA's). (See Annex 3.1).

Furthermore, since we are interested in the specific provisions of the DSA and DMA which qualify how information should be provided (e.g., 'clear', 'accessible'), we added all those terms from the proposals' informational provisions (ten terms in total, see Annex 3.1).

Finally, stakeholders use a variety of terms to refer to the same concept. For instance, our list might include 'self-preferencing', but we would miss differences on 'self-favoring'. Our pre-defined list of terms was not able to capture this variety. Since it was also not feasible to anticipate all these variations, we chose to manually code those results that are closely related to the terms and concepts of our list *ex post*.

To perform manual coding, we relied on the legal expertise of our team, with the aid of external assistance[80]. Finally, the terms that were added manually were a total of 204, while overall the computational analysis was performed over of a total of 323 words.

[77] The difference in terminology also derives from the fact that the 'NCT' inception IA was based on Art 106 TFEU (much focused on competition), while the 'Ex-ante regulation's legal base was Art 114 TFEU (internal market). Following the consultation, the DMA proposal had its own legal base (Art 114) and terminology.

[78] As are spheres of application of the DMA in comparison with the inception IAs.

[79] Glossaries are definitions of terms usually contained in Arts. 2 of EU Directives and Regulations. Namely, we added all the glossaries from: the GDPR, the NIS Directive, the Data Governance Act proposal, the E-commerce Directive and the Platform-2-Business directive.

[80] We are thankful to Andrea Ruffo, legal scholar and teaching assistant at Luiss University of Rome for his wonderful assistance in the manual coding activities. The legal analysis was performed by Tatjana Grote and Fabiana Di Porto.

3.2. *Results: Different Groups, Different Users?*

We found a statistically significant difference for

1,865 word pairs between corpora A and C,

2,184 between corpora A and B and

1,113 between B and C[81].

A detailed description of how this comparison was conducted and what 'significant' means in this context, is provided for in the Appendix (Annex 3). From all the 5,162 significant distances we found, we chose those that were relevant to our analysis, based on the selection procedure described above. This resulted in a list of 13 relevant terms for which we found significant differences in use and understanding.

[81] It needs to be noted that many of these words are not of particular interest for us because they might identify a specific service of a certain company (e.g., the 'Gmail' email service in Google's submissions). However, some of the key buzzwords surrounding competition and transparency obligations show statistically significant differences.

Table 1: Summary of results

Term	Distance AB	Distance BC	Distance AC	Close words A	Close words B	Close words C
Consumer-centric	1.557 (0.03)**	1.625 (0.02)**	1.247 (0.16)	privacy-protecting	systems	computing
Easy	1.444 (0.04)**	1.443 (0.07)*	1.451 (0.05)*			
Easy-to-use	1.450 (0.04)**	1.427 (0.08)*	1.522 (0.02)**	deregulation		cut-off
Meaningful	0.545 (0.627)	0.670 (0.648)	1.482 (0.04)**			
Precise	1.645 (0.01)**	0.878 (0.434)	0.747 (0.497)	cartel	checklist	
Privacy-friendly			1.468 (0.04)**	misconceptions		tailor-made
Ranking	1.182 (0.15)	1.644 (0.02)**	1.452 (0.05)*		guidelines, improve, oversight	appearance, disclosing
Readable	1.051 (0.237)	1.720 (0.01)**	1.394 (0.08)*		effective, specific, clear	entities
Self-regulatory	1.340 (0.09)*	1.536 (0.04)**	0.897 (0.37)		blacklisting, sanctions, obligations	benchmarking, codes, ameliorate
Simple	1.703 (0.01)**	1.504 (0.05)*	1.158 (0.20)	formats	precise	
Understandability		1.663 (0.02)**			single-homing, practice	informs
Unregulated	1.361 (0.07)*	1.566 (0.04)**	1.822 (0.00)***	not-sufficient		mitigation
Well-informed	0.943 (0.293)	1.734 (0.01)**	1.749 (0.00)***	Confusing, explainable		Inscrutability, implementation

Note: The asterisks indicate significance at a 0.001 (***), 0.05 (**), and 0.1 (*) level, respectively.

Table 1 shows these results. The 'Distance' columns report the distance between the vectors of the same words for each corpus pair, with the respective p-value in parentheses. A grey field in the 'Distance' columns indicates that a word was not used in both of the respective corpora.

The 'Close Words' columns shine a light on some of the concepts that were closely related with the term in question in the corpora for which there was a statistically significant distance between the terms. To be precise, we computed the ten words which were most similar to the term in question[82] and then hand-coded those words which were relevant to our analysis,

[82] Our similarity measure is the cosine distance between two vectors. R. Řehůřek, *Gensim: Store and query word vectors – Similarity*, 2019, https://radimrehurek.com/gensim/models/keyedvectors.html#gensim. models.keyedvectors.WordEmbeddingsKeyedVectors.similarity (accessed on Aug. 30, 2020).

based on the same procedure outlined above (see the last paragraph of 3.1.2). A grey field in the 'Close words' columns means that we did not look for close words because the respective corpus was not involved in any of the significant distances or there were no meaningful close words.

Moving on to our results, we start with some terms that are of importance on a meta-level, namely those related to the overall regulatory strategy employed. Since there are different regulatory paths to ensuring transparency (e.g. by regulation or self-regulation), this is of interest as well.[83]

3.2.1. *Words related to the regulatory 'meta-level'*

We observe that 'self-regulatory' is used differently by different stakeholders. Generally, we see that self-regulation seems to be a more prominent issue for medium and big companies (corpus C): while the term is only mentioned ca. 5,000 times by small companies (corpus B, with 810, 961)[84], it occurs more than 25,000 times in corpus C (which contains 1,177,120), where it is associated with the terms 'benchmarking', 'codes', and 'ameliorate'. This is reflected in Fig. 1, and could be read as a sign that self-regulation is seen as an important strategy by medium and big companies/organizations.

Differences in use also exist for the term 'unregulated'. For individuals (A) and small entities (B), an 'unregulated' digital single market does not seem like a favorable option, with 'not-sufficient' and 'precariousness' as closely related terms. (Fig. 2).

3.2.2 *Words related to informational duties*

With regards to *informational duties*, it is interesting to note that there is a statistically significant distance between the use of the word 'simple' between corpus A and B (Fig. 2). While individuals and micro-businesses/organizations seem to focus on 'formats' regarding simplicity, small companies/organizations in our dataset associate the attribute 'precise'. However, it needs to be noted that the term 'precise' also underlies some significant differences between corpora A and B, which is an important finding in light of the wording of Art. 23 DSA (Table 1).

Generally, individuals and micro-organizations (A) used the word 'simple'

[83] For a detailed discussion of regulatory strategies in disclosure regulation, see Di Porto & Zuppetta, *supra* note 14.

[84] Note that the corpus sizes indicated here refer to the overall corpus, i.e., the number of words in the documents as they were submitted. For corpus sizes indicated above we only considered the unique words for each corpus, which is why these numbers are much smaller.

roughly 20-times more often than small businesses and organizations (B).

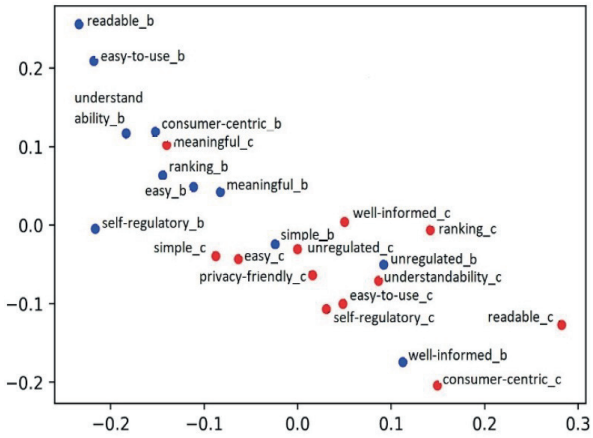


Figure 1: Aligned Vector Space Model - Corpora B & C

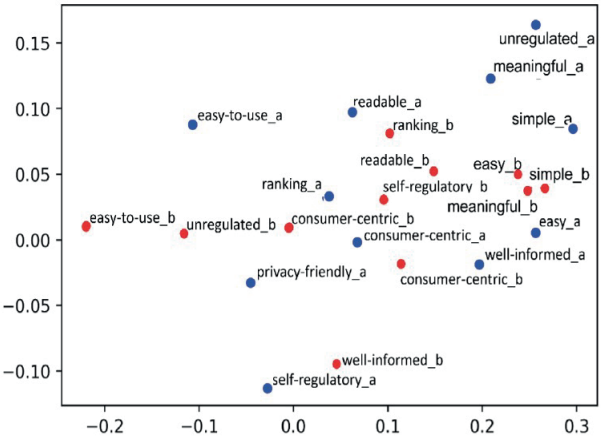


Figure 2: Aligned Vector Space Model - Corpora A & B

With regards to the obligation of advertisement system transparency laid down in Art 24 DSA, it is surprising to see that 'meaningful' is used very differently by individuals and micro-organizations/businesses (A) than by medium and big companies (C).[85] Again, this could potentially impact

[85] The use of 'Meaningful' for the corpus pair C and A might look close in Fig. 3 because the difference is not as pronounced as for some other terms, but it has p-value of 0.04, meaning that we can conclude there is a statistically significant difference.

the efficacy of said provision since what is deemed 'meaningful' by the drafters of the respective disclosures might be rather meaningless for their recipients.

In the comparison between corpora A and C, the term 'well-informed' is mentioned roughly 26,000 times by individuals and micro-contributors (A; in total 1,044,337 words) compared to 18,642 mentions in corpus C (in total 1,177,120 words) and is closely related to 'explainable'. Furthermore, we find a different utilization of the terms 'easy-to-use' and 'privacy-friendly', respectively (see Fig. 3).

The first is interesting with a view to rules like Art 17(2) DSA, which speaks of *easy to access, user-friendly* complaint mechanisms. The latter seems to be located within slightly different contexts by different stakeholder groups: while individuals (A) heed possible 'misconceptions', medium and large companies/organizations (C) associate 'privacy-friendly' with 'tailor-made' and 'reinforced'. Interestingly, the Commission explicitly mentions that 'privacy-friendly services' wereone key expected outcome of the DMA in the eyes of the consultation respondents. However, what might be missing is that not all stakeholders understand the same when speaking of 'privacy-friendly'.

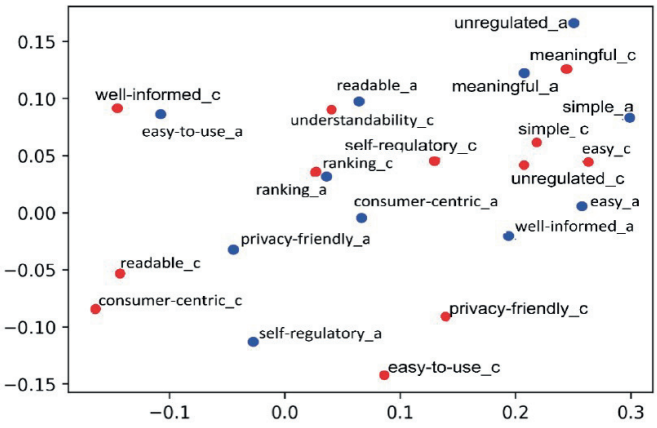


Figure 3: Aligned Vector Space Model - Corpuses A & C

Comparing small companies/organizations (B) and medium/big companies/organizations (C), we find a significant distance between the vectors for the terms 'well-informed' and 'consumer-centric' (Fig. 1, above). The latter word is closely related to the term 'systems' in corpus B, which is unsurprising. In corpus C, we see a close association with 'computing',

which is interesting since it seems to shift the focus of consumer-centric design to the processes happening behind the systems that consumers interact with.

Another intricate finding concerns the term 'ranking', which has been central in discussions about the transparency of online platforms. This close connection between transparency and rankings is also reflected in the close words we found: small companies (B) associate rankings with '*guidelines*', medium/big companies with *'disclosing'*.

As 'ranking' is not a crucial term for transparency duties as such, this difference will not necessarily impede the effectiveness of disclosures. Nevertheless, this finding shows that there are different perceptions of some key concepts of the DSA and DMA across stakeholders.

We further find differences for the terms 'understandability' and 'readable'. This should be a key concern for policymakers and legal scholars when debating transparency duties: if no uniform understanding of what 'readable' transparency disclosures look like can be reached, consumers will likely have to deal with strongly differing levels of readability and understandability.

3.3 *Challenges*

Our algorithmic analysis of the consultation process for the DSA and DMA has shown that there are statistically significant differences between stakeholders' use and understandings of some key concepts of transparency. To the best of our knowledge, we are the first to conduct such a 'close reading' of an EU rulemaking process and discern differences in the ways a consultation relates to the rules in the context of the DSA and DMA. Our results show that NLP techniques can allow the Commission to understand not only what stakeholders say, but what they actually mean; which could substantially improve stakeholder consultations' analysis as we did here. For instance, the Commission took note of demands for more 'simple' notice-and-action procedures for content removal.[86] Yet, we discovered that the term 'simple' might not be understood in the same way across all groups of stakeholders. This could offer a first signal to the Commission that it is premature to legislate on this matter; or that a one-size-fits-all measure may not be suitable.

Linking our results back to the discussion of transparency duties and

[86] DSA proposal, 8.

their importance for consumer protection in digital markets, our findings cast doubt on whether all stakeholders have a similar understanding and thus make similar uses of simple, meaningful, easy-to-understand, readable transparency statements. Given that the exact implementation of such duties often lies in the hands of different stakeholders, this might be one reason why transparency duties remain ineffective. For instance, our algorithm reveals that 'meaningful' is understood and used differently by the individual consumers and the medium/big platforms. This may cause Art. 24 DSA failure, as it obliges platforms to inform consumers in real-time that what is being displayed to them is an ad, in a clear and '*unambiguous* manner'. Since the literature on the failure of disclosure regulation has mostly focused on how transparency statements are perceived by consumers[87], our focus on all stakeholders, inclusive both the recipients and drafters of disclosure statements, adds a unique, novel perspective.

Having said that, there are challenges that need to be addressed, some of which are common to the computational law scholarship[88], others are specific to our analysis. Both offer room for improvement by future research[89].

Concerning the analysis, in the methodology, we make two assumptions for the statistical test we perform: that words in the control vocabulary used for the vector space alignment transformation do not have a semantic difference and that the distribution of distances has the same shape also for the other words. For instance, we assume that words like 'and' or 'one' are understood in the same way by all contributors in the consultation. While this seems plausible, we cannot entirely discard the possibility of errors in the creation of the models and their alignment due to shortcomings in these assumptions. Nonetheless, our assumptions are commonly accepted in the literature.[90]

Second, our corpora are relatively small and heterogeneous since they contain documents from many different authors with potentially different styles and focuses. For instance, feedback we analyzed are in English language only, but their authors might not be native English speakers. This could introduce a bias, meaning that results may be partially driven

[87] Above (n 11).
[88] D. Lim, *Can Computational Antitrust Succeed? Stanford Computational Antitrust*, https://law.stanford.edu/wp-content/uploads/2021/04/lim-computational-antitrust-project.pdf (accessed on June 22, 2021), pp. 10-13.
[89] More technical limitations are presented in the Appendix.
[90] See Nyarko & Sanga, *supra* note 25, p. 4.

by the particularities of our corpora. Hence, increasing the corpus size and the control vocabulary should be a top priority for future research. Another way to solve the problem would be using bootstrapping: by repeatedly and randomly changing some words in the corpora and then taking the mean value, the random term u_t^{AB} in the distribution of distances could be reduced.

Generally, it needs to be noted that our analysis focuses on the *identification* of semantically different terms. At this stage, we do not seek to provide insights into what the identified differences might be based on and how they impact the stakeholders' opinions. Therefore, it has some limitations as far as *interpretation* is concerned. Using word embedding alignment alone does not allow (yet) to show any causal relationship between differences in perceptions of transparency and specific factors. Although we compared the most similar vectors[91] corresponding to the word pairs of interest, gaining an idea of how the meanings might differ, this still requires a certain degree of *ad hoc* interpretation. Moreover, we used ex post manual coding when selecting the results to be presented here. In the future, fully replicable, *ex ante* criteria should be used to make this selection.

Due to these limitations, our results need to be treated with caution and should be complemented by further research. Nevertheless, they constitute a first step providing interesting insights into informational duties in the DMA and DSA.

4. *Concluding Remarks*

This paper sets out to explore whether different stakeholders participating in the consultation process for the latest Commission proposals on new rules for digital markets (the DSA and DMA) share a similar understanding of key concepts related to one integral pillar of the new proposals: informational duties. We analyzed the replies to questionnaires and feedback documents submitted in the consultation process using the NLP technique of Word Embedding Alignment, which allowed us to identify terms that are not used in the same way by all stakeholders.

We find significant differences in the way stakeholders use words that

[91] See note 82 above.

are central in transparency duties, like 'readable', 'simple', and 'privacy-friendly'. These differences are group-specific, and hold for individuals and micro organizations; small; and medium/large organizations. If that might seem obvious at first sight, it is surprising if one considers that those participating in the consultation process on the DSA and DMA constitute a rather small epistemic community, made of legal and economic scholars, digital companies, NGOs, and IP specialists who have a high stake interest in expressing their voice and are, therefore, well-informed about the subject they discuss.

Our results should be a key concern for policymakers and legal scholars for several reasons. Differences in understanding might mean (undesirable) differences in implementation. If there is no uniform use (and understanding) of what 'readable' transparency disclosures or 'simple' complaint mechanisms look like, users will likely have to deal with strongly differing levels of readability and simplicity.

Second, this could decrease the effectiveness of transparency duties in ensuring competitive and fair markets, given that those who replied to the consultation are also those who will draft and receive the disclosures.

Third, and strictly related, different understanding and uses of words that are relevant to informational duties might also help explain why such rules fail.

The last takeaway we want to stress is that rule-makers are recommended to consider another interesting finding: that understanding and use of relevant terms of transparency (like 'simple' and 'well-informed') do not differ between medium and big organizations (corpus C), as one would expect. That is to the point to make them a sole group for the sake of text analysis. Generally, if the Commission used tools like the one applied here to complement its impact assessments and rulemaking, it could not only hear what stakeholders *say* but understand what they *mean*, which might ultimately improve the functioning of the EU's new regulatory traffic lights for digital markets.

Looking at the perspectives this paper opens, we think that our analysis, if complemented with other computational techniques, will bevery useful in doctrinal studies of the future.

One scenario could be to investigate the 'rationale' of the DSA and DMA's rules. By the time the DSA and DMA will entry into force, their wording will change several times, depending on multiple interactions of the Commission, the Parliament, Council and stakeholders. Our analysis might be a first step in the direction of keeping records of textual

modifications and then tracing back the statements that influenced them the most (e.g., being the most similar). Clearly, our analysis alone would not be enough and would need to be complemented with other NLP techniques. For example, text similarity techniques could be employed to map out which stakeholder opinions might have influenced the EU institutions when drafting not only its proposals but also its final rules. This might allow gaining a precise understanding of why rules were drafted in a certain way and could greatly help the interpretation of rules in light of their *telos* and their drafting history.

A second research area that our analysis could inaugurate is that of improving the drafting of disclosure statements and transparency reports, as envisaged by the two new proposals. While we considered the use and understanding of information-related terms by firms and organizations together, one could zoom in on the use of concepts by individual consumers and firms only, which will certainly differ. For instance, the phrase 'easy to use' was used differently by all three clusters. If we already find this disagreement in large, aggregated groups, the understanding of such a phrase will most likely differ between individuals. Consequently, regulators might opt for clusterized disclosures, with messages adapted to the specific informational capabilities of users' groups (as identified by our computational analysis).

That might help to overcome many of the shortcomings of current disclosure statements. While this possibility was discussed in great detail elsewhere[92], our analysis suggests that the Commission and platforms would be well-advised to explore this possibility.

Our algorithm should be seen as the first building block of a fully-fledged tool for a more in-depth algorithmic analysis of EU rulemaking. The other building blocks might be:

'topic modeling'[93], which would allow rule-makers like the Commission and scholars to get an intuitive understanding of how the most important topics, that will become rules in a near future, are part of a shared view among different stakeholders or whether they emphasize different issues;

[92] See, e.g., F Di Porto, Algorithmic Disclosure Rules, in Artificial Intelligence and Law, (2020), https://ssrn.com/abstract=3705967 or http://dx.doi.org/10.2139/ssrn.3705967 (accessed on Oct. 27, 2021). More information on the implementation of clusterized disclosures is available at: www.lawandtechnology.it. See also: C. Busch, *Implementing Personalized Law: Personalized Disclosures in Consumer Law and Data Privacy Law*, in *The University of Chicago L. Rev.*, vol. 86(2), 2019, 309–332.

[93] D. M. Blei, A. Y. Ng & M. I. Jordan, *Latent dirichlet allocation*, in *The J. Machine Learning Research*, 2003, 3, 993–1022.

'document similarity'[94] could be used to cluster statements that are input to regulation before the Commission publishes a regulatory proposal. This could help to perceive certain similarities or alliances, between stakeholders, even across different groups like e.g., small companies and medium/large companies.

- Sentiment Analysis could be another means to understand if the parties to a rulemaking process agree or disagree with certain proposals or statements. In fact, we performed a first explorative sentiment analysis using a pre-trained model on those paragraphs in our documents which contain the terms of interest presented above (Table 1). While this analysis produced some interesting results[95], a fully-developed sentiment analysis is best left for future research. Furthermore, one could cluster each statement based on the overall sentiment of a group of contributors[96] to get a better understanding of how supporters and critics of a proposal are distributed and what their main concerns and arguments are.

Overall, while we believe that discerning latent differences in the use of certain terms is a crucial capability that could significantly enhance the consultation process at the EU level, the above-mentioned additions could be combined in a fully-fledged NLP toolbox that could substantially enrich the work of both the Commission and legal scholars and provide many new insights.

Be that as it may, it is hoped that our findings will enrich the positive and normative debate about transparency rules in digital markets, inspire

[94] See, e.g., B. K. Triwijoyo & K. Kartarina, *Analysis of Document Clustering based on Cosine Similarity and K-Main Algorithms*, in *Journal of Information Systems and Informatics*, 1(2), 2019, pp. 164–177. D. Gannemann, *Comparative Law: Study of Similarities or Differences?*, in M. Reimann and R. Zimmermann (eds.), *Oxford Handbook of Comparative Law* (2d ed.), Oxford University Press, 2019.

[95] For instance, we found that 'understandability' is seen much more favorably by small companies/organizations (B; 0.721) than by medium/big entities (C; 0.340). Similarly, we found a more positive attitude towards the terms 'well-informed' and 'consumer-centric' for individual and micro contributors (0.624) than for small companies/organizations (0.051). We also identified a negative sentiment of small companies/organizations towards the term 'unregulated' (-0.118). Lastly, 'simple' is viewed more favorably by individuals and micro contributors (A; 0.314) than by big and medium organizations/businesses (C; 0.220).

[96] See e.g., S. Feng, D. Wang, G. Yu, C. Yang & N. Yang, *Sentiment Clustering: A Novel Method to Explore in the Blogosphere*, in Q. Li, L. Feng, J. Pei, S. X. Wang, X. Zhou, & Q. M. Zhu (Eds.), *Advances in Data and Web Management*, 2009, pp. 332–344.

future research in the computational antitrust arena, and urge EU rule-makers to rethink their convictions about the use of computational tools in the consultations.

Addendum

Corrigendum - The authors also published a paper using the same dataset and methodology in Fabiana Di Porto, Tatjana Grote, Gabriele Volpi & Riccardo Invernizzi, "I see something you don't see": A computational analysis of the Digital Services Act and the Digital Markets Act", 2021 *Stanford Computational Antitrust journal, #5* https://law. stanford.edu/wp-content/uploads/2021/08/di-porto-computational-antitrust.pdf.

Appendix

T / D duty	Digital Services Act (DSA)	Recipient of info (r) Info to be provided (i)	'How' to disclose	Core service providers (Art 2(f) DSA)	Online platforms (Art 2(h) DSA)	Very Large online platforms (Art 25)
D	Terms of service include information on content moderation and use of algorithms	(r) Users; (i) potential restrictions to their services.	'easily accessible format' written in 'clear unambiguous language'	Art 12 (Terms and conditions)		
T	Yearly reports on content moderation providing key information specified in Art 13(1) DSA	(r) Users and the general public; (i) content moderation practices	written in 'clear and comprehensible language'; need to include specific information (a. 14, 17)	Art 13 (Transparency reporting obligations for providers of intermediary services)		
D	Reasons for removing the content or disabling access	(r) Users whose content was removed or access disabled	Clear and specific statement containing the information listed in Art 15(2)		Art 15 (Statement of reasons)	
T	Additional information (with reference to Art. 13) on content suspension actions taken, use of automated means for content moderation, and out-of-court dispute settlement	(r) Users and the general public, (i) esp. about automation of content moderation and ADR	Format potentially to be specified by Commission, Art 23(4)		Art 23 (Transparency reporting obligations for providers of online platforms)	
T/D	Advertising transparency duties	(r) Users and recipients of service; (i) display that info is an ad + personalization of ad	Provided in a 'clear and unambiguous manner'		Art 24 (Online advertising transparency)	
D	Main parameters used in recommender systems must be set out in terms and conditions	(r) Users; (i) use of algorithms for recommending content	Provided in a clear, accessible, and easily comprehensible manner	/	Art 29 (Recommender Systems)	
T	Additional advertisement transparency duties to maintain in the repository and made accessible	(r) Users and the general public; (i) advertisements and their display	Repository be made publicly available through an API		Art 30 (Additional online advertising transparency)	
T	Additional information on content moderation, risk management, and auditing	(r) Users, the general public, and Digital Service Coordinator; (i) results of risk assessments and audits	-		Art 33 (Transparency reporting obligations)	
	Digital Markets Act (DMA)	Recipient of info	'How' to disclose	Gatekeepers (as defined in Art 3 DMA)		
D	Information about advertising services provided by gatekeepers for advertisers and publishers	(r) Advertisers and publishers counter-parts	-	Art 5(g) (Obligations for gatekeepers)		
D	Provide free of charge access to performance measuring tools of gatekeepers and information necessary to enable advertisers to carry our independent verification	(r) Advertisers and publishers	-	Art 6(g) (Obligations for gatekeepers susceptible of being further specified)		

Note: Informational duties (Column 1) may include either transparency duties (T) or disclosure duties (D).

Table 1 Informational duties in the DMA and DSA

Annex 1 Groups identification

To analyze the replies to questionnaires and feedback documents, we created a special scraper algorithm, which allowed us to download all the files automatically, convert them into text, and split them into

three clusters. In doing this, we started by following the Commission's categorization scheme for the organization size of the feedback contributors. We then aggregated the different sub-categories into three corpora based on the typology and the dimension of the feedback contributor: Corpus A (individuals and micro organizations), B (small companies/organizations), and C (medium and large companies/organizations).

Our clustering choice is based on two considerations: First, a qualitative analysis of the questionnaires accompanying the feedback documents[97] allowed us to get an understanding of which aggregation would cluster comparable feedback contributors together. We mostly analyzed the types of feedback contributors in the sample and had a look at their replies to questions related to informational duties. Second, we conducted a quantitative analysis of the same questionnaires to ensure that our clusterization choices are solid. In particular, we sought to ensure that there is no statistically significant difference between medium and large entities in our sample since at least medium companies are often grouped with small, rather than large companies[98]. However, it needs to be noted that our feedback contributors are not only businesses but also other types of organizations. This diversity could "smooth" the differences we would have expected to find if our sample included companies only. In fact, our qualitative analysis of the questionnaires suggested that medium entities in our sample are more comparable to large businesses/ organizations both in terms of entity type (whether they are from academia, civil society, private economy, etc.) and in terms of how they perceive challenges arising from digital markets (in the sense that they gave more similar answers to the pertinent multiple-choice questions in the questionnaires)[99]. To test the robustness of this perception, we analyzed the answers provided for by medium and large entities to specific multiple choices questions[100]. We

[97] European Commission, Digital Services Act – deepening the internal market and clarifying responsibilities for digital services, 11 January 2021, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12418-Digital-Services-Act-package-ex-ante-regulatory-instrument-of-very-large-online-platforms-acting-as-gatekeepers/public-consultation (accessed on Jan. 28, 2021).

[98] Statistically significant refers to the hypothesis of the K-S test, that the data of both groups is originating from the same population.

[99] While this could be due to the idiosyncrasy of our sample, this finding also corresponds with scholarly literature. See e.g., R. Kemp & C. Lutz, *Perceived barriers to entry: Are there any differences between small, medium-sized and large companies*, in *International Journal of Entrepreneurship and Small Business*, vol. 3(5), 2006, pp. 538–553.

[100] The questions were selected manually based on two criteria: First, we manually identified all questions relating to informational duties and competition in digital markets. In a second

applied a Kolmogorov-Smirnov two-sample test[101] to understand if there is a statistically significant discrepancy between the distribution of the answers of the two groups. If that was the case, we would assume that these answers must be considered as provided by two different populations, not allowing us to treat them as a unique cluster. The results of the test are shown in Figure 1.

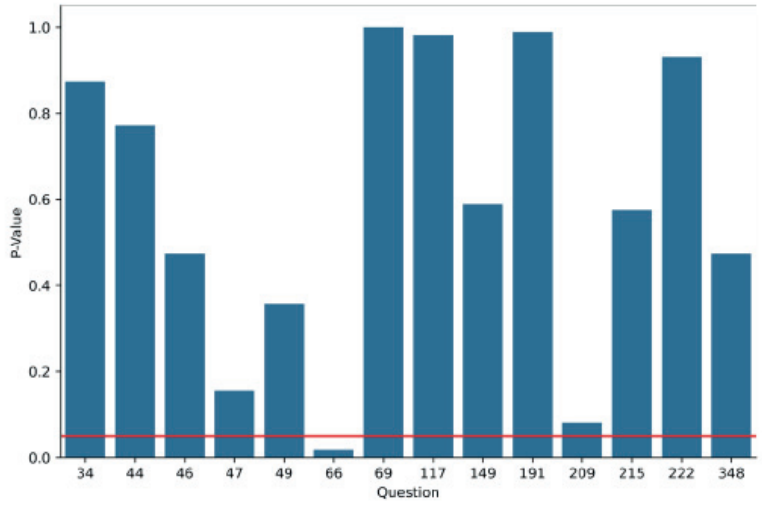


Figure 1: p-values resulting from KS-two sample test applied to the answer distributions of the considered questions. Red line highlights our significative tolerance value of p=0.05

Even using a very high tolerance p-value level of 0.05, only question no. 66 showed a statistically significant variation. This question alone however is mostly unrelated to our core research interest, and hence unlikely to compromise the validity of our clustering.

In total, we collected 744 documents with 35.949 words for corpus A, 393 documents with 32.100 words for corpus B, and 689 documents with 39.815 words for corpus C. We always compared two corpora, hence we analyzed three corpus pairs (A-B, B-C, A-C).

step, we singled out questions that had a categorical answer scale, i.e., non-text replies.

[101] L. Hoboes Jr., *The significance probability of the Smirnov two-sample test*, in *Matematica*, vol. 3(5), 1958, 469-486.

Annex 2 Training the algorithm

To discern differences in the use of certain key terms across stakeholder groups (i.e., a different semantic understanding of identical terms), we leveraged Word Embedding Models to quantify evidence of such differing understandings. This technique has already been used in various Natural Language Processing tasks, and recently also in the Computational Law literature[102]. It has been demonstrated to be very powerful and useful in providing insights into latent differences in how language is used.

The core of this technique consists in training a special neural network to convert each word contained in a corpus of texts into a vector, i.e., a set of numbers[103]. While a simple algorithm would require researchers to formulate explicit rules to somehow approximate the semantic meanings of words, ML (or the neural network, to be precise) learns the implicit rules directly from the data we feed it. This does not only increase the performance of the algorithm but alsoprevents an undue influence of the researchers' conscious or subconscious assumptions. The resulting vectors are based on the frequency of words occurring next to each other, meaning their relative positions in each phrase of the corpus and the correlation between words. The stronger two words are correlated (in their occurrence – and so in their semantic meaning)[104] in the corpus the model was trained in, the closer the corresponding vectors will be located to each other.

[102] See e.g., Nyarko and Sanga, *supra* note 25; E. Peramo, C. Cheng & M. Cordel, *Juris2vec: Building Word Embeddings from Philippine Jurisprudence*, in *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 121–125; I. Chalkidis & D. Kampas, *Deep learning in law: Early adaptation and legal word embeddings trained on large corpora*, in *Artificial Intelligence and Law*, vol. 27(2), 2019, 171–198; A. Mandal, K. Ghosh, S. Ghosh, & S. Mandal, *Unsupervised approaches for measuring textual similarity between legal court case reports*, in *Artificial Intelligence and Law*, vol. 29(1), 2021, 1-35.

[103] The Neural Network in particular is a LSTM (Long-Short Term Memory Network). See S. Hochreiter & J. Schmidhuber, *Long Short-term Memory*, in *Neural Computation*, vol. 9(8), 1997, 1735-80. More generally, see S. Lai, K. Liu, S. He & J. Zhao, *How to Generate a Good Word Embedding*, in *IEEE Intelligent Systems*, vol. 31(6), 2016, 5–14; Y. Li & T. Yang, *Word Embedding for Understanding Natural Language: A Survey*, in S. Srinivasan (Eds.), *Guide to Big Data Applications. Springer International Publishing*, 2018, 83–104.

[104] This is based on the 'distributional hypothesis', which assum es that words which frequently occur together are usually also semantically related. While this approach might seem too simple to capture complex semantic meanings, the success of algorithms relying on it suggests that the claim has some merit. E. Altszyler, M. Sigman, S. Ribeiro & D. F. Slezak, *Comparative study of LSA vs Word2vec embeddings in small corpora: A case study in dreams database*, in *Consciousness and Cognition*, 56, 2017, 178–187.

However, the meaning of the vectors in the model depends on their relative positions in the respective corpus; the vector of a single word alone does not give us any insights. To test if there is evidence of different semantic use of the same words between two texts, we had to assess the distance between vectors from the two different corpora corresponding to the same words. To align them, we transformed the two models geometrically[105]. This allows us to understand how a vector in one corpus relates to the vector of another corpus. After the transformation, the vectors of the two aligned corpora are comparable to each other.

For each corpus we trained a different word embedded space, and we aligned each pair of words occurring in both corpora through the means of Unsupervised Vector Space Alignment[106].

Annex 3 Making sense of semantic distance

3.1 *The Data*

1. *List of terms from glossaries*[107]

E-commerce directive, P2B regulation, glossary of terms for DSA' questionnaire:

1- Application Programming Interface
2- Collaborative Economy Platform
3- Competent Authorities
4- Content Provider
5- Digital Service

[105] To perform this transformation, we used a "control vocabulary", containing a list of words that we can safely assume that share the same semantical meaning. The list of 1,189 words we used is, in fact, composed mainly of numbers and stop-words (like e.g., 'the'). We are thankful to Professor Julian Nyarko from Stanford University for providing us with a first list of Control keywords, to which we further added almost 2000 numerals and stop-words from the different corpuses.

[106] We used a special algorithm provided by Facebook in the library FastText. (https://github.com/facebookresearch/fastText), used in Python. P. Bojanowski, E. Grave, A. Joulin, & T. Mikolov, *Enriching Word Vectors with Subword Information*, 2017. http://arxiv.org/abs/1607.04606 (accessed on Jan. 22, 2021).

[107] Terms gathered from glossaries attached to all legislation recalled by the DSA and DMA proposals plus terms taken from the glossary attached to the DSA questionnaire.

6- Harmful Behaviours
7- Activities Online
8- Hosting Service Provider
9- Information Society Service
10- Illegal Content
11- Illegal Goods
12- Illegal Hate Speech
13- Intermediary Service
14- Intermediation Services
15- Law Enforcement Authorities
16- Notice
17- Notice Provider
18- Online Advertising
19- Online Platforms
20- Online Platform Ecosystems
21- Recommender Systems
22- Scaleup, Smart Contracts
23- Start-up
24- Trusted Flagger
25- User
26- Gatekeeper
27- Core Platform Service
28- Digital Sector
29- Online Intermediation Services
30- Online Search Engine
31- Online Social Networking Service
32- Video-Sharing Platform Service
33- Number-Independent Interpersonal Communications Service
34- Operating System
35- Cloud Computing Services
36- Software Application Stores
37- Software Application
38- Ancillary Service
39- Identification Service
40- End User
41- Business User

42- Ranking, Data
43- Personal Data
44- Non-Personal Data
45- Undertaking
46- Control
47- Recipient
48- Consumer
49- Offer Services
50- Trader
51- Intermediary Service
52- Illegal Content
53- Dissemination
54- Distance Contract
55- Online Interface
56- Digital Services Coordinator Of Establishment
57- Digital Services Coordinator Of Destination
58- Advertisement, Recommender System
59- Content Moderation
60- Terms And Conditions
61- Service Provider
62- Established Service Provider
63- Commercial Communication
64- Regulated Profession
65- Coordinated Field
66- Business User
67- Provider
68- Corporate Website User
69- Ranking
70- Mediation
71- Durable Medium

From DGA proposal:

72- Access
73- Re-Use
74- Metadata
75- Data Altruism
76- Data User

77- Data Holder
78- Data Sharing Main Establishment
79- Public Sector Body
80- Bodies Governed by Public Law
81- Public Undertaking
82- Secure Processing Environment
83- Representative

From NIS (Network and Information Systems):[108]

84- Network And Information System
85- Security Of Network And Information Systems
86- National Strategy On The Security Of Network And Information Systems
87- Operator Of Essential Services
88- Digital Service Provider
89- Incident
90- Incident Handling
91- Risk
92- Standard
93- Specification
94- Internet Exchange Point (IXP)
95- Domain Name System (DNS)
96- DNS Service Provider
97- Top-Level Domain Name Registry
98- Online Marketplace

From GDPR:

99- Processing
100- Restriction Of Processing
101- Profiling
102- Pseudonymisation
103- Filing System
104- Controller
105- Processor

[108] EU rules on the security of Network and Information Systems (NIS) are at the core of the Single Market for cybersecurity. The Commission proposes to reform these rules under a revised NIS Directive to increase the level of cyber resilience of all relevant sectors, public and private, that perform an important function for the economy and society. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX-:32016L1148&from=EN

106- Third Party
107- Consent
108- Personal Data Breach
109- Genetic Data
110- Biometric Data
111- Data Concerning Health
112- Enterprise
113- Group Of Undertakings
114- Binding Corporate Rules
115- Supervisory Authority
116- Supervisory Authority Concerned
117- Cross-Border Processing
118- Relevant And Reasoned Objection
119- International Organisation

II. *Manually coded from the questionnaires on DSA and DMA*
Manually coded from Questionnaire for the public consultation on a New Competition Tool

1- Access to data
2- adjacent/neighbouring markets
3- aftermarket
4- algorithm-based technological solutions
5- alignment of prices
6- anti-competitive
7- appropriateness
8- barriers to enter
9- binding
10- case-by-case
11- choice
12- competition
13- concentrated market
14- conditions of competition
15- copyright
16- customer lock-in
17- customer switching costs

18- data accumulation
19- data dependency
20- digital markets
21- digitisation
22- dominance-based
23- dominant
24- dual role situations
25- economies of scale
26- economies of scope
27- extreme economies of scale
28- fixed operating costs
29- gatekeeper
30- global distribution footprint
31- homogeneity of products
32- incomplete or misleading information
33- increased transparency
34- incumbency advantages
35- incumbency advantages
36- information asymmetry
37- innovation
38- inspections
39- interim measures
40- investigative powers
41- judicial review
42- lack of access to data
43- lack of competition
44- lack of transparency
45- leveraging
46- lock-in effects
47- market concentration
48- market dominance
49- market entry
50- market player
51- market power
52- market share
53- market-sharing cartels

54- monopolisation
55- multi-homing
56- multi-sided markets
57- network effects
58- new competition tool
59- non-binding recommendation
60- oligopolist
61- oligopolistic market structures
62- oligopoly
63- online platform
64- patents
65- penalties
66- platform
67- policy options
68- price increases
69- price leader
70- price leader-follower behavior/behaviour
71- price-fixing
72- pricing algorithms
73- procedural safeguards
74- proportionality
75- recommendations
76- regulatory barriers
77- related market
78- request of information
79- single-home
80- start-up costs
81- structural lack of competition problem
82- structural risk for competition
83- switching
84- tacit collusion
85- tailored remedies
86- tipping
87- tipping markets
88- transparency
89- two-sided markets

90- vertical integration
91- voluntary commitments
92- zero-pricing

Terms manually coded from DSA questionnaire

1- accountability
2- advertisement
3- algorithmic process
4- app store
5- appropriate
6- auction
7- automated detection
8- banning
9- bargaining power
10- behavioural advertising
11- blog hosting
12- bullying
13- business users
14- child sexual abuse material
15- complaint
16- conglomerate
17- conglomerate effect
18- consumer rights
19- content moderation
20- contestable
21- contextual advertising
22- control mechanism
23- counter-notice
24- coverage
25- cyber security
26- data sharing
27- dependency
28- digital identity
29- disabling
30- discrimination
31- disinformation

32- disputes
33- dissemination
34- divisive messages
35- due diligence
36- effective
37- effective measures
38- enforcement
39- ex-ante rules
40- fast-track assessment
41- flagging
42- fundamental rights
43- gender equality
44- governance
45- grooming
46- harmful
47- hate speech
48- illegal content
49- illegal medicine
50- information disclosure
51- institutional cooperation
52- internal practices
53- interoperability
54- know your customer
55- large online platform companies
56- leverage
57- liability
58- manipulation
59- market entry
60- national level
61- non-discrimination
62- non-payment
63- notice-and-action
64- notice-and-takedown
65- notifications
66- operating systems
67- oversight

68- pet trafficking
69- platforms' content policies
70- political advertising
71- price comparison
72- primary activities
73- programmatic advertising
74- proportionate
75- quality standards
76- Rating and reviews
77- Real-time bidding
78- recommendation
79- redress
80- Referral
81- reinstated content
82- removal
83- remuneration
84- reporting procedure
85- search engines
86- sector specific rules
87- self-employed
88- sharing
89- social networks
90- solidarity
91- suspension
92- tailored
93- takedowns
94- terrorist propaganda
95- trusted organisations
96- trusted researchers
97- unfair
98- unfair practices
99- unfavorable
100- user base
101- very large online platform companies
102- video sharing

Terms manually coded from the DSA and DMA proposals:
1- easily accessible
2- clear
3- unambiguous
4- specific
5- easily comprehensible
6- available
7- detailed
8- easy to access
9- user-friendly
10- precise

3.2 *Statistical test*

To see if there is evidence for a statistically significant semantic difference between the use of a term between the different stakeholder groups, we must perform a statistical test of their relative distance. We can model the relative distance d_t^{AB} dABt of a word t in the corpus A and B be as:

$$d_t^{AB} = y_t^{AB} + \mu_t^{AB} + u_t^{AB}$$

This takes into account a semantical term y_t^{AB}, a non-semantical term (originated from the simple different words disposition in the two corpora) and a random term . More precisely, the semantic term is defined as the difference in the usage of the same word which is driven by different understandings of the meaning of this term. Hence, this is the term we are interested in. On the other hand, the non-semantic term is defined as the term capturing all the non-semantic differences in usage, which can emanate from more frequent use of the word in different contexts, different authors, or stylistic differences. Finally, we define the random term as random differences in usage unrelated to systematic differences between the corpora. These could arise from the document-production process or the randomness of the initialization of the word-embedding algorithm's training[109].

The statistical test we performed is based on two assumptions. Our first assumption is that words in the control vocabulary used for the

[109] Nyarko & Sanga, *supra* note 102.

Vector Space Alignment Transformation do not have a semantic difference, i.e., $\gamma_t^{AB} = 0$. Consequently, their relative distance can give an empirical distribution of the non-semantical distance between words, composed of the only two terms $\mu_t^{AB} + u_t^{AB}$ which is our secondassumption. In this manner, it is possible to construct an empirical cumulative distribution of these distances, distributed with the hypothesis of zero semantic difference.

We first built an empirical Fisher-Snedecor distribution of distances calculated with all the common words included in the Control Vocabulary. We then analyzed the distance between the vectors of a word in the two corpora, counting the number of times these values were smaller than the control words' distances in the distribution. If we accept the null hypothesis that the word we are analyzing shows no semantic difference between the different corpora, then the obtained (normalized) p-value tells us the probability to have a distance equal or greater than that. If this probability is small enough, we can refuse this null hypothesis with a small possibility of error. This is to say that the particular word has, indeed, a statistically significant semantic difference in the two corpora. A general acceptance value for the p-value is 0.05, which we will use as the critical threshold for our analysis.

Annex 4. Cumulative distribution of semantic differences

Figures 1 to 3 show the cumulative distribution of distances of control dictionary words (in blue) against the cumulative distribution of distances and similarities of analyzed words (in red) for each corpus pair (i.e., corpus X against corpus Y). The plot shows that the words we analyzed create a statistical distribution different from the one of the common words, as we can see from the different shapes. These differences suggest that there are significant semantic differences between the corpora.

Figure 1. Corpuses AB - Cumulative distribution of control distances (top) and similarities (bottom)

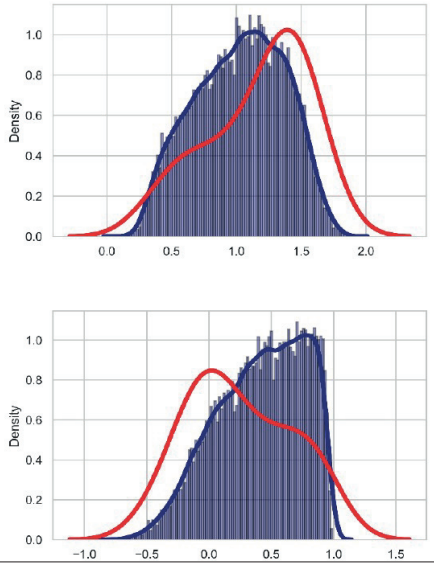


Figure 2. Corpuses BC - Cumulative distribution of control distances (left) and similarities (right)

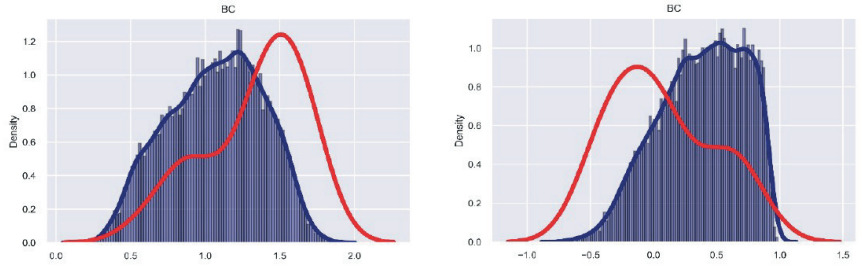
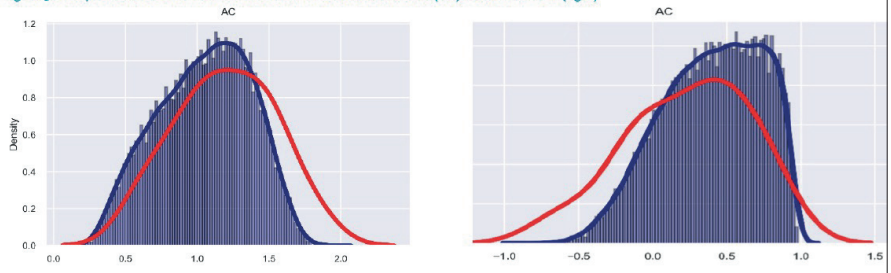


Figure 3. Corpus Pair AC - Cumulative distribution of control distances (left) and similarities (right)



Fabiana Di Porto, Tatjana Grote,
Gabriele Volpi, Riccardo Invernizzi

"I See Something You Don't See": A Computational Analysis of the Digital Services Act and the Digital Markets Act

Abstract: In its latest proposals, the Digital Markets Act (DMA) and Digital Services Act (DSA), the European Commission puts forward several new obligations for online intermediaries, especially large online platforms and "gatekeepers." Both are expected to serve as a blueprint for regulation in the United States, where lawmakers have also been investigating competition on digital platforms and new antitrust laws passed the House Judiciary Committee as of June 11, 2021. This Article investigates whether all stakeholder groups share the same understanding and use of the relevant terms and concepts of the DSA and DMA. Leveraging the power of computational text analysis, we find significant differences in the employment of terms like "gatekeepers," "self-preferencing," "collusion," and others in the position papers of the consultation process that informed the drafting of the two latest Commission proposals. Added to that, sentiment analysis shows that in some cases these differences also come with dissimilar attitudes. While this may not be surprising for new concepts such as gatekeepers or self-preferencing, the same is not true for other terms, like "self-regulatory," which not only is used differently by stakeholders but is also viewed more favorably by medium and big companies and organizations than by small ones. We conclude by sketching out how different computational text analysis tools, could be combined to provide many helpful insights for both rulemakers and legal scholars.

* This article was first published in *Stanford Computational Antitrust*, vol. 1, p. 85, 2021. This paper is part of a broader interdisciplinary research project on algorithmic disclosure regulation, funded by a three-year PRIN research grant awarded by the Italian Ministry of Research (no. of grant 2017BAPSXF, www.lawandtechnology.it). We are thankful to Prof. Julian Nyarko from Stanford University for providing some of the data we used in the empirical part and to a number of other people for insightful discussions: Dr. Roberto Marseglia, Proff. Daniela Piana, Giuseppe Italiano, Giorgio Monti, the discussants at the Ascola annual conference of 2021. All mistakes remain our own.

1. *Introduction*

"It is complex but we are looking forward to it."[1] This is the closing remark of European Commission Executive Vice-President Margrethe Vestager's statement when introducing the latest Commission proposals on digital platforms: the Digital Markets Act and the Digital Services Act[2].

Without a doubt, managing competition on digital markets is not a simple endeavor. This complexity is reflected by the lively scholarly debate on new rules for digital markets[3]. In December 2020, the Commission published what everyone involved in these debates had been looking forward to: twin proposals suggesting many new rules for "online platforms,"[4] "very large online platforms," and "gatekeepers[5]."

[1] M. Vestager, Statement by Executive Vice-President Vestager on the Commission Proposal on New Rules for Digital Platforms, Eur. Comm'n (Dec. 15, 2020), https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_20_2450.

[2] Eur. Comm'n, Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC, COM/2020/0825 (Dec. 15, 2020) [hereinafter, DSA Proposal; the draft regulation therein, hereinafter DSA]; Eur. Comm'n, Proposal for a Regulation of the European Parliament and the Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act), COM/2020/0842 (Dec. 15, 2020) [[hereinafter, DMA Proposal; the draft regulation therein, hereinafter DMA].

[3] *See, e.g.*, P. Ibáñez Colomo, *Whatever Happened to the 'More Economics-Based Approach'?*, in *J. Eur. Competition L. & Prac.*, vol. 11, 473, 2020, p. 473 (discussing the growing shift from the so-called "more economic approach" to ex-ante intervention against big digital platforms in the European legal community). For challenges related to competition law, see generally, A. Ezrachi & M. E. Stucke, *Virtual Competition: The Promise And Perils Of The Algorithm-Driven Economy*, 2016; and P. Marsden & R. Podszun, *Restoring Balance To Digital Competition——Sensible Rules, Effective Enforcement*, 1-87, 2020. On consumer protection and its relation to data protection and competition law, see W. Kerber, *Digital Markets, Data, and Privacy: Competition Law, Consumer Law, and Data Protection*, in *J. Intell. Prop. L. & Prac.*, vol. 11, 856, 2016. More recently, UNCTAD, Competition law, policy and regulation in the digital era, TD/B/C.I/CLP/57, 28 April 2021, https://unctad.org/system/files/official-document/ciclpd57_en.pdf (accessed on Jul. 7, 2021) (providing an excellent and up-to-date overview of the debate).

[4] There is no perfect overlapping between the legal definitions of "platforms" in the DSA and DMA. In the DSA, the widest concept is that of online "intermediary service," which covers all services within the scope of art. 1(3), including "online platforms" (providing hosting services), as described in art. 2(1)(h). Conversely, in the DMA, the widest concept is that of "core online platform," described in art. 2(2), which covers "online intermediation services" (inclusive of application stores), together with other services (for example, cloud computing services, social networking sites, video-sharing platforms, search engines, operating systems, and advertising services).

[5] See DSA Proposal, *supra* note 2; DMA Proposal, *supra* note 2.

These rules might serve as a "blueprint for regulation across the globe[6]," for at least two reasons. First, they are not only relevant to EU businesses and consumers, but to businesses around the world, especially well-known US tech companies[7].

Second, given that US lawmakers, agencies and state attorney generals have been investigating competition on digital platforms recently[8], and that new antitrust laws passed the House Judiciary Committee as of June 11, 2021[9], the newly proposed EU rules might have some visible repercussions on the US legal landscape. To name just a few: If the acts are adopted, "very large online platforms[10]," will need to *provide meaningful information* on how they manage, (DSA art. 23), and *rank* their content[11]. The DMA addresses platforms designated as gatekeepers"[12] with a list of

[6] A. Blankertz & J. Jaursch, *How the EU Plans to Rewrite the Rules for the Internet*, Brookings. (Oct. 21, 2020), https://www.brookings.edu/techstream/how-the-eu-plans-to-rewrite-the-rules-for-theinternet.

[7] See DSA art. 1(3); DMA art. 1(2).

[8] See Subcomm. On Antitrust, Com. & Admin. L., H. Comm. On The Judiciary, 116th Cong., Investigation Of Competition In Digital Markets, 2020, https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf; see, e.g., Complaint for Injunctive and Other Equitable Relief, Fed. Trade Comm'n v. Facebook, Inc., No. 1:20-cv-03590 (D.D.C. Dec. 9, 2020); Complaint, New York v. Facebook, Inc., No. 1:20-cv-03589 (D.D.C. Dec. 9, 2020); Complaint, United States v. Google LLC, No. 1:20-cv-03010-APM (D.D.C. Oct. 20, 2020).

[9] In fact, the US House Judiciary Committee only recently passed six acts dealing with this very topic. Most notably, H.R. 3849, ACCESS Act of June 11, 2021, 117th Congress (2021-2022), https://www.govtrack.us/congress/bills/117/hr3849 (Augmenting Compatibility and Competition by Enabling Service Switching) focuses on interoperability and data portability; H.R. 3816 (American Choice and Innovation Online Act) of June 11, 2021, 117th Congress (2021-2022), https://www.govtrack.us/congress/bills/117/hr3816 addresses self-preferencing and other anticompetitive conducts (all accessed on Aug. 3, 2021).

[10] Some online platforms are subject to additional obligations: these are the "very large online platforms," defined as those with at least 45 million monthly active users in the EU. See DSA art. 25.

[11] DSA art. 29 requires online platforms to disclose the main parameters used in recommender systems so as to prevent self-preferencing practices in rankings.

[12] In the DMA art. 3, some providers of core online services may be designated as "gatekeepers". There are two different mechanisms for this designation. The default one is based on three quantitative criteria set out in DMA art. 3(2): (1) a turnover equal to or above EUR 6.5 billion or an average market capitalization of at least EUR 65 billion; (2) more than 45 million monthly active end users, and more than 10.000 yearly active business users established in the Union; (3) the second criteria being met for three consecutive financial years. If a platform fulfills all three criteria, it must notify the Commission (DMA art. 3(3)) and will be designated as a gatekeeper within 60 days unless it presents proof that it does not match the description of a gatekeeper in

dos and don'ts. For instance, it prohibits selfpreferencing[13], while (obliging them to enable multi-homing and interoperability[14].

As can be seen from these examples, terms and concepts like "gatekeepers," "self-preferencing," and "interoperability" play central roles in designing new rules for online platforms. Inherently, this comes with the challenge of defining these new concepts to make them legally operable[15]. At the same time, there might be some confusion on how exactly some more "traditional" qualifiers, e.g., "dominant" position or "anti-competitive" conduct, apply to digital services[16].

Consequently, a large part of the debate on new rules for digital markets prior to the publication of the EC twin proposal has centered around how to define or understand certain essential concepts[17].

In this article, we aim to investigate how diffused is the consensus

DMA art. 3(1) despite fulfilling the criteria in DMA art. 3(2), DMA art. 3(4). Thus, that a company is a gatekeeper is based on a rebuttable presumption. There is a second mechanism to establish if a company is a gatekeeper. If the thresholds are not met, the provider of a "core platform service" can still be identified as a gatekeeper, DMA art. 3(6), after a market investigation by the Commission, DMA art. 15. Hence, not every very large platform is a gatekeeper, but it is likely that every gatekeeper will also be a very large online platform. See DMA art. 3(2)(b).

[13] To be precise, the DMA obliges gatekeepers to "refrain from treating more favorably in ranking services and products offered by the gatekeeper itself or by any third party belonging to the same undertaking compared to similar services or products of third party," DMA art. 6(1)(d).

[14] Multi-homing (DMA art. 5(e)) is the possibility to use the services of more than one platform simultaneously. Interoperability (DMA arts. 5(f)) is both the compatibility of protocols (protocol interoperability) and the possibility to access data in real-time for both the data subject and entities acting on the data subject's behest (data interoperability). See J. Crémer, Y-A De Montjoye & H. Schweitzer, *Competition Policy For The Digital Era*, 83-84, 2019, https://data.europa.eu/doi/10.2763/407537 [hereinafter Crémer Report].

[15] DMA arts. 7-8.

[16] For information on measuring market power, see Crémer Report, *supra* note 14, pp. 48-50.

[17] For instance, there has been a vivid debate on the scope of application of the Digital Services Act and the Digital Markets Act, in particular. For a comment on how to best set the criteria for application of the DMA and how to designate "gatekeepers," see D. Geradin, *The Digital Markets Act: How Should Ex Ante Rules Look Like?*, in *Platform L. Blog* (Oct. 23, 2020), https://theplatformlaw.blog/2020/10/23/thedigital-markets-act-how-should-ex-ante-rules-look-like. For a discussion of the relationship between dominance and gatekeeping power, see T. Käseberg, *Antitrust 2.0—Governance of Oversight Over Digital Gatekeepers*, in *Kluwer Competition L. Blog* (Dec. 14, 2020), http://competitionlawblog.kluwercompetitionlaw.com/2020/12/14/antitrust-2-0-governance-ofoversight-over-digital-gatekeepers.

about shifting towards new tools and concepts of competition law among the stakeholders. Do all stakeholder groups share the same understanding and use of the relevant terms and concepts of the DSA and DMA? Or can we identify different attitudes towards these issues?

This is relevant for several reasons. First, a different understanding of legal terms may jeopardize their application by the stakeholders. Think for instance to the many transparency obligations and disclosure duties that require information to be given in a "clear" and "easy-to-understand" manner. The precise enforcement of such duties utterly depends on the behavior of the drafter (mainly medium and big firms) and users, which in turn should have a common understanding and use of said terms. Second, not having a homogeneous understanding and use of terms may cause big reforms like the ones we are discussing to fail to achieve their goals.

For example, defining the scope of application of the new rules involves the clear identification of new terms, like gatekeepers. The same can be said with regard to self-preferencing, an anticompetitive conduct which definition is blurry. Third, on a rule-making level, unshared use and understanding of relevant terms may puzzle the consultation process hold by the EC over what became the DSA and DMA[18].

According to the Better Regulation agenda 2021[19], consultations are key to enhance transparency and participation among stakeholders, thus building consensus around new regulatory intervention[20].20 If some (new) terms are not consistently used, then can consensus really be reached?

For instance, in assessing the feedback it received, the Commission concluded that all stakeholders demanded new rules for gatekeepers

[18] The DSA and DMA were developed from three different Inception Impact Assessment documents: Eur. Comm'n, Digital Services Act Package: Deepening The Internal Market And Clarifying Responsibilities For Digital Services, Inception Impact Assessment, 2020, available at https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12417-Digital-Services-Actdeepening-the-internal-market-and-clarifying-responsibilities-for-digital-services_en; Eur. Comm'n, Digital Services Act Package: Ex Ante Regulatory Instrument For Large Online Platforms With Significant Network Effects Acting As Gate-Keepers In The European Union's Internal Market, Inception Impact Assessment, 2020, available at https://ec.europa.eu/info/law/betterregulation/have-your-say/initiatives/12418-Digital-Services-Act-package-ex-ante-regulatoryinstrument-of-very-large-online-platforms-acting-as-gatekeepers; and Eur. Comm'n, New Competition Tool ('Nct'), Inception Impact Assessment, 2020, available at https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12416-New-competition-tool.

[19] https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulationwhy-and-how_en.

[20] C. M. Radaelli, *Regulating Rule-making via Impact Assessment*, in *Governance*, vol. 23, 89, 2010.

which explicitly prohibit "anti-competitive" practices[21]. This sounds like a consensus at first, but how can we know all respondents are having the same platforms and practices in mind when the key terms they are using are themselves still up to debate?

While some might think of gatekeepers as dominant platforms in the sense of Treaty on the Functioning of the European Union art. 102 others might set a much lower threshold[22].22And while some might think that self-preferencing leads to foreclosure, others might view it as necessary to penetrate the market.

In order to investigate these issues, we leverage the power of computational tools. More specifically, we use supervised and unsupervised Machine Learning to analyze the debate preceding the publication of the DSA and DMA proposals and address the following questions: (1) Does the use and understanding of terms related to competition in digital markets differ across different groups of stakeholders? (2) What are stakeholders' attitudes towards certain contentious terms?

On top of these substantiative questions, we further discuss (3) the extent to which computational tools can help automate and enrich the analysis of documents that are used as inputs in the rulemaking process. If computational tools really can help in this analysis, then such approaches should be recognized as valuable additions to the "manual" qualitative analysis the Commission currently undertakes.

This Article is structured as follows. The following Part I sets the larger context by outlining some novel competition challenges posed by digital markets that led to the proposed ex-ante regulatory response set out in the DSA and DMA. In Part II, we present our methodology and show, first, that similar opinions are expressed by groups usually belonging to different clusters (i.e., medium and big organizations), and second, groups of stakeholders use central terms of the DSA and DMA in different ways. Lastly, in Part III, we conclude by suggesting what this evidence should tell us about the two proposals and, on a more overarching tone, how computational methods could support EU targeting rules, although very cautiously.

[21] See DMA proposal, 7–8.

[22] Dominance has been defined as "a position of economic strength enjoyed by an undertaking, which enables it to prevent effective competition being maintained on a relevant market". While the assessment of market power is based on several factors, an undertaking is generally not considered dominant if it has a market share of below 40% in the relevant market. Eur. Comm'n, Communication From The Commission — Guidance On The Commission's Enforcement Priorities In Applying Article 82 Of The Ec Treaty To Abusive Exclusionary Conduct By Dominant Undertakings, 2009/C 45/02, at 8-9.

2. *Competition In The Digital Era And The Proposed Regulatory Response In The Dsa And Dma*

There are manifold challenges arising from digital markets that standard competition rules seem not to tackle adequately. This Section will give a snapshot of those inadequacies that were most debated, highlighting what ex-ante rules received the greatest support, as a consequence (A). We will then have a closer look at the DSA and DMA proposals, to see in what way such debate and the perceived needs translated into actual norms (B).

A – Why We Need the DSA and DMA: New Buzzwords

Consumers benefit in many ways from the impressive development of digital markets[23]. However, lawmakers and scholars alike have been emphasizing that digital markets show certain characteristics which are likely to favor highly concentrated markets.

1. *Seizing Market Power of Digital Platforms*

First, many business models in the digital realm are characterized by strong returns to scale[24]. Second, incumbents in online markets are particularly hard to dislodge due to substantial network effects[25]. Third, due to the data dependency of many online services, established players might hold a competitive edge over small contestants by leveraging the

[23] See DSA recital 1. On the DSA see A. de Streel & S. Broughton Micova, Centre On Regulation In Europe, *Digital Services Act – Deepening The Internal Market And Clarifying Responsibilities For Digital Services* at 39 (2020), https://cerre.eu/publications/digital-services-actresponsibility-platforms/ (also referred as Cerre DSA report, in the following).

[24] Crémer Report, *supra* note 14, p. 3; M. S. Gal & N. Petit, *Radical Restorative Remedies for Digital Markets*, in *Berkeley Tech. L. J*., vol. 36, 617, 2021, (manuscript 5-6), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3687604#); Stigler Center For The Study Of The Economy And The State, *Report Of The Committee For The Study Of Digital Platforms, Market Structure And Antitrust Subcommittee*, 2019, p. 14 [hereinafter Stigler Report]; Org. for Econ. Coop. & Dev. [OECD], Executive Summary of the Roundtable on Algorithms and Collusion, at 5, OECD Doc. DAF/COMP/M(2017)1/ANN3/FINAL (Sept. 26, 2018), https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DAF/COMP/M(2017)1/ANN3/FINAL&docLanguage=En.

[25] Crémer Report, *supra* note 14, p. 3; P. Georg Picht & G. Tazio Loderer, *Framing Algorithms: Competition Law and (Other) Regulatory Tools*, in *World Competition*, vol. 42, 391, 2019, p. 406.

power of the large amounts of data they accumulate[26].

While the extent of these advantages could be limited by multi-homing and interoperability[27], there still is a very real chance that certain platforms accumulate some kind of "gatekeeping" power and impose the prices, conditions, and level of transparency they deem appropriate, for their own convenience. In its investigation report on competition in digital markets, the US Congress Subcommittee on Antitrust, Commercial Law and Administrative Law addressed the last-mentioned aspect: "Without transparency or effective choice, dominant firms may impose terms of service with weak privacy protections that are designed to restrict consumer choice, creating a race to the bottom[28]." This would further increase the danger of "tipped markets" in which one company takes the large majority of the market share[29].

While there is broad consensus that these dynamics can be highly problematic, the exact threshold for a tipped market, or the precise definition of a "gatekeeper" position arising from such a situation is still up to debate[30]. Especially the relation between the legal concept of dominance and the definition of gatekeepers has been the subject of a vivid debate among scholars[31].

[26] Crémer Report, *supra* note 144, p. 2.

[27] When users can multi-home, they might switch to better service providers or use services in parallel, which will increase competition. Interoperability often is a pre-condition for multi-homing and might further allow users to unbundle or establish complementary services. Crémer report, *supra* note 14), at 23, 83-84. For definitions of multi-homing and interoperability, see note 11.

[28] Subcomm. On Antitrust, Com. & Admin. L. Of The Comm. On The Judiciary, H. Comm. On The Judiciary, 116th Cong., Investigation Of Competition In Digital Markets, 2020. The Subcommittee report also mentions manipulative design interfaces, so-called dark patterns, nudging consumers into certain choices. *Id*. at 53.

[29] Digit. Competition Expert Panel, Hm Treasury, Unlocking Digital Competition: Report Of The Digital Competition Expert Panel 3-4, 88 (2019), https://doi.org/10.17639/wjcs-jc14.

[30] Discussions about how to best designate "gatekeepers" started before the proposal was published and continued in the aftermath of its publication. For a critical view on the gatekeeper definition of Art. 3 DMA, see L. Cabral, J. Haucap, G. Parker, G. Petropoulos, T. Valletti, & M. Van Alstyne, Eur. Comm'n, *The EU Digital Markets Act: A Report from a Panel of Economic Experts*, 2021, p. 9, https://publications.jrc.ec.europa.eu/repository/handle/JRC122910; Cristina Caffarra & Fiona Scott Morton, *The European Commission Digital Markets Act: A Translation*, VOX EU CEPR (Jan. 5, 2021), https://voxeu.org/article/european-commission-digital-markets-act-translation (accessed 3 Apr. 2021).

[31] For a comprehensive discussion of possible criteria for gatekeepers, see A. De Streel, Ctr. On Regul., in Eur., *Digital Markets Act: Marking Economic Regulation Of Platforms*

2. *Tackling Anti-Competitive Conduc*t

Besides these structural challenges, a lot of the competition law debate on digital markets has centered around some specific anti-competitive conducts. First, the issue of *algorithmic* collusion and *tacit* collusion was heavily discussed. Art. 101 TFEU notoriously only condemns collusive arrangements, leaving uncoordinated parallel behavior untouched. For a long time, this was not perceived as a problem since tacit collusion is almost never stable on non-digital markets[32]. However, the increased concentration, transparency, entry barriers and interaction frequency of online markets make tacit collusion, especially through algorithmic means, a theoretically very plausible scenario[33]. Although the practical relevance of this phenomenon is still disputed[34], there is no denying that collusion has received considerable attention among scholars and competition authorities.

Second, the list of digital competition buzzwords further includes "selfpreferencing."[35] Whenever platforms somehow influence consumer choice, e.g., by presenting offers from business customers in a certain order, they have the possibility to favor their own products or services. The *Google Shopping* case established that self-preferential placements are,

Fit For The Digital Age 35-44 (Nov. 24, 2020) [hereinafter Cerre DMA report]. While UK authorities intentionally eschewed the traditional notion of dominance (see, e.g., HM Treasury, *Unlocking digital competition, Report of the Digital Competition Expert Panel*, Mar. 13, 2019, https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digitalcompetition-expert-panel (accessed 25 February 2021), at 55, Germany's tenth amendment of its Act Against Restraints of Competition (Gesetz gegen Wettbewerbsbeschränkungen, GWB) explicitly includes dominance in one or several markets as part of its definition of "paramount significance" (§ 19(1) no. 1 GWB), the equivalent to the DMA's "gatekeepers."

[32] F. Beneke & M-O Mackenrodt, *Remedies for Algorithmic Tacit Collusion*, in *J. Antitrust Enf't*, vol. 9, 152, 2021, 158-159.

[33] Bundeskartellamt & Autorite De La Concurrence, Algorithms And Competition, at II, 2019, https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Berichte/Algorithms_and_Competition_Working-Paper.html;jsessionid=BCD024895C5890BEDCAAB208FBCC6B8F.2_cid387?nn=3591568.

[34] For a critical view, see S. K. Mehra, *Robo-Seller Prosecutions and Antitrust's Error-Cost Framework*, in *Competition Pol'y Int'l: Antitrust Chron*, (May 15, 2017), https://www.competitionpolicyinternational.com/robo-seller-prosecutions-and-antitrusts-error-cost-framework/ (accessed on Apr. 3, 2021).

[35] For an in-depth discussion, see D. Geradin & D. Katsifis, *"Trust Me, I'm Fair": Analysing Google's Latest Practices in Ad Tech from the Perspective of EU Competition Law*, in *Eur. Competition J.*, vol. 16, 11, 2020. Self-preferencing was at the heart of the Microsoft saga (on which, see J. P. Jennings, *Comparing the US and EU Microsoft Antitrust Prosecutions: How Level Is the Playing Field?*, in *Erasmus L. & Econ. Rev.*, vol. 2, 71, 2006).

indeed, not compatible with competition law[36]. Only if the parameters used to rank products are transparent[37], will it be possible to know whether an online platform is distorting competition by preferencing certain offers, leaving consumers in the dark about the "trade-offs they are facing," and hence inhibiting competition in a significant manner.

3. *Regulatory Reform: The Need For Ex-Ante Rules*

Existing competition and consumer protection laws often fall short in addressing these issues, both in the EU and the United States. For instance, the US Subcommittee on Antitrust has noted that "some of these business practices are a detriment to fair competition, but they do not easily fit the existing categories identified by the Sherman Act, namely 'monopolization' or 'restraint of trade' or the Clayton Act[38]. The American Choice and Innovation Online Act of June 11, 2021[39], can thus be read as a means to prohibit the usage of exploitative practices by large online platforms and to promote users.

In the debate surrounding the DSA and DMA proposals, general shortcomings of EU competition rules when dealing with opaque online platforms have been highlighted[40]. Relying exclusively on Arts. 101 and 102 TFEU might mean that the Commission could only act on a case-by-case basis, for a very limited set of platforms, ex post, and only after lengthy investigations[41].

[36] In the Google Shopping case, self-preferential placements were deemed not compatible with competition law: Commission Decision C(2017) 4444 of 27 June 2017, Google Search (Shopping), 2018 O.J. (C 9) ¶¶ 9-10 of summary decision. See Pinar Ackman, The Theory of Abuse in Google Search: A Positive and Normative Assessment Under EU Competition Law, 2017 J. L., TECH. & POL'Y 301 (2017).

[37] Picht and Loderer, *supra* note 18, p. 416.

[38] Subcomm. On Antitrust, *supra* note 9, p. 396.

[39] US House Judiciary Committee, supra note 10. See e.g. Sect. 2.3 of the Act prohibiting discriminatory conduct against business users by those operating 'covered platforms' (a concept largely assimilable to the European 'gatekeeper').

[40] The Crémer report points out several criticalities: (1) Not all gatekeepers enjoy a dominant position in the sense of Art. 102 TFEU; (2) the relevant market might be substantially harder to define than in nondigital cases; (3) not every problematic practice has a demonstrable effect on the relevant market. The authors conclude that greater emphasis should be put on the theory of harm, instead. Crémer report, *supra* note 14, p. 3-4. Moreover, digital markets are often moving at a rapid pace, which is not necessarily a characteristic they share with competition law. Hence, there are concerns whether competition law could be applied with the necessary speed to address urgent competition needs. Cerre DMA report, *supra* note 24, p. 59; Recital 5 DMA

[41] This is not to say that existing competition norms are generally useless to address

Given that Section 2 of the Sherman Act (15 U.S.C. § 2) and Section 7 of the Clayton Act (15 U.S.C. § 18) are even more narrow than comparable EU competition norms, their suitability to achieve a satisfactory level of fairness and competition might be more limited, especially regarding effective remedies[42].

In light of these interconnected challenges for consumer protection and competition, a consensus has been reached on the need for new ex-ante rules[43] to complement the existing legal framework[44].

novel competition issues; in fact, EU case law has shown the opposite. E.g., Google Search (Shopping) Case C(2017) 4444, 27 June 2017; Commission Decision C(2018) 4761 of 18 July 2018., Google Android, 2019 O.J. (C 402). On the national level the German competition authority has taken action against certain data collection practices of Facebook (see Bundeskartellamt [Federal Cartel Office] 6 Feb. 2019, B6-22/16, Bundeskartellamt, p. 6, https://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Entscheidungen/Missbrauchsaufsicht/2019/B6-22-16.pdf?__blob=publicationFile&v=5). Also with a focus on data collection, the Commission has opened an investigation against Amazon in 2019 (European Commission Press Release IP/19/4291, Antitrust: Commission Opens Investigation into Possible Anti-Competitive Conduct of Amazon (17 July 2019), https://ec.europa.eu/commission/presscorner/detail/en/ip_19_4291). For older cases, see S. Weber Waller, *Access and Information Remedies in High-Tech Antitrust*, in *J. Competition L. & Econ.*, vol. 8, 575, 2012, p. 576. Regarding the United States, some argue in a similar vein: "This is not to say that the use of the antitrust laws should be abandoned. If history is a guide, there is a meaningful possibility that antitrust enforcement activities will produce value commensurate with their costs." T. Wheeler, P. Verveer, & G. Kimmelman, Shorenstein Ctr. On Media, Pol. & Pub. Pol'y, Harvard Kennedy Sch., *New Digital Realities; New Oversight Solutions In The U.S.*, 2020, p. 26, https://shorensteincenter.org/wp-content/uploads/2020/08/New-Digital-Realities_August-2020.pdf.

[42] Wheeler et al., *supra* note 33, p. 24-26. On fairness in competition law see A. Ezrachi, *EU Competition Law Goals and the Digital Economy*, Oxford Legal Studies Research Paper No. 17/2018, 2018, (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3191766) (accessed 27 January 2021), at 16; M. Stucke & A. Ezrachi, *Competition Overdose: How Free Market Mythology Transformed Us From Citizen Kings To Market Servants*, 2020; S. Zuboff, *The Age Of Surveillance Capitalism: The Fight For A Human Future At The New Frontier Of Power*, 2019.

[43] Discussions on ex-ante rules to complement ex post antitrust enforcement in digital markets can be found in Cerre DMA report, *supra* note 31, p. 26; L. Cabral; J. Haucap; G. Parker; G. Petropoulos; T. Valletti; & M. Van Alstyne, *The EU Digital Markets Act: A Report from a Panel of Economic Experts*, JRC122910, Feb. 9, 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3783436# (accessed on July 1, 2021); J. B. Baker, *Protecting and Fostering Online Platform Competition: The Role of Antitrust Law*, in *J. of Competition Law & Economics*, vol. 17, 493, 2021, (discussing the role of regulation to supplement antitrust law in fostering competition in online platforms).

[44] See, e.g., Algorithm Watch, *Governing Platforms – Final Recommendations*, 2020, https://algorithmwatch.org/wp-content/uploads/2020/10/Governing-Platforms_

B – How the DMA and DSA Proposals Respond to the Quest for New Pro-Competitive Rules

The European Commission's vision of these rules was first outlined in three separate inception impact assessments[45], that were open to consultation by stakeholders. As a result, the DMA and DSA normative proposals were adopted that encapsulate such discussion. In the following, we will briefly present their content.

1. D*efining The Scope: The Gatekeepers*

The DSA applies to all "intermediary services," while the scope of the DMA is limited to "core platform services" offered by "gatekeepers" as defined in Article 3 of the DMA. The DSA replicates the GDPR's approach to applicability, hence applying to all services provided to EU citizens "irrespective of the place of establishment of the providers of those services."[46] This is a characteristic it shares with the DMA[47].

However, the DMA is much more limited in scope since its goal is not to lay down a fundamental framework, but rather to complement existing competition law norms with respect to a very specific set of market players (the gatekeepers)[48].

Regarding its *ratione materiae*, the DMA has a more limited scope (the core platform services listed in Article 2 Section 2)[49], to be in line with its much more specific objective of "ensuring contestable and fair markets … *where gatekeepers are present.*"[50] Given the centrality of the "gatekeeper" concept, different understandings of this term could cast doubts at whether all stakeholders mean the same when they express their support for additional rules for "gatekeepers."

2. *Regulating Conducts: New Ex-Ante Rules*

DSARecommendations.pdf., p. 1; Cerre DSA report, *supra* note 23, p. 39; Comm. on the Internal Mkt. & Consumer Prot., Eur. Parliament, Draft Report with Recommendations to the Commission on Digital Services Act: Improving the Functioning of the Single Market, para. 15, at 5, 2020/2018(INL) (24 Apr. 2020); Comm. On Legal Affairs, Draft Report with recommendations to the Commission on a Digital Services Act: adapting commercial and civil law rules for commercial entities operating online, para. 2, at 5, 2020/2019(INL) (Apr. 22, 2020).

[45] See *supra* note 16.

[46] DSA art. 1, § 3.

[47] DMA art. 1, § 2.

[48] See *supra* notes 7-11 and accompanying text.

[49] Note that electronic communication network and services markets are exempted from the scope of the proposal (DMA art. 1 §3).

[50] DMA art. 1 §1 (emphasis added).

Turning to some of the substantive rules for said gatekeepers, under the new law gatekeepers will have to grant data access to business users for the data they generated on the platform and allow for multi-homing and interoperability[51]. Selfpreferencing will be explicitly prohibited[52], as will preventing users from uninstalling certain pre-installed tools[53]. With a view to advertisement markets, Article 5 Section g of the DMA would oblige gatekeepers to provide information about pricing and performance measuring tools to consumers of core platform services within the meaning of Article 3 Section 7 of the DMA[54]. This would allow consumers to assess how satisfied they are with the advertisement product they are paying for[55].

With a view to the connection between transparency and competition mentioned above, a look at the DSA is helpful to complete the picture. Its Articles 12(1), 13, 23, 24, 29, and 33 establish comprehensive but differentiated disclosure and reporting duties regarding ranking, advertisement, and content moderation practices[56]. The more pronounced transparency obligations for very large online platforms within the meaning of Article 24 of the DSA reflect the differentiated approach the Commission took for the design of the DSA, explicitly mentioned in Recital 39 of the proposal.

To sum up, this section has shown that certain characteristics of digital services might impair competition. It has also sketched out the debate on why existing rules might be insufficient to prevent this and how the DMA and DSA proposals of the European Commission seek to change this.

[51] DMA art. 5, §§ f, h, g.
[52] DMA art. 6, §1(d).
[53] DMA art. 5, §b.
[54] DMA art. 5, §g. See also DMA art. 3, §7.
[55] DMA art. 6, §1(g). While these obligations are rather specific, Article 10 of the DMA would open the door to add further duties in the future if a market investigation pursuant to Article 17 of the DMA identified a need to do so for the sake of safeguarding fair competition.
[56] Just to name those duties for very large online platforms within the meaning of Article 25 of the DSA. Article 29 of the DSA would entail the obligation to provide information on the use of recommender systems and their parameters. Article 33 sets out comprehensive transparency obligations for very large online platforms.

3. *A Computational Analysis Of The Dsa And Dma Consultation Process*

In motivating its twin proposals, the Commission reports that the "vast majority of respondents" in the consultation process "considered that dedicated rules on platforms should include prohibitions and obligations for gatekeeper platforms.[57]"

More specifically, in the DMA this majority believed that the "the proposed list of problematic practices, or 'blacklist,' should be targeted to clearly unfair and harmful practices of gatekeeper platforms.[58]" In the DSA, the quest for "algorithmic accountability and transparency audits, especially with regard to how information is prioritized and targeted" online comes from "a wide category of stakeholders," and is particularly voiced by "civil society and academics.[59]"

In this chapter, we ask whether these ex-ante rules for very large platforms and gatekeepers are what stakeholders demanded in the consultation process, and whether their actual wording in the DSA and DMA proposals reflects the way each stakeholder group uses the relevant terms. This is a relevant step, as it is important that the addressees of such duties (typically digital firms) and the beneficiaries (individuals, micro and small organizations using platforms) agree on their meaning.

To do so, we use computational text analysis techniques to analyze the replies to the questionnaires and position papers submitted by stakeholders to the EU consultation process for both proposals.

A. *Data and Methodology*

To analyze these stakeholder documents, we created a special scraper algorithm, which allowed us to download all the files automatically, convert them into text, and split them into three clusters. In doing this, we relied on the Commission's categorization, based on the organization size of the feedback contributors.[60] We then aggregated the different sub-

[57] DMA, Explanatory Memorandum at 8 (summarizing the results of stakeholder consultations and impact assessments).

[58] Ibid.

[59] DSA, explanatory memorandum at 9. See also Algorithm Watch, *supra* note 46, p. 1.; Cerre DSA Report, *supra* note 23, p. 39; Draft Report of the Committee on the Internal Market and Consumer Protection with Recommendations to the Commission on Digital Services Act: Improving The Functioning of the Single Market at 5, 2020/2018(INL) (n 39); Eur. Comm'n, White Paper on Artificial Intelligence - A European approach to excellence and trust, at 15, COM(2020) 65 final

[60] The Commission distinguishes between (1) individuals, micro (< 10 employees), (2)

categories into three corpuses, based on type and organizational size of the feedback contributor:

Corpus A (individuals and micro-organizations),

Corpus B (small companies/organizations), and

Corpus C (medium and large companies/organizations).

This clustering scheme was informed by a preliminary analysis of our data, which is explained in more detail in the Appendix (Annex 1). Interestingly and importantly, it showed that the classical "small and medium enterprises" group was not equally applicable to our data, since the feedback contributors do not only comprise enterprises, but a more diverse set of actors.

To discern differences in the use of certain key terms across stakeholder groups (i.e., a different semantic understanding of identical terms), we leveraged Word Embedding Models to quantify evidence of such differing understandings. This technique has already been used in various Natural Language Processing tasks,[61] and recently also in the Computational Law literature. More specifically, it has been demonstrated to be very powerful and useful in providing insights into latent differences in how language is used in legal scholarship[62] and is starting to be discussed in the computational antitrust literature[63].

small (< 50 employees), (3) medium (< 250 employees), and (4) large (250 or more) organizations as well as between different types of feedback contributors (i.e., in the DSA, p. 8 you find that respondents are: the general public (66%), companies/businesses organizations (7.4%), business associations (6%), and NGOs (5.6%) authorities (2.2%), academic/research institutions (1.2%), trade unions (0.9%), and consumer/environmental organizations (0.4%)). DSA, Explanatory Memorandum at 8.

[61] For instance, word embeddings have been used by sociologists to investigate the meaning of the term 'class', to predict conflict, or to classify documents. A. C. Kozlowski, M. Taddy & J. A. Evans, *The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings*, in *American Sociological Review*, vol. 84, 905, 2019; A. Kutuzov, E. Velldal & L. Øvrelid, *Tracing Armed Conflicts With Diachronic Word Embedding Models* in (T. Caselli et al., eds.) *Proceedings Of The Events And Stories In The News Workshop*, 31–36; X. Yang, C. Macdonald & I. Ounis, *Using Word Embeddings in Twitter Election Classification*, in *Information Retrieval Journal*, vol. 21, 183, 2018.

[62] Julian Nyarko & Sarath Sanga, *A Statistical Test for Legal Interpretation: Theory and Applications* (Nov. 25 2020) (unpublished article) (on author's website) https://juliannyarko.com/wpcontent/uploads/other/nyarko_sanga_legal_interpretation.pdf.

[63] For a discussion of computational techniques in antitrust law and their implications, see the following: T. Schrepel, *Computational Antitrust: An Introduction and Research Agenda*, in *Stan. Computational Antitrust*, vol. 1, 1, 2021; G. Massarotto & A. Ittoo, *Gleaning Insight from Antitrust Cases Using Machine Learning*, in *Stan. Computational Antitrust*, vol. 1, 16, 2021.

The core of this technique consists in training a special neural network to convert each word contained in a corpus of texts into a vector, i.e., a set of numbers[64]. While a simple algorithm would require formulating explicit rules to somehow approximate the semantic meanings of words, ML (or the neural network, to be precise) learns the implicit rules directly from the data we feed it.

The resulting vectors are based on the frequency with which words occur next to each other. The neural network tracks their relative positions in each phrase of the corpus and the correlation between words. The stronger two words are correlated (in their occurrence – and so in their semantic meaning)[65] in the corpus the model was trained in, the closer the corresponding vectors will be located to each other.

However, models trained on different corpuses are not directly comparable as they depend on the corpus the model was trained on. The vector of a single word alone does not provide meaningful insights. To test if there is evidence of a different semantic use of the same words between two texts, we assess the distance between vectors from two different corpuses corresponding to the same words. To align them, we transform the two models geometrically[66]. This allows to understand how a vector in one corpus relates to the vector of another corpus. After the transformation, the vectors of the two aligned corpuses are comparable. Therefore, for each corpus we trained a different word embedded space, and we aligned each pair of words occurring in both corpuses through the

[64] The Neural Network used in this research in particular is a LSTM (Long-Short Term Memory Network). See S. Hochreiter & J. Schmidhuber, *Long Short-term Memory*, in *Neural Computation*, vol. 9, 1735, 1997.

[65] This is based on the "distributional hypothesis," which assumes that words which frequently occur together are usually also semantically related. While this approach might seem too simple to capture complex semantic meanings, the success of algorithms relying on it suggests that the claim has some merit. E. Altszyler, M. Sigman, S. Ribeiro, & D. Fernandez Slezak, *Comparative Study of LSA vs Word2vec Embeddings in Small Corpora: A Case Study in Dreams Database*, in *Consciousness And Cognition*, vol. 56, 178.

[66] To perform this transformation, we used a 'control vocabulary', containing a list of words that we can safely assume that share the same semantical meaning. The list of 1,189 words we used is, in fact, composed mainly of numbers and stop-words (like e.g. 'the'). We are thankful to Professor Julian Nyarko from Stanford University for providing us with a first list of control keywords, to which we further added almost 2.000 stop-words and numerals we took from the different corpuses. We manually selected our control vocabulary. We used the glossaries contained in all EU directives and regulations recalled in the DSA and DMA proposals and published in the OJUE. Further, we manually coded the questionnaires (used in the consultation) and selected terms of interest.

means of Unsupervised Vector Space Alignment[67].

Having aligned our three corpuses, we are able to compute the distance between the same terms from different corpuses. However, how do we know that the distances we find are not just random, but actually based on substantial uses and understandings between stakeholders? To see if there is evidence for a statistically significant semantic difference between the use of a term between the different stakeholder groups, we must perform a statistical test (see Appendix, Annex 2 for detailed description). Otherwise, we would not be able to tell whether the differences we find between corpuses are actually relevant or merely signs of random, non-semantic differences between our corpuses. By modeling the theoretical distribution of non-semantic differences for the terms in our corpus, we can compare the distance we would expect to see if there was no semantic difference with the distance we observe. In this way, we can conclude with a certain level of confidence that the distance we observe between our corpuses is more than just random or syntactical.

In order to gain deeper insights into possible reasons for a semantic difference in the use of key words between different corpuses, we also leveraged the tool of Sentiment Analysis[68], applied to sentences of the two compared corpuses where the specific key word appears. Sentiment analysis is a Natural Language Processing technique, which classifies a sentence, or a paragraph based on the use of specific words and their location inside the text, giving as a score a value of the positive or negative sentiment inside the particular text. These two values are aggregated to a compound value, which gives a score to the overall sentiment ranging from -1 (totally negative) to +1 (totally positive).

B. Results and Discussion: Different Groups, Different Uses?

With these tools at hand, we were able to find a significant difference for 1,865 word pairs between Corpuses A and C. Between Corpuses A and B we found 2,184 statistically significant differences and 1,113 between stakeholder groups B and C[69].

[67] We used a special algorithm provided by Facebook in the library FastText. (https://github.com/facebookresearch/fastText), used in Python. P. Bojanowski, E. Grave, A. Joulin & T. Mikolov, *Enriching Word Vectors with Subword Information* (June 19, 2017) (unpublished manuscript) (on file at http://arxiv.org/abs/1607.04606) (accessed on Jan. 22, 2021).

[68] F. Khan et al., *Sentiment Analysis of Twitter Data*, in *International Journal Of Engineering Research* vol. 16, 15, 2018.

[69] Note that many of these words are not of particular interest for us as they might be specific to a position paper of a certain company (e.g. 'Gmail' in Google's submissions).

In the following, we will only discuss the most interesting differences we found, that are relevant to (1) key actors and structure of digital markets, (2) their anticompetitive conduct, and (3) the identified remedies and ex-ante rules. The results are summarized in Tables 1 to 3 respectively.

Summary of results

Table 1: ***Key actors and market structures***

Term	Distance AB	Distance BC	Distance AC	Close words A	Close words B	Close words C
Dominant	0.515 (0.65)	0.797 (0.50)	1.530 (0.02)**	Self-preferencing		Policymakers, quasi-judicial
Gatekeepers	1.064 (0.22)	1.367 (0.11)	1.486 (0.03)**	Content		Unsatisfactory
Monopolistic	1.023 (0.25)	1.473 (0.06)*	1.631 (0.00)**	Discouraging, profitability	Vulnerability, linking	Higher-cost, welfare
Monopolization*	1.323 (0.109)	1.432 (0.078)	1.609 (0.00)**	Endangers, non-dominant, data		Operations
Newcomers	1.469 (0.03)**	1.571 (0.04)**	1.192 (0.18)	Non-existent, none	Tech	Start-ups, destroyed

**"Monopolization" was used a relatively small number of times (only 285), compared to other reported words. We decided to include it to show the semantic distance with neighboring concepts.*

However, some of the key buzzwords surrounding competition law and new ex-ante remedies show statistically significant differences.

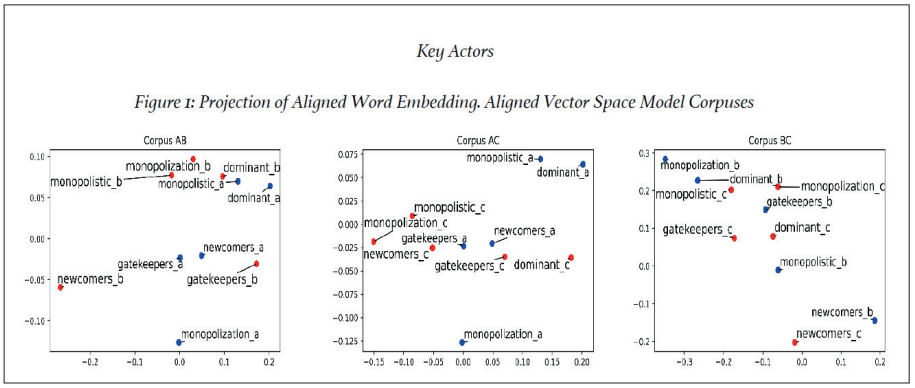
Table 2: ***Anticompetitive conduct***

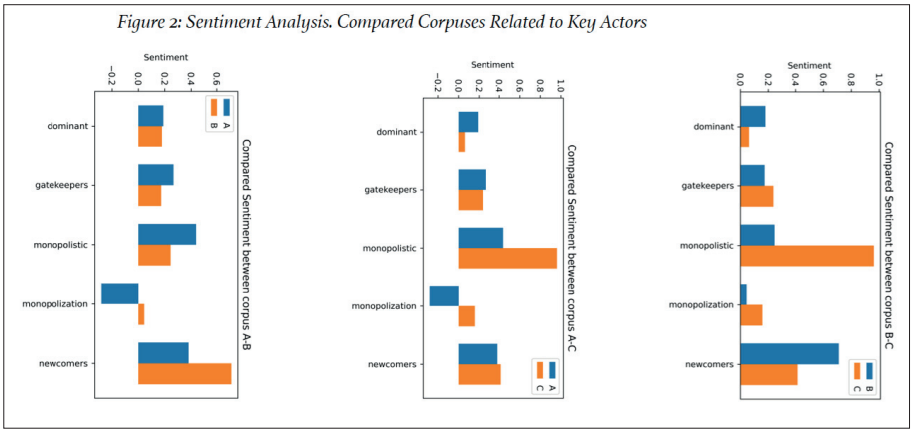
Term	Distance AB	Distance BC	Distance AC	Close words A	Close words B	Close words C
Abusive	1.216 (0.14)	1.162 (0.20)	1.543 (0.04)**	Competing		Tacit, monopoly
Collusion	1.590 (0.01)***	1.420 (0.08)*	1.396 (0.08)*	policing	tacit, data	
Coordinated	1.395 (0.06)*	1.110 (0.23)	1.541 (0.02)**	Cooperation, dialogue	Tacit, legitimate	Future-proof, EU-level
Pro-competitive	1.445 (0.04)**	1.295 (0.13)	1.022 (0.29)	comparison	user-friendly	
Ranking	1.182 (0.15)	1.644 (0.02)**	1.452 (0.05)**		guidelines, improve, oversight	appearance, disclosing
Self-favoring	.	.	1.589 (0.00)**	combating		debates
Self-preferencing	1.393 (0.04)**	1.457 (0.07)*	1.190 (0.18)	monopolizing	over-regulation, reports	
Tacit	1.439 (0.04)**	0.923 (0.40)	1.274 (0.13)		Threatens	
Tipping	1.509 (0.03)**	1.584 (0.03)**	1.213 (0.17)		bottleneck, nudge	unwanted
Uncompetitive	1.478 (0.03)**	1.199 (0.18)	1.499 (0.03)**	single-sided, state-run	EU-commission, institutions	

Table 3: ***Remedies and ex-ante antitrust reform***

Term	Distance AB	Distance BC	Distance AC	Close words A	Close words B	Close words C
Blacklist	1.402 (0.05)	1.063 (0.27)	1.580 (0.01)**	functionalization		Dominance-based, problem
Data-sharing	1.566 (0.02)**	.	.	recycled	differentiation	
Disproportionally	1.590 (0.01)**	1.044 (0.28)	1.154 (0.20)	Public-interest, bans	Ensure, competition, incentives	
Interoperability	1.665 (0.01)**	1.150 (0.21)	0.756 (0.49)	reliability, trustworthy	Licensing	
Overregulated	1.500 (0.03)**	.	.	Sellers	Tax-like	
Pro-competitive	1.445 (0.04)**	1.295 (0.13)	1.022 (0.29)	comparison	user-friendly	
Underenforcement	1.463 (0.03)**	1.283 (0.13)	1.513 (0.02)**	Complaints, unbureaucratic	Consensus	Misconceptions, improvements
Unregulated	1.361 (0.07)*	1.566 (0.04)**	1.822 (0.00)***	not-sufficient		mitigation
Welfare	0.989 (0.26)	1.543 (0.04)**	0.626 (0.65)		Economic, rights	mobility

Note: The "Distance" columns report the distance between the vectors of the same words for each corpus pair with the respective p-value in parentheses. A grey field indicates that a word was not used in both of the respective corpuses. The "Close Words" columns shine a light at some of the concepts that were closely related with the term in question in the corpuses for which there was a statistically significant distance between the terms. The asterisks indicate significance at a 0.001 (***), 0.05 (**), and 0.1 (*) level, respectively.





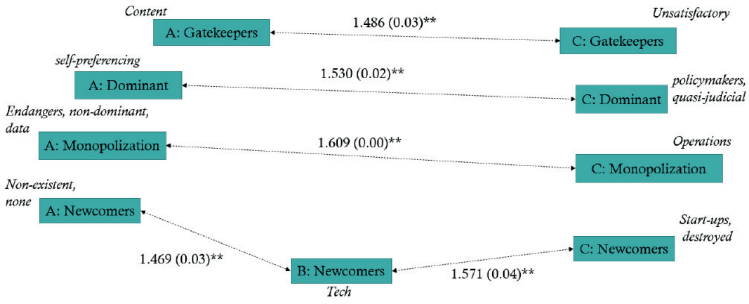
Starting with terms related to key actors, and especially the applicability of competition law and the DMA, we see a statistically significant difference between micro companies/organizations and the medium/big companies and organizations, for an absolutely central term: "gatekeepers." This can be linked back to the Commission's summary of the consultation, where it pointed out that some stakeholders considered the concept of a gatekeeper too broad.

Here, our computational analysis confirms the qualitative assessment of the Commission. The question of how "gatekeeper" should be defined was hotly debated[70] and it seems like the differing standpoints also translated into a different use of the term "gatekeeper" in the position papers for all three corpuses.

[70] See DSA, Explanatory Memorandum, 8.

Another term that has risen in importance in the recent debates on digital markets is "newcomers," the counterpart of "gatekeepers." We find that the term is used differently between both Corpus A and B as well as between Corpus B and C. Interestingly (Fig. 3), it is closely related with "non-existent" in Corpus A and "destroyed" in Corpus C, which seems to hint at the difficult standing of small tech start-ups on certain digital markets.

Figure 3. Distances and closest words for terms related to "key actors"



Remarkably, we also find a significant difference for the term "dominant," with Corpus A showing a close connection to the term "self-preferencing." This is noteworthy with a view to Art. 102 TFEU, which applies only to firms holding a "dominant" position. This finding suggests that that not only new competition concepts, but also established ones are perceived differently, in this case by microorganizations/individuals and medium/big organizations.

Unlike "dominant," EU competition law does not make use of the term "monopoly" or related concepts. However, they are frequently used in debates on competition law, including those on the DSA and DMA. Our analysis shows that there is no perfect agreement between different stakeholder groups concerning the precise content of terms like "monopolistic" or "monopolization." While the novelty of a term like "gatekeeper" might to some extent explain differences in its use, it is surprising that more established terms are used in an equally non-uniform way. This could be due to the fact that many respondents are non-Europeans and therefore not used to the EU competition jargon. However, as the close words analysis for "monopolization" reveals, this finding might also be related to the debates on how even non-dominant firms might gain a position that enables them to, e.g., effectively bar start-ups ("newcomers") from entering the market. The fact that also "data" is

closely related to "monopolization" could be interpreted as a hint of how new technologies can be a drive for economic giantism, while introducing uncertainty around relatively well-established terms.

Nevertheless, it needs to be noted that we did not find a significant difference for terms like "platform," "market power," or "dominance" which should be encouraging.

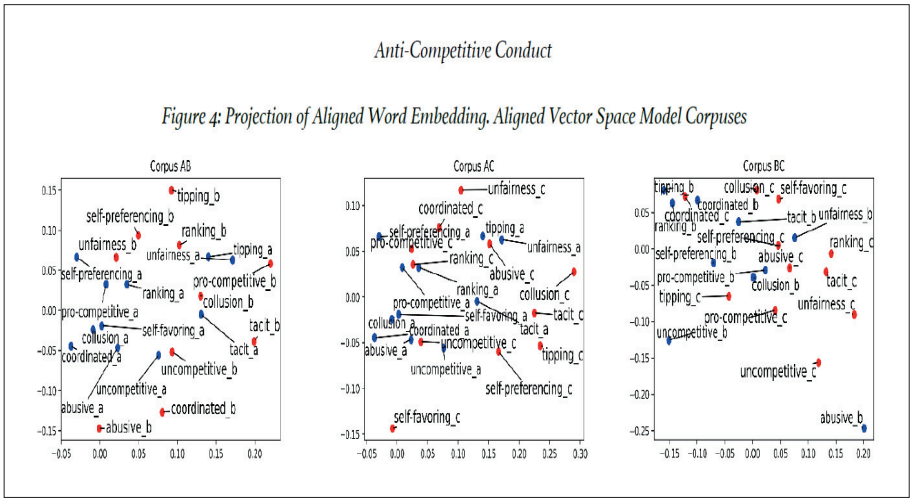


Figure 4: Projection of Aligned Word Embedding. Aligned Vector Space Model Corpuses

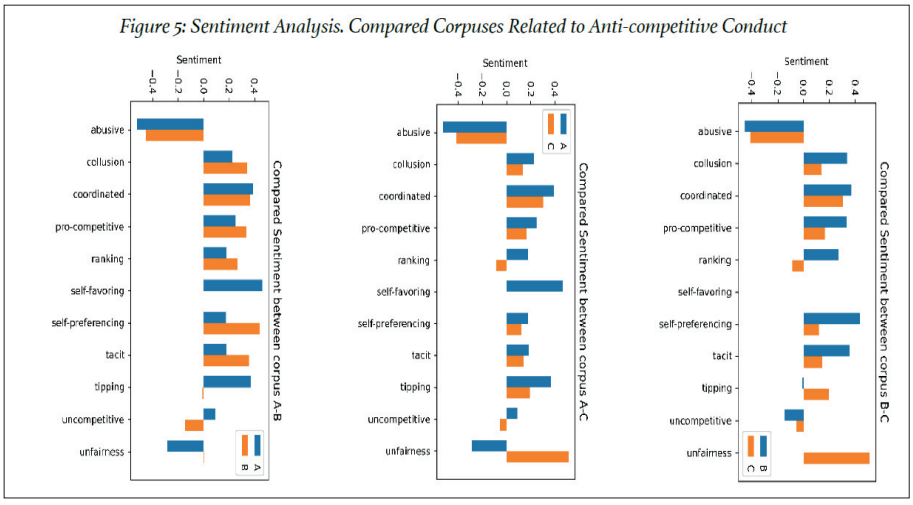


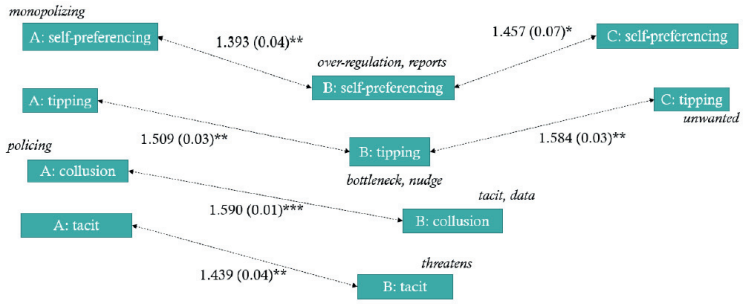
Figure 5: Sentiment Analysis. Compared Corpuses Related to Anti-competitive Conduct

Moving on to terms related to potentially anti-competitive conduct (Fig. 4), it is remarkable to see that general terms like "uncompetitive" and "pro-competitive" are not used in the same way. The same holds true for the more drastic, but also commonly used term "abusive."

Another interesting finding is that "pro-competitive" is closely associated with "user-friendly" for small companies and organizations (b), which could be interpreted as a hint that this group of stakeholders, in particular, perceive a strong link between consumer protection and friendliness (or fairness, in the dma and dsa jargon) and competition. Individuals and micro-organizations (a), on the other hand, connect "pro-competitive" with the term "comparison". This could emanate from the centrality of comparing offers to boost competition and innovation, and seems to confirm the importance of provisions like article of the 6(1)d dma[71] or article 29 dsa[72].

Strictly related to that, is the statistically significant distance we find between the vectors of "self-preferencing" for corpuses a and b as well as for b and c. For corpus pair ac, the alternative term "self-favoring" is used differently as well, suggesting that it is the general concept behind these two terms that seems to be still elusive. In corpus a, "self-preferencing" is most closely related to "monopolizing," (see fig. 6) indicating that the anti-competitive outcomes of selfpreferencing are a key concern for individuals and micro-organizations.

Figure 6: Distances and closest words for terms related to "anti-competitive practices"



[71] DMA art. 6 §1(d) deals with self-preferencing in the context of ranking services (a practice addressed in the Google Shopping case, *supra* note 36).
[72] *Supra* note 13.

Small companies, on the other hand, associate this practice with "overregulation" and "reports." This is interesting not only with a view to the prohibition on self-preferencing in Article 6(1)d of the DMA, but also in light of the comprehensive reporting duties on rankings in the DSA (Articles 12 and 29). Our results could be interpreted as a clue that small companies and organizations might fear comprehensive transparency and reporting duties, thus highlighting the need for a differentiated approach[73].

Another anti-competitive practice which yields a significant distance between corpuses A and B is "collusion." For small businesses and organizations (B), this keyword is closely related to "tacit" which well-reflects the many debates on "tacit collusion," driven both by the doctrine[74] as well as competition authorities[75], and lawmakers[76]. Note that the term "tacit" itself is used in a similarly idiosyncratic manner, which might reflect the debate on what exactly constitutes an agreement or otherwise unlawfully coordinated behavior. The term "coordinated" is also used differently by different stakeholders, although only small companies seem to connect it with tacit collusion. Interestingly, another term that shows up in the vicinity of "collusion" is "data," which is in line with the conclusion drawn by many experts regarding the central role of data availability as an enabling factor of tacit collusion[77].

It is also worth noting that no reference is made in the DMA to "data" or "algorithmic collusion," even though both locutions were referred to in the consultations. Given that other close relations exist with the words "harms" and "barred," we could deduce a rather negative attitude towards the emergence of collusion in corpus B. Generally, collusion is relevant with a view to existing competition law (esp. Art. 101 TFEU), but was not picked up in the DMA, despite the long debate.

A potential avenue for anti-competitive conduct of online platforms that is being addressed by both the DSA (Article 29) and the DMA (Article 6(d)) are "rankings," for which we find a statistically significant distance between corpuses A and C. A similar discrepancy was found for "tipping," which is an important concept related to the need for ex-ante rules. In the

[73] However, such a proportional approach is not easy to find, especially if there is no consensus on what "disproportionally" burdensome provisions look like. This seems to be the case at least for corpus pair AB.

[74] See e.g. A. Ezrachi & M. E. Stucke, *Sustainable and Unchallenged Algorithmic Tacit Collusion*, in *Northwestern J. Tech. & Intell. Prop.* 217, vol. 17, 217, 2020.

[75] See e.g. Bundeskartellamt & Autorité de la Concurrence (*supra* note 35).

[76] Crémer report, *supra* note 14, p. 68.

[77] See, for instance, Crémer report, Ibid, at 8.

small companies/organizations corpus, "tipping" is surrounded by the terms "bottleneck" and "nudge." This mirrors the debate around gatekeepers, which are often considered to gain their dominant or quasi-dominant role from their position as bottlenecks, nudging consumers into certain choices[78]. Interestingly, our sentiment analysis (Fig. 5) shows that "tipping" is viewed more favorably by medium and big companies/organizations C (0.338) than by small companies/organizations B (-0.296). The negative attitude of small companies and organizations towards "tipping" might reflect the potentially even higher costs of uncompetitive markets small businesses face, because they might often depend on uncontestable platform markets.

Figure 7: Projection of Aligned Word Embedding. Aligned Vector Space Model Corpuses

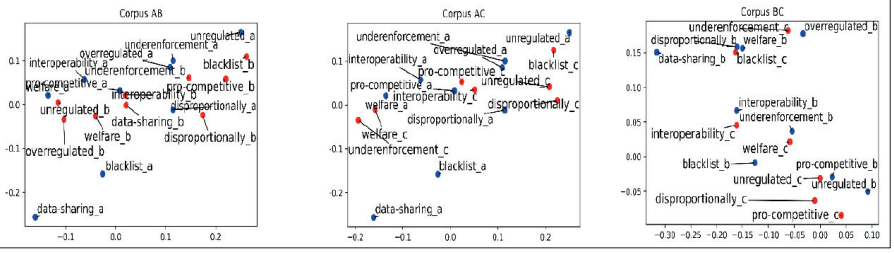
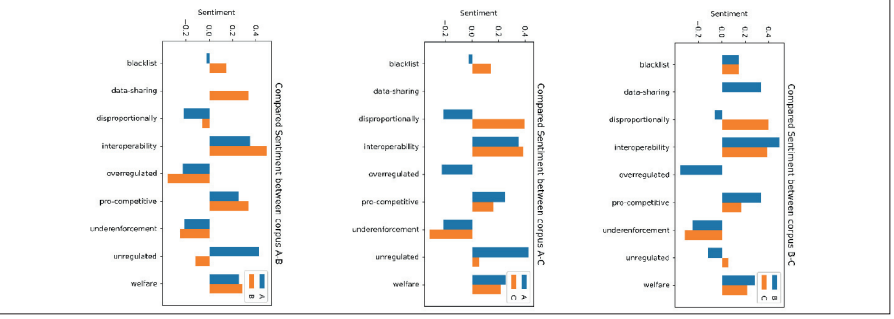


Figure 8: Sentiment Analysis. Compared Corpuses Related to Regulatory Reform and Remedies

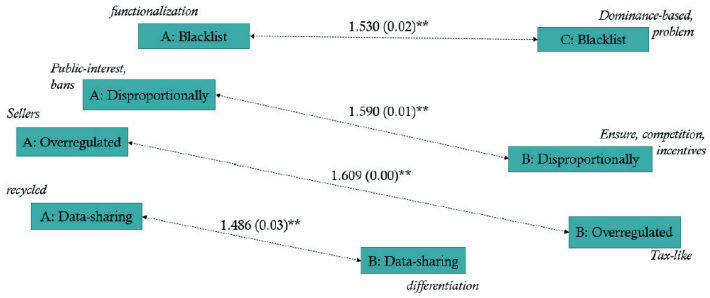


[78] There has even been empirical evidence for problematic nudges, so called "dark patterns." A. Mathur, G. Acar, M. J. Friedman, E. Lucherini, J. Mayer, M. Chetty & A. Narayanan, *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*, in *Proceedings Of The Acm On Human-Computer Interaction, (CSCW)*, vol. 3, 81. Concerning bottlenecks, the UK Furman report incorporated the concept in its 'significant market status' threshold, which is comparable to the "gatekeeper" designation in the DMA. Her Majesty's Treasury, Unlocking Digital Competition, Report Of The Digital Competition Expert Panel (2019), https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digitalcompetition-expert-panel (accessed on Feb. 25, 2021), p. 10.

Having discussed concepts related to competition challenges on digital markets, the logical next question is: What to do about it? Regarding the general level of regulatory intensity, there seems to be no perfect alignment on the meaning of "overregulated," with small organizations appearing to fear "tax-like" measures. At the same time, at least micro-organizations and individuals look to perceive a lack of regulation as "not sufficient." Generally, for the term "unregulated" we find significant differences between all corpuses. In a similar vein, we find a considerable distance between the vectors of "underenforcement" for Corpuses A and B as well as A and C.

Looking at how stakeholder groups associate words (Fig. 9), while microorganizations and individuals (A) seem to emphasize the importance of "unbureaucratic" procedures, medium and big organizations/companies (C) interestingly speak of "misconceptions" and "improvements." While the precise meaning of these associations remains obscure, it is interesting to note that not only individuals and micro-entities but also bigger organizations seem to demand some kind of reform.

Figure 9. Distances and closest words for terms related to "regulatory reform" and "remedies"



Regarding the concrete measures of reform proposed by the Commission, we find that stakeholders have different understandings of three key concepts (Fig. 7). First, the term "blacklist," which is highly relevant for the DMA[79], is not used in the same way by micro entities as by medium/big organizations. Interestingly, the former mentioned the concept far less than the latter (1,972 occurrences in Corpus A, compared to 36,161 in Corpus C).

[79] To be specific, DMA Article 5(1)(a), (d)-(f) and DMA Article 6(1)(a), (d)-(e) "blacklist" certain actions.

This disparity could be explained by a certain sense of alarm on the side of larger entities that is hinted by the associated term "problem." While this interpretation is highly tentative and would need to be confirmed by further analysis, the second closely associated term, "dominance-based" is a bit more telling and highly noteworthy. In fact, it might reflect the demand to set the threshold of application for new, stricter rules at the "dominance" threshold[80]. Linking this finding with our results for the terms "gatekeeper" (the approach chosen by the Commission) and "dominant," it seems reasonable to conclude that the disagreement on the applicability *ratione personae* of new rules is also reflected in the use of certain pivotal terms related thereto.

Furthermore, our sentiment analysis (Fig. 8 above) shows that "blacklists" are seen more favorably by small, medium, and large companies than by individuals/micro entities. This might explain some of significant differences we find in the use of the term. Second, moving from the "don'ts" of the blacklist part of the DMA to the positive obligations complementing these prohibitions, we find two noteworthy differences between Corpuses A and B. For instance, "interoperability" was used in a substantially different way, which might be a reflection of the debate on how far-reaching interoperability obligations should be[81]. Interestingly, the focus on "reliability" and trustworthiness in the closely associated words in Corpus A matches the spirit of Article 6(1)(c) of the DMA, which grants gatekeepers the opportunity to restrict interoperability to the extent necessary to ensure "the integrity of the hardware or operating system provided."

Lastly, we found significant differences in the use of the term "data-sharing." While individuals and micro-organizations (A) emphasize the possibility to "recycle" the data gatekeepers are obligated to share under, e.g., Article 5(1)(g) and Article 6(1)(g) of the DMA, small companies and organizations (B) heed the need for "differentiation." Potentially, this association is an expression of a fear of being overburdened with costly, potentially disproportional data-sharing duties.

However, what exactly "disproportionally" burdensome remedies look like seems to be a matter of debate for itself, given that the term shows a significant difference between corpuses A and B.

[80] See e.g., Cerre, DMA report, *supra* note 28, p. 52.

[81] Currently, Article 6(f) of the DMA only covers ancillary services, which many would like to see changed. S. Stolton, *EU SMEs in Bid for Greater Interoperability in Digital Markets Act*, EURACTIV, https://www.euractiv.com/section/digital/news/eu-smes-in-bid-for-greater-interoperability-in-digitalmarkets-act/ (accessed on May 13, 2021).

4. *Naming Is Taming? Drawing Legal Lessons From Computational Analyses*

Using computational tools to analyze the DSA and DMA consultation feedback documents allowed us to reach three results. First, we found that considering small and medium entities as a homogenous group of stakeholders might not always be recommendable, at least not for natural language data. In our computational analysis, it turned out that medium organizations are more comparable to big contributors than small ones. This highlights the need for data-driven procedures when determining the optimal units of analysis. As outlined below, there are several computational tools the Commission could use to respond to this need.

Second, our algorithmic analysis has shown that there are statistically significant differences between stakeholder groups in the use and understanding for some key concepts of competition policy. On the one hand, our analysis reproduced some of the results the Commission outlined in the summaries of the consultations. For instance, the differences in the terms "gatekeepers" or certain remedial strategies between different groups can be linked back to the debate on these concepts in the sense that they were not entirely uncontroversial[82].

On the other, we spotted substantial differences between stakeholders for terms the Commission perceived as uncontroversial. For example, the Commission concluded that "[t]he large majority of stakeholders believed that the proposed list of problematic practices, or "blacklist," should be targeted to clearly unfair and harmful practices of gatekeeper platforms"; while we found that the use and understanding of "blacklist" differs significantly among different feedback contributor groups. The same holds for some important anticompetitive practices, such as "self-preferencing."

The second finding of our computational analysis is that the consensus the Commission identified over the ex-ante measures proposed in the DSA and DMA might not be as unanimous as it seems at first sight: although stakeholders might say the same, they could mean different things.

These findings are relevant for two reasons. First, exposing "hidden misunderstandings" can enhance the quality of EU consultation processes. Since differences in understanding could lead to (undesirable) differences in implementation of those provisions that require stakeholders to act, it should be in the interest of the Commission to identify such differences before drawing conclusions from the consultations. Second, the findings of our analysis shine a new light on the scholarly debate about central terms

[82] DMA, Explanatory Memorandum, 7-8.

of the DSA and DMA. While theoretical discussions are without a doubt the pivotal starting point for any kind of reform, our analysis adds some empirical insights regarding the clarity of certain concepts, which might in turn inspire new theoretical arguments.

Yet, our analysis certainly also has some limitations. First, from a technical point of view, our corpus is rather small and heterogenous due to the great number of different feedback contributors. One way to help mitigate this shortcoming would be to enlarge the corpus, on the one hand, and applying techniques like bootstrapping[83], on the other. Second, we had to do some manual coding to select the most interesting terms for which we found significant distances. This part of the research process should be automatized or made more easily replicable in the future. Lastly, and most importantly: this analysis identified differences in understanding and use of some relevant terms, but cannot explain *why* they occur.

However, although it does not allow us to empirically prove a causal link between how clearly defined a term is, either by law or jurisprudence, and inconsistencies in its use, our results do suggest that such a connection exists.

One idea could be to run the same algorithms on new data, namely the submissions to the current consultations[84]. Now that the term 'gatekeeper' is defined using at least some quantitative criteria, it would be interesting to see whether stakeholders use the term more consistently than before the proposal was published, which is the consultation phase we looked at. If we found that 'gatekeepers' is used more consistently, while e.g., self-preferencing, which was not clearly defined is still used inconsistently, it would be a very strong sign that such quantitative or at least clear criteria help to avoid misunderstandings.

[83] T. Joseph, *Bootstrapping Statistics. What it is and why it's used*. https://towardsdatascience.com/bootstrapping-statistics-what-it-is-and-why-its-used-e2fa29577307) Jun 17, 2020 (accessed on Aug. 10, 2021).

[84] We refer to the public consultation on the Data Act opened by the EC on June 3, 2021 and currently undergoing (https://ec.europa.eu/eusurvey/runner/Data_Act#, accessed on Aug. 10, 2021) (covering many subjects and terms of the DSA and DMA, and addressing the same stakeholders. E.g. Section VI of the questionnaire explicitly refers to 'gatekeepers').

5. *Concluding Remarks*

This paper set out to explore whether different stakeholders share a similar understanding of the many new competition challenges coming with the increased importance of digital markets. To do so, we contrasted the debate spurred by the consultation on the three EC inception impact assessments with the corresponding norms in the DSA and DMA proposals. We then employed computational tools to gain a fine-grained understanding of the stakeholders' feedback documents.

Analyzing replies to the EU Commission's public consultation, we find significant differences in stakeholders' use of central terms of competition law like for instance "gatekeepers," "procompetitive," "collusion," and "self-preferencing."

While we believe that discerning latent differences in the use of certain terms competition law is a crucial capability that could significantly enhance the consultation process, both lawmakers and legal scholars could benefit even further from quantitative text analysis if the tool we present in this paper is complemented by other NLP techniques.

Hence, the tools presented here should be seen as only a first building block of a fully-fledged NLP toolbox. For instance, "topic modeling"[85] could be used to get an intuitive understanding of which topics are the most relevant to stakeholders. Another powerful tool to be utilized more systematically in the future is "sentiment analysis". While we employed this technique to investigate the general attitude of each group of contributors towards certain words of interest, we did so by using a pre-trained model. While this analysis produced some interesting results, one could use the same idea to cluster statements based on the sentiment of a group towards a certain concept or proposal to get a better understanding of how supporters and critics of a proposal are distributed and what their main concerns and arguments are[86].

If statements that are inputs to regulation are clustered based on "document similarity measures"[87], this could help to perceive certain

[85] See D. M. Blei, A. Y. Ng & M. I. Jordan, *Latent dirichlet allocation*, in *The Journal of Machine Learning Research,* 3, 993–1022, 2003.

[86] See e.g., S. Feng, D. Wang, G. Yu, C. Yang & N. Yang, *Sentiment Clustering: A Novel Method to Explore in the Blogosphere*, in Q. Li, L. Feng, J. Pei, S. X. Wang, X. Zhou, & Q.-M. Zhu (Eds.), *Advances in Data and Web Management. Springer*, 332–344, 2009.

[87] See, e.g., B. Krismono Triwijoyo, & K. Kartarina, *Analysis of Document Clustering based on Cosine Similarity and K-Main Algorithms*, in *Journal of Information Systems and Informatics*, 1(2), 164–177, 2019. G. Dannemann, *Comparative Law: Study of*

similarities or alliances between stakeholders, even across different groups like e.g., small companies and medium/large companies. As our analysis has shown, these clusters might not always look like what one would expect *prima facie*.

On a more legal ground, computational tools could be used to trace back the influence of certain stakeholders by identifying those statements which are the most similar to the rules the Commission decided to propose. This could allow to gain a precise understanding of why rules were drafted in a certain way and greatly help the interpretation of norms in light of their *telos* and their drafting history.

Consequently, computational analysis could uncover novel insights into the provenance of a provision: which stakeholders asked for it, where does it come from? Given that the analysis of the drafting history and objectives of a norm is an essential part of its exegesis, these insights are essential for legal scholars. And especially when it comes to proposals as complex as the DMA and DSA, they could be of great value both for the Commissions and legal scholars, and hence something to look forward to.

That might suggest both to further the inquiry, and to make an effort to spread a common understanding of relevant terms.

Appendix

Annex 1 – Results of preliminary data analysis to motivate corpus construction

Our clustering choice is based on two considerations: First, a qualitative analysis of the questionnaires accompanying the feedback documents[88] allowed us to get an understanding of which aggregation would cluster comparable feedback contributors together. Second, we conducted a quantitative analysis of the same questionnaires to ensure that our clusterization choices are solid. In particular, we sought to ensure that there is no statistically significant difference between *medium* and *large* entities in our sample since at least medium *companies* are often grouped

Similarities or Differences?, in M. Reimann and R. Zimmermann (eds.). *Oxford Handbook of Comparative Law* (2d ed.), Oxford: Oxford University Press, 390-422, 2019.

[88] For an archive of the feedback materials, see European Commission, Digital Services Act – Deepening the Internal Market and Clarifying Responsibilities For Digital Services, Public Consultation 11 January 2021, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12418-Digital-Services-Actpackage-ex-ante-regulatory-instrument-of-very-large-online-platforms-acting-as-gatekeepers/publicconsultation (accessed on Jan. 28, 2021).

with small, rather than large *companies*.

However, it needs to be noted that our feedback contributors are not only businesses, but also other types of organizations. This diversity could "smooth" the differences we would have expected to find if our sample included companies only.

In fact, our qualitative analysis of the questionnaires suggested that medium entities in our sample are more comparable to large businesses/organizations[89]. To test the robustness of this perception, we analyzed the answers provided by medium and large entities to specific multiple choices questions[90]. We applied a Kolmogorov-Smirnov two-sample test[91] to understand if there is a statistically significant discrepancy between the distribution of the answers of the two groups.

If that was the case, we would assume that theses answers must be considered as provided by two different populations, not allowing us to treat them as a unique cluster. The results of the test are shown in Figure 1:

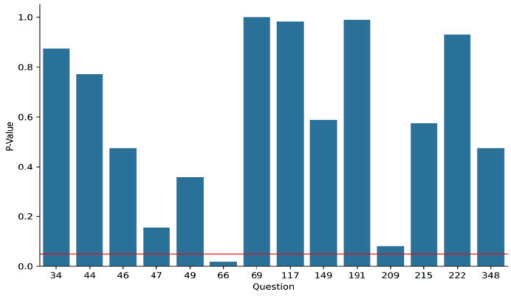


Figure 1: p-value distribution associated with the KS-two sample test applied to the answer distributions of the considered questions. Red line highlights our significative tolerance value of p=0.05

Even using a very high tolerance p-value level of 0.05, only question no. 66[92] showed a statistically significant variation. This question alone

[89] While this could be due to the idiosyncrasy of our sample, this finding also corresponds with scholarly literature. See e.g. R. Kemp & C. Lutz, *Perceived Barriers to Entry: Are There Any Differences Between Small, Medium-sized and Large Companies*, in *International Journal Of Entrepreneurship And Small Business*, vol. 3, 538, 2006.

[90] We selected said questions based on what could be considered interesting for our research. A full list of the questions we selected can be found in the appendix.

[91] J. Lawson Hodges Jr, *The Significance Probability of the Smirnov Two-sample Test*, in *Matematica*, vol. 3, 469, 1958.

[92] See Annex 1.

however is mostly unrelated to our core research interest, and hence unlikely to compromise the validity of our clustering.

Annex 2 – A statistical test to identify semantic differences

We can model the relative distance d_t^{AB} of a word t in the corpus A and B be as:

$$d_t^{AB} = \gamma_t^{AB} + \mu_t^{AB} + u_t^{AB}$$

This takes into account a semantical term Y_t^{AB}, a non-semantical term (originated from the simple different words disposition in the two corpus) and a random term u_t^{AB}.

To isolate the semantic difference in the distances between words we found, we need to set two assumptions. Our first assumption is that words in the control vocabulary used for the Vector Space Alignment Transformation do not have a semantic difference, i.e., $\gamma_t^{AB} = 0$. This means we assume that stakeholders mean the same when they use words like "and" or "one."

Consequently, based on the distances between these "control" words, we can construct an empirical distribution of the non-semantic distance between words. In this manner, we can get an idea of what a distance would look like if there was no semantic difference.

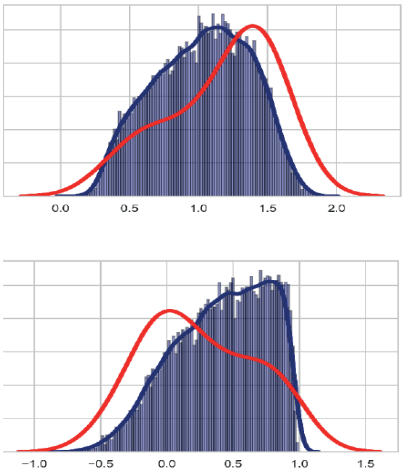
In a next step, we can hence compare the distance we observe for certain words with what we would expect if there was no semantic difference. Our p-value represents the probability to have a distance equal or greater than what we observe if our null-hypothesis, that there is no semantic difference between the same term in different corpuses, is actually true. If this probability is small enough, we can reject this null hypothesis with a small possibility of error. This is to say that the particular word has, indeed, a statistically significant semantic difference in the two corpuses. A general acceptance value for the p-value is 0.05, which we will use as the critical threshold for our analysis.

Annex 3 – Cumulative distribution of control distances

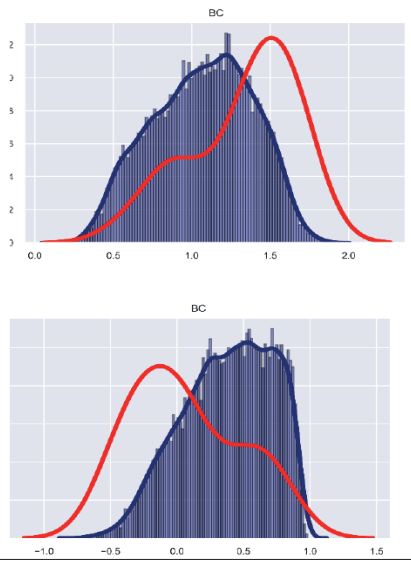
Figures 1 to 3 show the Cumulative Distribution of distances of control dictionary words (in blue) against cumulative distribution of distances of

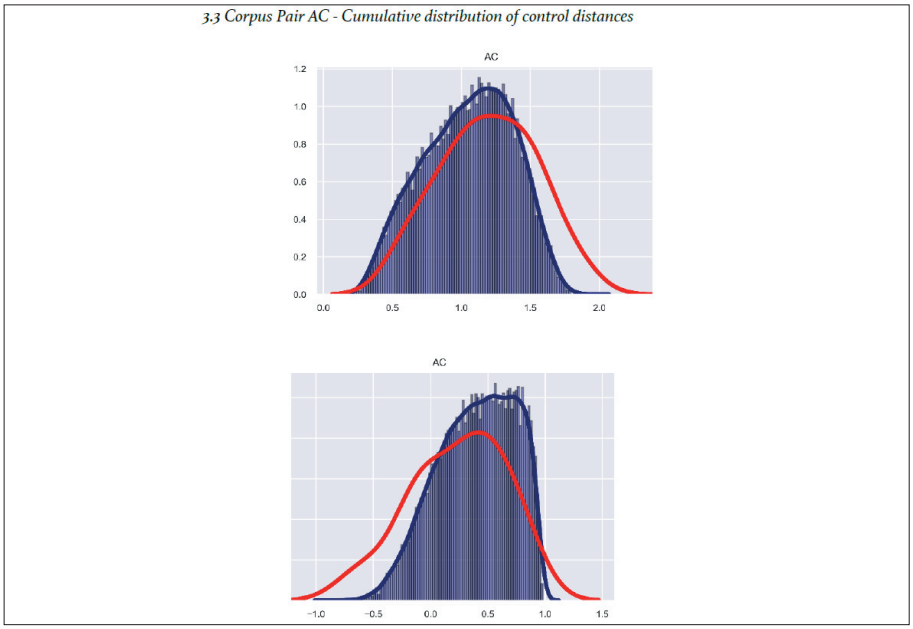
analyzed words (in red) for each corpus pair (i.e. corpus X against corpus Y).

3.1 Corpuses AB - Cumulative distribution of control distances



3.2 Corpuses BC - Cumulative distribution of control distances





Annex 4 – Questions manually selected from the DSA questionnaire[93] *to run the Kolmogorov-Smirnov two-sample test*

Question 34: Did you ever come across illegal content online (for example illegal incitement to violence, hatred or discrimination on any protected grounds such as race, ethnicity, gender or sexual orientation; child sexual abuse material; terrorist propaganda; defamation; content infringing intellectual property rights, consumer law infringements)?

Question 44: Do you consider these measures([94]) appropriate?

Question 46: If your content or offering of goods and services was ever removed or blocked from an online platform, were you informed by the platform?

Question 47: Were you able to follow-up on the information?

Question 49: If you provided a notice to a digital service asking for the removal or disabling of access to such content or offering of goods or services, were you informed about the follow-up to the request?

[93] Note that the questions have been re-enumerated consecutively. The content remains the same.

[94] Question 43: 'What actions do online platforms take to minimize risks for consumers to be exposed to scams and other unfair practices (e.g. misleading advertising, exhortation to purchase made to children)?'.

Question 66: Does your organisation access any data or information from online platforms?

Question 69: Do you use WHOIS information about the registration of domain names and related information?

Question 117: What information would be, in your view, necessary and sufficient for users and third parties to send to an online platform in order to notify an illegal activity (sales of illegal goods, offering of services or sharing illegal content) conducted by a user of the service?

a) Precise location: e.g., URL
b) Precise reason why the activity is considered illegal
c) Description of the activity
d) Identity of the person or organisation sending the notification. Please explain under what conditions such information is necessary
e) Other, please specify

Question 149: In your view, is there a need for enhanced data sharing between online platforms and authorities, within the boundaries set by the General Data Protection Regulation? Please select the appropriate situations, in your view:

Question 191: Do you believe that in order to address any negative societal and economic effects of the gatekeeper role that large online platform companies exercise over whole platform ecosystems, there is a need to consider dedicated regulatory rules?

Question 209: Which, if any, of the following characteristics are relevant when considering the requirements for a potential regulatory authority overseeing the large online platform companies with the gatekeeper role:

a) Institutional cooperation with other authorities addressing related sectors –e.g., competition authorities, data protection authorities, financial services authorities, consumer protection authorities, cyber security, etc.
b) Pan-EU scope
c) Swift and effective cross-border cooperation and assistance across Member States
d) Capacity building within Member States
e) High level of technical capabilities including data processing, auditing capacities
f) Cooperation with extra-EU jurisdictions
g) Other

Question 215: Taking into consideration the parallel consultation on a proposal for a New Competition Tool focusing on addressing structural competition problems that prevent markets from functioning properly and tilt the level playing field in favor of only a few market players. Please rate the suitability of each option below to address market issues arising in online platforms ecosystems. Please rate the policy options below from 1 (not effective) to 5 (most effective).:

a) Current competition rules are enough to address issues raised in digital markets
b) There is a need for an additional regulatory framework Imposing obligations and prohibitions that are generally applicable to all large online platforms with gatekeeper power
c) There is a need for an additional regulatory framework allowing for the possibility to impose tailored remedies on individual large online platforms with gatekeeper power, on a case-by-case basis
d) There is a need for a New Competition Tool allowing to address structural risks and lack of competition in (digital) markets on a case-by-case basis.
e) There is a need for combination of two or more of the options 2 to 4.

Question 222: When you see an online ad, is it clear to you who has placed the advertisement online?

Question 348: In your view, is there a need to ensure similar supervision of digital services established outside of the EU that provide their services to EU users?

Giulia Ferrari, Mariateresa Maggiolino

GAFAM's power across markets: how should we deal with it?

Abstract: Large technology companies, such as Google, Amazon, Facebook, Apple and Microsoft, enjoy various competitive advantages that allow them to form ecosystems. Antitrust law does not have the appropriate tools to manage the power that these companies exercise "across markets". However, European and German legislators have just prepared new solutions to face this phenomenon. This paper compares and briefly discusses them.

Abstract: Grandi imprese tecnologiche, come Google, Amazon, Facebook, Apple e Microsoft, godono di diversi vantaggi competitivi che consentono loro di formare ecosistemi. Il diritto antitrust non dispone degli strumenti più appropriati per gestire il potere che queste imprese vengono così ad esercitare "across markets". Tuttavia, i legislatori europeo e tedesco hanno appena predisposto nuove soluzioni per far fronte a questo fenomeno. Il presente lavoro le confronta e discute brevemente.

1. *Premessa*

Nell'era della quarta rivoluzione industriale, il dibattito in tema di concorrenza è stato sempre più spesso animato da una preoccupazione: che in un prossimo futuro alcune imprese, come Amazon, Google, Facebook, Apple o Microsoft, arrivino a dominare tutti i settori dell'economia; ed in effetti, nel corso degli ultimi anni, le c.d. GAFAM hanno per lo meno dimostrato di godere del dono dell'ubiquità, entrando in tanti e diversi mercati, ancorché tutti merceologicamente distanti dalla propria primigenia area di elezione, ossia – in termini più manageriali – dal proprio *core business*. Si consideri, ad esempio, che ad oggi Amazon può rendere disponibili ai propri clienti tanto strumenti di pagamento e mezzi di finanziamento quanto forme di intelligenza artificiale, sebbene questa

* Questo articolo è stato originariamente pubblicato in *Orizzonti del diritto commerciale*, 2021, 463-488. Il presente contributo è stato scritto prima dell'approvazione del Digital Markets Act (DMA).

piattaforma sia nata quantomeno intenzionalmente "giusto" per fare da intermediario nella vendita e distribuzione di beni e servizi.

A fronte di un possibile dominio "*across markets*" delle suddette imprese[1], chi studia il diritto antitrust è stato sollecitato a individuare delle vie – anche originali – per aggredire tali giganti[2], quasi si desse per scontato che i loro ecosistemi siano il frutto di condotte anticoncorrenziali o – ipotesi forse più veridica – quasi si dovesse impedire l'avverarsi del pronosticato scenario, di là dalla eventualità che esso risulti o meno dal connubio tra il naturale operare delle forze di mercato e le caratteristiche strutturali del mercato medesimo[3].

Detto ancora più esplicitamente, diffusa è la sensazione che il caso di poche imprese tra loro concorrenti in tanti e diversi mercati debba essere scongiurato "*whatever it takes*", ossia anche qualora il successo delle GAFAM fosse decretato da consumatori avveduti e informati che, consapevolmente, preferiscono la proposta commerciale delle società statunitensi a quella delle altre imprese[4].

Ebbene, questa volontà di intervento "ad ogni costo" – che evidentemente rivela una granitica (!) fiducia nelle capacità dei rivali delle GAFAM di produrre beni e servizi migliori rispetto a quelli resi disponibili da questi giganti – ha sulle prime alimentato e nutrito le più disparate discussioni: da quella, invero consueta, sugli obiettivi della disciplina antitrust[5] a quella,

[1] N. Petit, *Tech Giants, Competition And Innovation: The Seen And The Unseen*, 2020.

[2] M. Cappai, G. Colangelo, *Navigating the Platform Age: the 'More Regulatory Approach'to Antitrust Law in the EU and the U.S.*, Stanford-Vienna TTLF Working Paper No.55, 10 aprile 2020, consultabile all'indirizzo https://ssrn.com/abstract=3572629.

[3] I principali fenomeni digitali ed i loro impatto sui tradizionali concetti dell'analisi e della normativa a tutela della concorrenza sono stati approfonditi nell'ambito di autorevoli report e studi di recente pubblicazione; tra questi si veda a titolo di esempio J. Crémer, Y.A. De Montjoye, H. Schweitzer, *Competition Policy for the Digital Era*, Report finale per la Commissione Europea, 2019; J. Furman *et alii*, *Unlocking digital competition*, *Report of the Digital Competition Expert Panel*, 2019; Autorité de la Concurrence, Contribution de l'Autorité de la concurrence au débat sur la politique de concurrence et les enjeux numériques, 2020; Autorità Garante della Concorrenza e del Mercato *et alii*, *Indagine conoscitiva sui big data*, 2020.

[4] Si permetta di rinviare a G. Colangelo, M. Maggiolino, *From fragile to smart consumers:Shifting paradigm for the digital era*, in *Computer Law & Security Review*, 2019, vol. 35, 173 ss.

[5] A. Ezrachi, *EU Competition Law Goals and the Digital Economy*, Oxford Legal Studies Research Paper No. 17/2018, 6 giugno 2018 consultabile all'indirizzo https://ssrn.com/abstract=3191766 e D. Sokol, *Antitrust's Curse of Bigness Problem*, University of Florida Levin College of Law Research Paper No. 20-12, 15 marzo 2020, in *Michigan Law Review*, 2020, vol. 118, in corso di pubblicazione, consultabile all'indirizzo https://ssrn.com/

più di nicchia, in merito alle critiche alla direttiva denominata PSD2, la quale avrebbe avuto il demerito di favorire proprio le big tech a danno delle banche, sebbene alcune tra queste ultime siano certo dominanti nei relativi mercati[6].

In mesi più recenti, invece, in sede tedesca ed europea si sono elaborate delle prime risposte, rispettivamente di matrice antitrust e regolatoria, per provare a governare proprio il caso di queste imprese che operano su diversi mercati, costituendo ecosistemi digitali. In questa sede si offrirà una breve – anzi brevissima – descrizione delle nuove norme che, per lo meno, hanno il merito di aver sgombrato il tavolo dall'eventualità che le GAFAM siano indiscriminatamente tenute a condividere con terzi, concorrenti e non, i dati e le altre risorse – prima tra tutte, la tecnologia – in cui si radicano i loro vantaggi competitivi.

2. *I tratti distintivi della quarta rivoluzione industriale e la produzione di valore: perché diamo così tanta importanza ai* big data

Nel corso degli ultimi cinque anni si è spesso evidenziato come l'avvento delle tecnologie digitali e il potenziamento delle capacità dei computer abbiano permesso alle imprese di acquisire nuovi strumenti per rappresentare e conoscere ciò che li circonda, nonché per riprodurre i percorsi decisionali degli individui.

Sotto il profilo della rappresentazione, ad oggi, molti dispositivi consentono di tradurre i comportamenti umani e i fatti del mondo in stringhe di codice binario. Ad esempio, non solo le imprese che si sono dotate di una infrastruttura IT possono godere di una riproduzione digitale di tutte le loro interazioni e nel caso in cui operino on line riescono a disporre di una traccia digitale di tutte le condotte dei propri clienti che navigano in Internet, ma – ben al di là di quanto forse si sarebbe portati a credere – persino le imprese che operano offline possono raggiungere il medesimo risultato, servendosi di oggetti intelligenti, ossia di apparecchi capaci di trasformare quanto accade intorno a loro in sequenze di uno e zero. Insomma, la quarta rivoluzione industriale ha innanzitutto consentito di procedere con la c.d. datificazione del reale, ossia ha permesso di

abstract=3554728.

[6] O. Borgogno, G. Colangelo, *The data sharing paradox: BigTechs in finance*, 2 maggio 2020, consultabile all'indirizzo https://ssrn.com/abstract=.

generare una quantità quasi incommensurabile di dati che altro non sono che le registrazioni, in formato digitale, dei fatti e dei comportamenti che si consumano nella realtà.

Sotto il profilo della conoscenza, poi, la possibilità di dare una veste unitaria – le stringhe di uno e zero – alle appena descritte osservazioni, qualsiasi ne sia l'oggetto (dalla temperatura di una stanza all'età di un Internet-nauta), aiuta ad accomunarle tutte in un unico "calderone", per poi sviluppare una modalità, altrettanto omogenea, di processarle. Più chiaramente, la circostanza che vuole che i più disparati contenuti si traducano tutti in sequenze di codice binario permette, per un verso, di mescolarli, assemblarli e combinarli tra loro dando luogo ai c.d. *big data* e, per altro verso, di assoggettarli tutti a un sistema unitario di analisi: le c.d. tecniche di *big data analytics*. Queste ultime costituiscono un insieme di software capaci di estrarre significati dalle menzionate stringhe individuando anche (ma non solo) correlazioni altrimenti nascoste tra i fatti, apparentemente tra loro indipendenti, che quelle stringhe rappresentano.

Le tecnologie che caratterizzano la quarta rivoluzione industriale, grazie alla velocità dei computer ed alle loro potenziate capacità di immagazzinamento e calcolo, hanno quindi reso possibile ampliare il dominio della conoscenza esplicita, facendo emergere anche informazioni che altrimenti sarebbero rimaste nascoste.

Infine, in tema di percorsi mentali, bisogna dare conto di come le imprese stiano oggi investendo non poche risorse per sviluppare "cervelli" che, sebbene non umani, riescano a porre in essere i descritti processi di rappresentazione e conoscenza. Nel corso degli ultimi anni, l'avvento delle tecnologie digitali e il potenziamento delle capacità dei computer hanno permesso di sviluppa re la c.d. intelligenza artificiale, vale a dire un insieme di programmi per computer capaci, in virtù della sequenza di ordini e comandi in cui si articolano (i c.d. algoritmi), di simulare l'intelligenza umana per offrire risposte a domande specifiche, individuare correlazioni tra fenomeni, formulare predizioni nonché, nel tempo, migliorare l'esecuzione di queste e altre attività connesse alla conoscenza e comprensione del mondo. E giacché – come tutti noi sappiamo – l'intelligenza umana procede anche elaborando giudizi informati in risposta agli interrogativi più disparati, cogliendo somiglianze e differenze, riconoscendo modelli di comportamento e ricorrenze, nonché immaginando scenari futuri, non deve stupire se, passando in rassegna alcune delle più famose tecnologie di intelligenza artificiale, ci si imbatte in: programmi per l'elaborazione di inferenze logiche e alberi decisionali, evidentemente utili alla formulazione

di risposte e predizioni; software di apprendimento automatico, affinché le macchine imparino a modificare le proprie condotte in rapporto alla mutevole realtà che circonda loro; sensori e programmi per la percezione e l'esecuzione del movimento, nonché programmi di riconoscimento del linguaggio, delle immagini e delle voci, tutti ugualmente funzionali a che i computer sappiano da sé "osservare" la realtà, distinguendo gli oggetti che la compongono, anche qualora si trovino in testi scritti[7].

Certo, anche per mitigare l'impatto della quarta rivoluzione industriale, si potrebbe argomentare che l'essere umano, percependo i fatti del mondo attraverso i propri sensi e i propri strumenti, da sempre produce osservazioni e registrazioni – in una parola, dati – che poi utilizza per formarsi un'idea di ciò che lo circonda, per rispondere cioè a domande come "cosa?", "in che modo", "perché?" e dunque per inferire informazioni da quei dati. Analogamente, essendo gli esseri umani muniti di intelligenza, si potrebbe contestare l'utilità di riprodurre artificialmente dei percorsi cognitivi che già si danno in natura. Rispetto al passato, però, l'avvento delle tecnologie digitali e il potenziamento delle capacità dei computer hanno permesso di disporre non solo di un numero infinitamente più grande di registrazioni del reale, ma anche di applicare tecniche assai più sofisticate e potenti per analizzarle, così da estrarne informazioni sempre più precise ed accurate. Analogamente, in un prossimo futuro, un massiccio ricorso all'intelligenza artificiale potrebbe consentire di raggiungere obiettivi assai ambiziosi in tempi e con un grado di efficienza non replicabili dagli esseri umani.

Ora – per quanto qui più interessa – occorre evidenziare come, allo stato, le imprese sfruttino le possibilità e opportunità tecnologiche offerte loro dalla quarta rivoluzione industriale utilizzando i big data come input di diversi processi industriali volti a produrre valore.

In primo luogo, le imprese possono utilizzare le informazioni inferite dai dati raccolti come base di lavoro per programmi di intelligenza artificiale[8],

[7] L'intelligenza artificiale è proliferata nel mercato dei prodotti di consumo con la crescita della c.d. «Internet of Things» («IoT»). I dispositivi IoT sono prodotti di consumo «intelligenti » che incorporano l'intelligenza artificiale e la connettività Internet nelle loro funzionalità. Esempi di IoT includono i prodotti Echo di Amazon con Alexa, i prodotti Apple con Siri, ed i dispositivi «smart home» per il controllo di funzioni come temperatura, illuminazione e sicurezza domestica.

[8] Cfr. Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *L'intelligenza artificiale per l'Europa*, COM(2018) 237 final, Bruxelles, 25 aprile 2018. Qui la Commissione evidenzia come la disponibilità di ingenti quantità di dati rappresenti una precondizione necessaria per lo sviluppo dell'intelligenza artificiale: «[s]ono necessari ingenti volumi di dati per sviluppare l'IA. L'apprendimento automatico, un tipo di IA, opera mediante l'individuazione

ossia per sviluppare degli algoritmi decisionali che successivamente impiegheranno o "al proprio interno" – per individuare le scelte da adottare con riferimento a loro stesse, come ad esempio accade nel caso dei c.d. robo-managers e dei robo-directors – oppure "all'interno di servizi rivolti a terzi". In quest'ultima circostanza, ovvero nel caso dei c.d. servizi di robo-advisors, gli algoritmi sviluppati sulla scorta dei dati raccolti e che lavorano su quei dati servono infatti a suggerire, nonché a gestire, i portafogli di strumenti finanziari e patrimoniali in cui i clienti potrebbero investire in ragione delle proprie preferenze[9].

Inoltre, le imprese possono scegliere di vendere pacchetti di dati e informazioni insieme a software capaci di processarli per ricavarne ulteriori informazioni ad alto contenuto aggiunto, nonché insieme a servizi addizionali che aiutino i clienti a comprendere e valutare dette informazioni. Nel mercato dei dati finanziari, ad esempio, con lo scopo di offrire ai più disparati operatori (dai promotori finanziari a coloro che offrono servizi fiduciari e di custodia) strumenti personalizzati di analisi, sono nati non pochi provider indipendenti di dati che possono avere ad oggetto azioni, obbligazioni, valute, indici, derivati, o fondi.

In aggiunta, le imprese possono utilizzare le informazioni estratte dai big data per meglio conoscere o i propri concorrenti – e allora si porrà un profilo di market intelligence che potrebbe sconfinare in un'ipotesi di collusione algoritmica – oppure i propri clienti, come nel caso della loro profilazione finalizzata all'offerta di beni e prezzi personalizzati anche tramite quell'attività di c.d. *targeted advertisement* ad oggi divenuta imprescindibile per le attività che ambiscono ad operare on-line. In ambito finanziario, ad esempio, sebbene i prodotti e servizi da offrire continuino ad essere descritti sulla scorta della propensione al rischio e della solvibilità del cliente, queste caratteristiche vengono ad oggi individuate con un maggiore grado di accuratezza e dettaglio proprio per il livello di granularità dei dati analizzati e per la potenza degli strumenti computazionali impiegati a tal senso.

di modelli a partire dai dati disponibili e la successiva applicazione di questa conoscenza ai dati nuovi. Quanto più è grande il set di dati, tanto più accurata sarà l'individuazione delle relazioni anche impercettibili tra i dati. Quando si tratta di utilizzare l'IA, gli ambienti ad alto contenuto di dati offrono anche le maggiori opportunità, perché i dati sono il mezzo attraverso il quale l'algoritmo apprende e interagisce con il suo ambiente. Per esempio, se tutte le macchine e i processi in uno stabilimento producono continuamente dati, è probabile che con l'aiuto dell'IA si possano realizzare ulteriore automazione e ottimizzazione. In un contesto analogico, per esempio in un lavoro basato su documenti cartacei senza dati digitalizzati sulle operazioni in corso, tale automazione non è possibile».

[9] AA.VV., *La digitalizzazione della consulenza in materia di investimenti finanziari*, in CONSOB, *Quaderni Fintech*, n. 3, gennaio 2019.

Infine, le imprese possono diffondere nel mercato l'informazione stessa che inferiscono dai dati sotto forma di classifiche, recensioni e di risultati dei motori di ricerca, delineando la possibilità che si consumino ipotesi di manipolazione dell'informazione così ottenuta e resa disponibile al mercato.

Pertanto, è guardando a come l'informazione inferita dai big data viene utilizzata nella prassi commerciale che nasce spontaneo concettualizzare i big data come una risorsa indispensabile allo sviluppo delle attività imprenditoriali che aspirino a trovare uno spazio nell'economia digitale, nell'economia della quarta rivoluzione industriale. Meglio, è alla luce di quanto sinora osservato che si è indotti ad associare i big data ad una generica idea di potere da intendersi come la capacità delle imprese di utilizzare queste risorse per intervenire sul reale e avvantaggiarsi del valore così prodotto.

Tuttavia, non è detto che tale capacità sia riconducibile a quella nozione di "potere di mercato" con cui il diritto antitrust è solito confrontarsi; ecco perché nelle prossime pagine occorrerà svolgere qualche rapida – anzi, rapidissima – considerazione sulle forme di potere che possono accostarsi ai big data.

3. *Alcune possibili concettualizzazioni della relazione tra i* big data *e il potere*

La relazione tra i big data e il potere può essere concettualizzata da diversi punti di vista. Dapprima, si può considerare come i big data possano essere utilizzati dalle imprese per manipolare le preferenze dei consumatori, ossia per fare in modo che questi ultimi vengano falsamente o surrettiziamente indotti a preferire delle imprese invece di altre, in ragione dell'ordine secondo cui sono proposti i risultati dei motori di ricerca, oppure sulla scorta delle recensioni diffuse e/o delle classifiche elaborate[10]. Tuttavia, come altrove argomentato[11] e come il celebre caso *Google Shopping* dimostra[12], il diritto della concorrenza non può perseguire

[10] T. Zarsky, *Online privacy, tailoring and persuasion*, in K.J. Standburg, D.S. Raicu, *Privacy and technologies of identity*, Springer, 2006, 209 e A. Quentin *et alii*, *Consumer Choice and Autonomy in the Age of Artificial Intelligence and Big Data*, in *Customer Needs and Solutions*, 2018, 28.

[11] Sul punto si permetta di rinviare a M. Colangelo, M. Maggiolino, *La manipolazione dell'informazione come illecito antitrust*, in *Riv. dir. comm.*, 2019, 159.

[12] CE, 27 giugno 2017, caso AT.39740.

la manipolazione dell'informazione in quanto tale, a meno che essa non integri gli estremi delle fattispecie antitrust, risultando ad esempio l'oggetto di un'intesa, come nel caso *Hoffman La Roche Novartis*[13] o l'effetto di una condotta unilaterale capace di escludere i concorrenti e di far aumentare i prezzi di mercato. Del resto, esiste pur sempre la disciplina a tutela dei consumatori, la quale potrebbe ben giocare un ruolo nel fronteggiare attività illecite tese a privare acquirenti di beni e servizi della capacità di autodeterminarsi[14].

In aggiunta, sempre altrove si è già osservato come i big data non siano fonte di potere di mercato[15], semplicemente perché il loro utilizzo non consente ad un'impresa di peggiorare la sua offerta, pur mantenendo elevati i propri livelli di profitto. Al contrario, le informazioni inferite dai big data permettono alle imprese di migliorare i beni e i servizi che offrono ai consumatori, modificando questi prodotti in modo da incontrare maggiormente le loro preferenze attuali e potenziali, presenti e future.

Ebbene, questa terza forma di potere può ben essere catturata dal diritto antitrust, qualificandola come una barriera all'ingresso, ossia argomentando che un'impresa che volesse competere con chi detiene dei big data dovrebbe essere altrettanto capace di ampliare la propria conoscenza e, così, migliorare la propria offerta, perfezionando i propri prodotti e/o servizi in termini di prezzi, qualità, o livello di innovazione. Si è intenzionalmente detto "catturata" e non "perseguita" perché, in generale e a differenza di altre specie di intervento dei pubblici poteri nel mercato, il diritto antitrust non agisce per abbattere le barriere all'ingresso, ma prende atto dell'esistenza delle stesse per, al più, apprezzare la potenziale durata del potere di mercato delle imprese protette da quelle barriere. In particolare, poi, finché ad essere analizzate sono condotte che, come la raccolta e l'accumulazione dei dati, determinano comunque un miglioramento dell'offerta, non esiste modo per considerare gli effetti escludenti propri di ogni barriera come preponderanti rispetto all'aumentato soddisfacimento

[13]AGCM, Provvedimento n. 24823 del 27 febbraio 2014, I760 – *Roche-Novartis/Farmaci Avastin E Lucentis*. Il caso è stato oggetto di ricorso al TAR ed al Consiglio di Stato e di un rinvio pregiudiziale alla Corte di Giustizia su iniziativa di quest'ultimo; cfr. TAR Lazio, sentenza n. 12168/2014; Cons. St., sentenza n. 4990/2019; Corte Giust., sentenza 23 gennaio 2018, C-179/16.

[14] A. Jabłonowska *et alii*, *Consumer Law and Artificial Intelligence: Challenges to the EU Consumer Law and Policy Stemming from the Business'Use of Artificial Intelligence – Final report of the ARTSY project*, 2018, EUI Department of Law Research Paper No. 2018/11, disponibile all'indirizzo https://ssrn.com/abstract=3228051.

[15] In tema sia consentito richiamare quanto scritto in M. Maggiolino, *I big data e il diritto antitrust*, Napoli, Egea, 2018, 266 ss.

dei consumatori, cioè per ritenere la raccolta e l'accumulazione dei dati alla stregua di una pratica anticoncorrenziale, anche quando posta in essere da un'impresa già in posizione dominante[16].

Infine, esiste una quarta forma di potere che le imprese possono derivare dal controllo dei big data – una forma di potere che, grazie ad un costante monitoraggio di ciò che accade in un dato settore o mercato, finisce con il consistere nella capacità di capire prima e meglio dei propri rivali come evolverà il mercato e come i consumatori potrebbero essere soddisfatti. E ciò con l'effetto ultimo di consentire a queste imprese di posizionarsi con anticipo rispetto ai concorrenti, in modo da cogliere gli effetti positivi dei prefigurati sviluppi.

4. *Il vero potere che risiede nei* big data*: la capacità di cogliere nuove opportunità di business.*

In tempi recenti, le grandi imprese come Google, Amazon, Facebook, Apple e Microsoft sembrano godere del dono dell'ubiquità, poiché esse hanno rapidamente costruito i loro ecosistemi entrando, con rapidità e semplicità, in tanti e diversi mercati. In particolare, grazie ad un'intensa proliferazione di prodotti e servizi differenti più che diversificati, le *big tech* hanno raggiunto fatturati non comprensibili guardando ai singoli mercati nei quali esse sono nate, sebbene proprio detti mercati siano ad oggi oggetto del loro dominio. In altri termini, il potere economico vantato da queste imprese è ben più significativo del potere di mercato, ancorché già dominante, di cui esse dispongono nei rispettivi mercati rilevanti, vale a dire nei mercati – si badi – non dei big data, ma dei motori di ricerca, dei servizi di intermediazione per la compra-vendita online, dei servizi di *social networking*, o nei mercati per lo sviluppo di alcuni programmi per computer.

Posto che le *big tech* entrano nei nuovi mercati essenzialmente in due modi, acquisendo un'altra impresa o sviluppando un nuovo prodotto/servizio, il diritto antitrust è chiamato ad interrogarsi sulla liceità di comportamenti che si qualificano o come concentrazioni o come condotte unilaterali di imprese in posizione dominante.

Tuttavia, le operazioni di concentrazione per effetto delle quali un'impresa si integra in un mercato differente da quello in cui opera sono

[16] *Ibidem*, 204 ss.

considerate anticompetitive solo in circostanze decisamente specifiche, ovvero quando comportano la soppressione di nuova tecnologia[17], oppure precludono l'accesso o a un input o a un canale distributivo essenziale. A quest'ultimo proposito è emblematico il caso *Facebook/WhatsApp*. In tale fattispecie, la Commissione Europea ha scelto di autorizzare la concentrazione tra le due società statunitensi dopo aver verificato come essa non desse luogo ad alcuna forma di preclusione anticompetitiva. I bacini di dati così sommati continuavano cioè ad ammettere dei sostituti, dopo aver escluso che, per effetto dell'acquisizione di WhatsApp, Facebook diventasse l'unico custode (c.d. *gatekeeper*) dei dati digitali relativi agli utenti. La Commissione ha chiaramente affermato che, anche dopo l'operazione, «there will continue to be a large amount of Internet user data that are valuable for advertising purposes and that are not within Facebook's exclusive control»[18], finendo così con il sostenere come i concorrenti delle parti, dalle compagnie telefoniche alle altre piattaforme digitali, fossero comunque in grado di accedere a delle fonti alternative di *Internet user data* utili a fini commerciali.

Analogamente, è assai raro che una condotta unilaterale che comporta la produzione di un nuovo prodotto e/o servizio e dunque l'ingresso di un nuovo rivale in mercato possa considerarsi anticompetitiva, sebbene il nuovo entrante sia un'impresa in posizione dominante in un altro mercato. Sempre emblematico, al riguardo, è il caso dei servizi di pagamento. L'ampia disponibilità di dati circa variabili quali le consuetudini di pagamento, i vincoli di bilancio e la solvibilità degli utenti, ha recentemente consentito a non poche big tech di inserirsi nel mercato dei servizi di pagamento, pur non possedendo alcuna pre-gressa competenza specifica in questo settore. Ebbene, questo ingresso non poteva che essere salutato con favore dal diritto antitrust, perché esso ha comportato:

ridotti prezzi al consumo – i servizi finanziari e bancari sono offerti dalle *Big Data Companies* a prezzi nulli; (*ii*) un'aumentata qualità dei prodotti – questi servizi incontrano le preferenze dei consumatori in termini di velocità e semplicità; e (*iii*) l'evidente aumento del grado di innovazione.

Sotto quest'ultimo profilo occorre, infatti, ricordare che – come accennato discutendo della flessibilità con cui le *big tech* si spostano da un

[17] M. Bourreau, A. De Streel, *Digital Conglomerates and EU Competition Policy*, 11 marzo 2019, consultabile all'indirizzo https://ssrn.com/abstract=3350512.

[18] CE, 3 ottobre 2014, caso COMP/M.7217, §§ 188-189. Si vedano anche il caso *Tom-Tom/Tele Atlas* (CE, 14 maggio 2008, caso COMP/M.4854) e il caso *Google/DoubleClick* (CE, 11 marzo 2008, caso COMP/M.4731) per altre ipotesi in cui la Commissione ha negato l'esistenza di una preclusione anticompetitiva.

mercato all'altro – queste imprese trovano spazio nei mercati dei servizi bancari e finanziari anche a causa dei limiti tecnologici delle banche e degli istituti finanziari, i quali non contemplano nel loro *core business* la produzione di software.

Per contro, diverso è il caso in cui l'ingresso – o meglio l'affermazione – in un nuovo mercato sia frutto di una strategia escludente che, a prescindere dai meriti della piattaforma, sia in grado di far leva su alcuni dei vantaggi competitivi detenuti dalla piattaforma medesima nel mercato già dominato (quali dati, tecnologia, forza del brand, duplice ruolo della piattaforma) per acquisire spazio in nuovi segmenti[19].

Ora, posto che la prassi antitrust potrebbe anche modificarsi, il punto che merita di essere esaminato è che i big data devono per certo annoverarsi tra i fattori che hanno reso possibile tale "ubiquità di mercato", giacché la possibilità di analizzare in tempi assai rapidi enormi quantità di dati e di estrarre informazioni da contenuti eterogenei e distanti rispetto ai contenuti raccolti ha consentito alle big tech non solo di individuare opportunità di business altrimenti non conoscibili, ma anche di accorciare i tempi di apprendimento.

È questa la forma più eclatante di potere che i big data portano con sé. Ecco perché sotto il profilo competitivo, lo scenario che più preoccupa non è tanto e non è soltanto quello di imprese che abusano del proprio potere di mercato – scenario certo non auspicabile, ma con cui gli ordinamenti europei e nordamericani sono soliti confrontarsi, sempre a patto che vogliano intervenire, ma questo non è un tema di regole, ma di enforcement.

Lo scenario che più spaventa è quello di pochissime imprese che riescono a godere di un vantaggio competitivo impareggiabile che consenta loro di collocarsi ovunque prima e meglio dei propri rivali, andando così a sviluppare degli ecosistemi che, seppur confrontandosi aspramente tra loro, potrebbero rimane-re gli unici a campeggiare nel più generale sistema economico. In altri termini, il vero pericolo competitivo di cui la quarta rivoluzione industriale è portatrice non riguarda le modalità illecite con cui le imprese acquisiscono dati, né il controllo di specifici insiemi di dati che alcune imprese potrebbero precludere ai rivali – tutte condotte delle quali il diritto antitrust può interessarsi; il vero pericolo competitivo insito nella quarta rivoluzione industriale concerne l'eventualità che molte imprese

[19] Emblematici in questo senso appaiono i casi italiani aperti (ed in procinto di definizione) contro Amazon e Google; cfr. AGCM, Provvedimento n. 27623 del 10 aprile 2019, A528 – *Fba Amazon* e Provvedimento n. 27771 dell'8 maggio 2019, A529 – *Google/Compatibilità App Enel X Italia Con Sistema Android Auto*.

non riescano a formarsi quella conoscenza del mondo e dei comportamenti umani che invece risulta già disponibile a poche rivali, le quali non per caso risultano tra le prime a sviluppare nuovi prodotti e servizi.

Ebbene, a fronte di detto rischio competitivo, il diritto antitrust contemporaneo può poco, ma non perché esso tutela il corretto funzionamento del mercato in luogo di un'equa distribuzione della ricchezza o di altri valori dal sapore più squisitamente politico[20]. Il diritto antitrust può poco perché, come sopra osservato parlando di concentrazioni conglomerali o di ingressi in nuovi mercati, esso assume una prospettiva di "equilibrio economico parziale", ossia procede "mercato per mercato", accertando il potere di cui un'impresa gode in un dato mercato e verificando che la condotta di quella impresa non peggiori, stesso mercato.

Inoltre, al momento, dottrina e giurisprudenza non sono riuscite a coniare delle categorie di analisi capaci di abbracciare l'effetto aggregato e, per ovviare al problema delle pari opportunità circa la conoscenza del mondo e dei comportamenti umani, propongono una soluzione regolamentare, cioè l'apertura di tutti i dati da chiunque essi siano controllati, così che tutte le imprese possano partire da questa base comune per poi sviluppare i propri prodotti e servizi, compresa l'intelligenza artificiale. In questo contesto si colloca, ad esempio, la disciplina della *Public sector information*, quale prescritta dalla Direttiva 2003/98/EC. Inoltre, in questa prospettiva, si spiegano l'art. 20 del GDPR e la PSD2. L'art. 20 del GDPR – una soluzione chiaramente bottom-up – responsabilizza i consumatori, chiedendo loro di essere sufficientemente consapevoli da far circolare i propri dati tra due o più imprese. La PSD2 – una soluzione altrettanto chiaramente top-down – impone una generale obbligazione a condividere i dati ritagliata su uno specifico settore e subordinata al consenso dei correntisti. E con ciò richiede la spendita di un certo capitale politico.

Infatti, non vi è chi non veda come il favore per una generica e indiscriminata apertura dei dati ponga almeno due ordini di problemi. In primo luogo, con riferimento ai dati personali, comporta la neutralizzazione del

[20] In dottrina si è sviluppato un vivace dibattito circa il contributo che il diritto antitrust può apportare al fine di ridurre le crescenti disuguaglianze nella distribuzione della ricchezza osservate nel mondo occidentale – anche a causa di fenomeni, quali i big data, l'avvento delle tecnologie dell'informazione, e la globalizzazione – ritenute idonee a minare il funzionamento del meccanismo democratico. Cfr., ad esempio, T. Piketty, *Capital In The Twenty-First Century*, Cambridge, MA, 2014 e J.E. Stiglitz, *The Price Of Inequality: How Today's Divided Society Endangers Our Future*, W.W. Norton & Co., 2012, 338 dove l'A. sostiene che la disuguaglianza potrebbe essere meglio affrontata tramite delle «stronger and more effectively enforced competition laws».

principio del consenso, ossia la possibilità che i titolari di quei dati finiscano per vederli condivisi anche tra soggetti non espressamente autorizzati a trattarli. In secondo luogo, anche laddove si volesse escludere *ab ovo* l'esistenza di un diritto di proprietà sui dati non personali raccolti da una impresa, potrebbe essere vero che l'impresa costretta a condividere quei dati e dunque indotta a perdere il vantaggio competitivo insito negli stessi, possa di conseguenza non solo smettere di raccoglierli, ma anche di investire nei prodotti e nei servizi che l'analisi di quei dati può generare. Se così fosse, saremmo di fronte a un dilemma: lasciare che alcune imprese godano di un vantaggio competitivo forse impareggiabile, disponendoci dunque a uno scenario in cui queste poche imprese competeranno tra loro con i propri ecosistemi, oppure annullare tale vantaggio, di modo che le imprese siano chiamate a competere solo sulla tecnologia e non sull'accesso ai dati, con ciò rischiando che si riducano in generale gli incentivi ad innovare e, conseguentemente, anche a produrre intelligenza artificiale.

Ma vi è comunque di più. Se anche optassimo per la condivisione dei dati, così da azzerare il vantaggio cognitivo delle *big tech* rispetto alle loro rivali, non è chiaro dove ci dovremmo fermare, giacché le GAFAM non sopravanzano le altre imprese soltanto in termini di conoscenza, bensì anche rispetto ad altri fattori.

4.1. *L'ipotesi di muoversi oltre le pari opportunità in tema di dati*

Le menzionate società statunitensi non godono soltanto di big data, ma anche di enormi capitali che quindi permettono loro di sostenere un'elevata propensione al rischio. Ora, laddove un'impresa ordinaria potrebbe rinunciare a sfruttare un'opportunità di business disvelata dai big data per ragioni di convenienza, una *big tech* ha importanti capitali da destinare a molteplici percorsi innovativi che procedono secondo lo schema del tentativo e dell'errore. Quindi, non è detto che, condividendo i dati, il divario competitivo di cui si parla verrebbe a colmarsi.

Inoltre, non bisogna dimenticare che i (pur differenti) prodotti e servizi realizzati dalle GAFAM presentano un importante elemento in comune: essi implicano la scrittura di codice. Di fatto, le menzionate società sono innanzitutto società digitali, perché la competenza principale che esse preservano e potenziano riguarda lo sviluppo di codice: un codice che potrebbe tradursi in uno strumento di pagamento, in un mezzo di finanziamento o, ancora, in un programma di intelligenza artificiale, ma che in ultima analisi resterebbe un codice. Diversamente, le altre imprese

– quelle che sopra si sono chiamate ordinarie, perché abituate a operare nei settori non digitali – non possiedono competenze così trasversali da permettere loro di saltare da un mercato all'altro con la medesima rapidità e semplicità. A titolo di esempio, si pensi a un mutuo e a un frigorifero intelligente. Per un'impresa non digitale, si tratta evidentemente di prodotti molto differenti tra loro; per un'impresa digitale sono dei risultati di due programmi software che, in quanto tali, presentano almeno un tratto – *rectius*, una funzione aziendale e una competenza – comune: le linee di codice da scrivere. Di conseguenza, anche da questo punto di vista si potrebbe avanzare il dubbio che, da solo, il disvelamento di nuove opportunità di business reso possibile dalla condivisione dei dati non possa colmare lo iato competitivo che separa le GAFAM dalle altre imprese.

Di qui, dunque, alcune provocazioni: se l'obiettivo che si vuole raggiungere ad ogni costo è evitare che in un prossimo futuro alcune imprese raggiungano un dominio trasversale a diversi mercati, posto che il diritto antitrust può ben poco, si potrebbe anche procedere imponendo alle *big tech* – e, per portare il ragionamento fino ai suoi estremi, solo alle *big tech* – di condividere i propri dati con chiunque ne faccia richiesta. Tuttavia, non è chiaro se una volta riconosciuta alle imprese la stessa dotazione iniziale in termini di dati, questo impegno nel senso delle pari opportunità potrebbe continuare, riguardando anche la tecnologia o altre risorse di cui solo alcune imprese potrebbero disporre – o, *rectius*, di cui solo alcune imprese come le GAFAM già dispongono. Nasce cioè il sospetto che questa quarta rivoluzione industriale possa indurre – e stia inducendo le autorità pubbliche – a diffidare del meccanismo concorrenziale sino al punto da far loro argomentare che, finché a vincere non saranno soggetti diversi dalle GAFAM, allora bisognerà intervenire per garantire ai loro rivali risorse equivalenti a quelle di cui già dispongono queste imprese. Insomma, nasce il sospetto che il tema non sia lasciare il mercato libero di agire, ma sostituirsi al mercato per scegliere i vincitori.

In modo meno provocatorio, le istituzioni tedesche ed europee hanno predisposto delle soluzioni comportamentali che, più che appuntarsi sulle dotazioni iniziali delle imprese, si focalizzano sulle loro condotte dal momento in cui dette imprese si qualificano come *gatekeeper* o come imprese di fondamentale importanza proprio per la concorrenza "*across markets*"[21].

In particolare, già qui vale la pena di menzionare quelle soluzioni

[21] M. Cappai, G. Colangelo, *Taming digital gatekeepers: toward the 'more regulatory' approach to antitrust law*, in corso di pubblicazione.

comportamentali che, individuate sia dal Digital Market Act europeo (d'ora in poi, DMA) sia dalla novella della legge tedesca in tema di concorrenza, si focalizzano sulla raccolta dei dati, per: (i) garantire la correttezza e trasparenza della medesima raccolta; (ii) impedire a una piattaforma integrata su mercati collegati a quello dominato di sfruttare il proprio vantaggio informativo a danno dei concorrenti che, pur presenti in quei mercati secondari, non godono dei medesimi dati; (iii) consentire, pur sempre entro certi limiti, che i dati ottenuti da una particolare piattaforma siano oggetto di condivisione con altri soggetti.

Così, in primo luogo, contiamo gli artt. 5, lett. a) e 6, lett. h) del DMA. La prima disposizione, ispirandosi non troppo velatamente ai fatti sottesi al caso *Facebook* concluso dal Bundeskartellamt[22], stabilisce che il *gatekeeper* si astenga dal combinare dati personali ricavati dai servizi di piattaforma di base con dati personali provenienti da qualsiasi altro servizio offerto dal *gatekeeper* o con dati personali provenienti da terzi, a meno che sia stata presentata all'utente finale la scelta specifica e che quest'ultimo abbia prestato il proprio consenso ai sensi del GDPR. La seconda disposizione, invece, si preoccupa in modo generico di ricordare il rispetto dell'art. 20 del GDPR obbligando le piattaforme a garantire e promuovere l'effettiva portabilità dei dati generati mediante l'attività di un utente commerciale o di un utente finale.

In secondo luogo, richiamando alla mente i fatti che paiono essere oggetto dell'istruttoria su Amazon[23], l'art. 6, primo comma, lett. a) e il connesso secondo comma dello stesso art. 6 DMA impongono il divieto per la piattaforma di utilizzare, in concorrenza con gli utenti commerciali, dati non accessibili al pubblico generati attraverso le attività degli utenti commerciali medesimi, ossia tutti i dati aggregati e non aggregati generati dagli utenti commerciali che possono essere ricavati o raccolti attraverso le attività commerciali degli utenti commerciali o dei loro clienti sul servizio di piattaforma di base del *gatekeeper*.

In terzo luogo, le lett. g), h) e j) dell'art. 6 DMA intervengono in materia di condivisione dei dati. La prima disposizione, sebbene non menzioni espressamente i dati, appare infatti garantire ad inserzionisti ed editori l'accesso diretto e gratuito ai dati che riguardano i loro business, attribuendo

[22] Bundeskartellamt, caso B6-22/16, *Facebook*, Comunicato stampa, https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2019/07_02_2019_Facebook.html.

[23] Cfr. CE, AT.40462, *Antitrust: Commission opens investigation into possible anticompetitive conduct of Amazon*, comunicato stampa, https://ec.europa.eu/commission/presscorner/detail/en/ip_19_4291.

a tali soggetti non solo la piena facoltà di accedervi ma anche di analizzare quei dati; ciò comporta che l'analisi dei dati può essere svolta in autonomia da inserzionisti ed editori che quindi non dovranno necessariamente comprarla sotto forma di servizio offerto dalla piattaforma[24].

La seconda previsione impone ai *gatekeeper* di fornire *a titolo gratuito* agli utenti commerciali un "accesso efficace, di elevata qualità, continuo e in tempo reale" a dati aggregati e non aggregati"; la piattaforma è tenuta altresì a garantire alle stesse condizioni l'uso di tali dati aggregati e non aggregati che sono forniti o generati nel contesto dell'uso dei pertinenti servizi di piattaforma di base da parte di tali utenti commerciali e degli utenti finali che si avvalgono di prodotti o servizi forniti da tali utenti commerciali. Con riferimento ai dati personali, poi, viene specificato che l'accesso e l'utilizzo deve essere garantito solo se si tratta di dati direttamente connessi con l'uso effettuato dall'utente finale in relazione ai prodotti o servizi offerti dal pertinente utente commerciale mediante il pertinente servizio di piattaforma di base e solo nel caso in cui l'utente finale abbia accettato tale condivisione.

Infine, la terza disposizione interviene con specifico riferimento ai dati prodotti dai motori di ricerca e stabilisce che il *gatekeeper* sia tenuto a garantire ai fornitori terzi l'accesso a condizioni eque, ragionevoli e non discriminatorie a dati relativi a posizionamento, ricerca, click e visualizzazione per quanto concerne le ricerche gratuite e a pagamento generate dagli utenti finali, fatta salva l'anonimizzazione dei dati personali.

Sulla falsariga del DMA, anche la riforma tedesca prevede il potere di intervento del *Bundeskartellamt* al fine di: (i) vietare quelle pratiche che comportando il trasferimento del potere di mercato a mercati precedentemente non dominanti, utilizzando e/o collegando dati esistenti rilevanti per la concorrenza, qualora questo crei o aumenti le barriere all'ingresso o ostacoli in altro modo altre imprese; nonché (ii) proibire di trattenere i dati – che le piattaforme ottengono in virtù dei loro servizi – in modo da creare dipendenze ingiustificate. In altre parole, per quanto riguarda questo secondo comportamento, le piattaforme sono tenute a fornire ad altre aziende informazioni insufficienti sulla portata, la qualità o il successo del servizio fornito o commissionato; co-sì, ad esempio, nel caso della pubblicità online, al cliente pubblicitario non può essere impedito senza motivo di valutare da solo i suoi indicatori chiave di performance (KPI)[25].

[24] Letteralmente la lett. g) dell'art. 6 dispone che il *gatekeeper* «fornisce a inserzionisti ed editori, su loro richiesta e a titolo gratuito, l›accesso ai propri strumenti di misurazione delle prestazioni e le informazioni necessarie agli inserzionisti e agli editori affinché possano effettuare una verifica indipendente dell'offerta di spazio pubblicitario».

[25] Cfr. art. 19 a), (2), nn. 4 e 6.

5. *Le risposte tedesca ed europea agli ecosistemi delle GAFAM*

Sul finire del 2020, il parlamento tedesco ha finalizzato la *Gesetz zur Änderung des Gesetzes gegen Wettbewerbsbeschränkungen für ein fokussiertes, proaktives und digitales Wettbewerbsrecht 4.0 und anderer Bestimmungen* (*GWB-Digitalisierungsgesetz*), vale a dire una riforma della legge della concorrenza tedesca che prende atto della quarta rivoluzione industriale e del conseguente sviluppo dell'economia digitale. Nel gennaio del 2021 è così entrato in vigore il nuovo art. 19, lett. a), il quale è proprio diretto a disciplinare le condotte delle imprese che possono dirsi di "fondamentale importanza per la concorrenza tra mercati" ed, in particolare, sette classi di pratiche che il *Bundeskartellamt* potrà considerare presumibilmente vietate, salvo che le imprese non riescano a produrre una giustificazione oggettiva a sostegno delle medesime condotte.

Più nel dettaglio, quanto all'elemento strutturale della fattispecie, la norma prevede che un'impresa possa qualificarsi come di fondamentale importanza per la concorrenza *across markets* alla luce di fattori ulteriori rispetto alla mera posizione dominante che l'impresa già detenga su uno o più mercati che pure viene presa in considerazione. Tra questi fattori si contano elementi che, evidentemente, catturano il fenomeno degli ecosistemi digitali e le tante frecce che, come si diceva sopra, le GAFAM hanno al proprio arco: dalla capacità di accedere al credito o ad altre risorse di natura finanziaria e non; al grado di integrazione verticale e orizzontale che consente all'impresa di operare su molti mercati tra loro collegati; dalla possibilità per l'impresa di accedere a dati rilevanti per la concorrenza; all'importanza che le attività dell'impresa rivestono per l'accesso dei terzi ai mercati di fornitura e vendita, nonché la relativa influenza sulle attività commerciali di terzi.

Con riguardo all'elemento comportamentale del divieto, invece, come in rapporto alla più tradizionale fattispecie dell'abuso di posizione dominante, le imprese saranno sempre legittimate a dimostrare di non aver commesso un illecito, offrendo delle ragioni che possano giustificarlo in termini oggettivi. Tuttavia, il compito del *Bundeskartellamt* sarà enormemente semplificato rispetto a quello di chi dovrebbe applicare il più consueto art. 102, perché l'autorità tedesca non dovrà mostrare gli effetti anticompetitivi prodotti da una impresa di importanza fondamentale per la concorrenza *across markets* che abbia scelto di:

a) fare discriminazioni quanto all'accesso ai mercati di approvvigionamento e vendita tra sé e i propri concorrenti privilegiando,

in particolare, le proprie offerte nella loro presentazione al pubblico[26] o preinstallando esclusivamente le proprie offerte sui dispositivi tramite i quali quelle medesime offerte raggiungono il pubblico[27];

b) ostacolare altre imprese nelle loro attività commerciali sui mercati di approvvigionamento o di vendita adottando, in particolare, misure che portino a una preinstallazione o integrazione esclusiva delle offerte dell'impresa o impedendo ad altre imprese di pubblicizzare le proprie offerte anche attraverso punti di accesso diversi da quelli forniti o mediati dall'impresa[28];

c) ostacolare i concorrenti su mercati in cui l'impresa può espandere rapidamente la propria posizione, anche senza essere dominante (in particolare, combinando l'uso di un'offerta con l'uso automatico di un'altra offerta dell'impresa, senza concedere sufficienti possibilità di scelta, o facendo dipendere l'uso di un'offerta dell'impresa dall'uso di un'altra offerta dell'impresa);

d) trattare i dati sensibili alla concorrenza raccolti per creare o innalzare barriere all'entrata nel mercato o per richiedere termini e condizioni per tale utilizzo;

e) impedire l'interoperabilità dei prodotti/servizi o la portabilità dei dati[29];

f) richiedere vantaggi per il trattamento delle offerte di un'altra azienda che siano sproporzionati al motivo della richiesta.

Le condotte affrontate dal nuovo art. 19, lett. a) della legge tedesca sono sostanzialmente simili a diverse pratiche vietate dalla proposta di Digital Markets Act (DMA), elaborata in sede europea con la differenza che mentre la lista tedesca si pone come esaustiva, quella europea risulta solo esemplificativa.

[26] Qui pare immediato il riferimento alla pratica del *self-preferencing* di cui al caso *Google Shopping* consistente nella sistematica retrocessione nei risultati di ricerca di Google Search, per il tramite un algoritmo ad hoc, dei servizi di comparazione degli acquisti offerti dai *competitor* di Google; retrocessione che, di contro, era accompagnata dalla visualizzazione di Google Shopping, con notevole evidenza grafica, in cima nella prima pagina dei risultati di ricerca. Cfr. CE, 27 giugno 2017, caso AT. 39740.

[27] Anche qui sembra immediato il riferimento ai molti casi di pre-installazione di software, anche nella forma di applicazioni, nei sistemi operativi di personal computers o dispositivi palmari, come i telefoni intelligenti. Tra questi, si pensi ad esempio, al caso *Google Android*, nell'ambito del quale la Commissione ha accertato l'esistenza di meccanismi contrattuali tramite i quali Google è riuscita ad imporre, tra le altre cose, la preinstallare della sua applicazione di ricerca (Google Search) e di *browsing* (Google Chrome) sui dispositivi mobili basati sul sistema operativo Android. CE, 18 luglio 2018, caso AT. 40099.

[28] Qui il caso rilevante sembra quello dei videogiochi.

[29] Anche qui la memoria corre al caso Microsoft europeo e al tema della compatibilità.

Più esattamente, nel dicembre del 2020 la Commissione europea ha pubblicato l'attesa proposta di regolazione dei mercati digitali, al fine di integrare gli strumenti a disposizione della Commissione per applicare in modo efficace le regole a tutela della concorrenza nei mercati digitali[30]. Tuttavia, occorre da subito chiarire come la Commissione spieghi che la proposta persegue un obiettivo diverso da quello tipico del diritto antitrust. Essa non vuole proteggere il funzionamento del mercato, ma garantire che nei settori digitali in cui sono presenti i c.d. "gatekeeper" i mercati restino *contendibili* ed *equi*, indipendentemente dagli effetti reali, probabili o presunti che i comportamenti di dati *gatekeeper* producono sul mercato[31].

In via ancora più specifica, il DMA svela la sua natura eminentemente regolatoria quando chiarisce che, sebbene gli artt. 101 e 102 TFUE rimangano applicabili al comportamento dei *gatekeeper*, non solo la loro portata è limitata a determinati casi di potere di mercato e di comportamenti anticoncorrenziali,ma la loro applicazione avviene ex post e richiede un'indagine approfondita, caso per caso, di fatti spesso molto complessi [32]. Diversamente, a tutela del mercato interno[33], il DMA avrebbe il merito di imporre una serie di divieti e obblighi applicabili, a prescindere dai loro effetti, in capo ai soggetti che ai sensi della medesima disciplina si qualificano alla stregua di *gatekeeper*, di là dal fatto che questi ultimi risultino o meno dominanti.

Per quello che qui più interessa, occorre sottolineare come l'ambito di applicazione del DMA *ratione personae* venga individuato guardando a: (i) la natura dei servizi forniti dalla piattaforma online e (ii) la designazione di quest'ultima come *gatekeeper*.

Con riferimento al primo requisito, viene introdotto il concetto di "*core platform services* "(ossia "servizi di piattaforma di base") al fine di isolare quei servizi digitali in relazione ai quali – a causa delle caratteristiche economiche dei servizi medesimi – sarebbe più frequente osservare sia la scarsa contendibilità dei mercati, sia il verificarsi di pratiche sleali. L'elenco di tali servizi di base include i servizi di intermediazione online (compresi, ad esempio, i mercati, i negozi di applicazioni software e i servizi di intermediazione online in altri settori come la mobilità, i trasporti o

[30] Commissione europea, «Proposta di Regolamento relativo a mercati equi e contendibili nel settore digitale» (Legge sui Mercati Digitali), COM(2020) 842 final.

[31] Considerando 10.

[32] Considerando 5.

[33] La base giuridica del DMA è costituita dall'art. 114 TFUE, e non dall'art. 103 che invece costituisce la base giuridica per l'implementazione delle disposizioni antitrust ai sensi degli artt. 101 e 102 TFUE.

l'energia), i motori di ricerca online, i servizi di social network, i servizi di piattaforma per la condivisione di video, servizi di comunicazione elettronica interpersonale indipendente dal numero, sistemi operativi, servizi cloud e servizi di pubblicità[34].

In buona sostanza, dunque, innanzitutto sono *gatekeeper* le imprese che, offrendo questi servizi, vengono a ricoprire la posizione di chi finisce per governare il mercato e chi riesce ad operarvi [35].

Per quanto invece riguarda il secondo requisito, una piattaforma online raggiunge lo status di *gatekeeper* sulla base di tre criteri cumulativi, vale a dire nel caso in cui l'impresa: (i) abbia un "impatto significativo" sul mercato interno, (ii) gestisca un servizio di piattaforma di base che costituisce un punto di accesso (*gateway*) importante affinché gli utenti commerciali raggiungano gli ut enti finali e (iii) detenga una "posizione consolidata e duratura" nell'ambito delle proprie attività o è prevedibile che acquisisca siffatta posizione nel prossimo futuro[36]. In aggiunta, la proposta introduce dei criteri quantitativi *ad hoc* (basati sul fatturato e sul numero di utenti attivi) corrispondenti a ciascun requisito qualitativo: laddove tali soglie siano soddisfatte si presume che ognuno dei criteri qualitativi sia automaticamente soddisfatto[37]. La piattaforma, d'altro canto, può vincere la presunzione presentando "argomentazioni sufficientemente fondate" per dimostrare che, nonostante rientri nelle soglie, di fatto non soddisfa i requisiti di tipo qualitativo sopra riportati. Tuttavia, come evidenziato in dottrina[38], questa è l'unica via tramite cui le imprese possono sottrarsi alla nozione di *gatekeeper*, visto che sarebbero prive di valore salvifico sia le efficienze economiche prodotte dalle imprese per i servizi svolti[39], sia il particolare modello di business da loro adottato.

Se poi non bastasse, la Commissione si riserva la facoltà di identificare come *gatekeeper q*ualsiasi fornitore di servizi di piattaforma di base che, sebbene non raggiunga le soglie quantitative, soddisfi gli elementi di tipo qualitativo[40]. Inoltre, al fine di evitare fenomeni di c.d. *market tipping*, la

[34] Art. 2(2). La Commissione si riserva la facoltà di allargare la lista dei servizi a seguito di una *market investigation*.
[35] Cfr. Considerando 2 e 12.
[36] Art. 3(1).
[37] Art. 3(2).
[38] M. Cappai, G. Colangelo, *Taming digital gatekeepers*, cit., 20.
[39] Considerando 23.
[40] Art. 3(6). Nell'effettuare tale valutazione la Commissione tiene conto dei seguenti elementi «a) le dimensioni, compresi fatturato e capitalizzazione di mercato, le attività e la posizione del fornitore di servizi di piattaforma di base; b) il numero di utenti commerciali che dipendono

proposta di DMA prevede anche la possibilità di designare un "*gatekeeper* emergente" quando una piattaforma soddisfa i primi due criteri qualitativi (impatto significativo e importante; *gateway*) ma il criterio della posizione consolidata e duratura sia solo prevedibile e non già realizzato[41].

Definiti i soggetti destinatari, la proposta di legge introduce una serie fissa di diciotto obblighi *ex ante* suddivisi in due liste: una lista di obblighi c.d. *selfenforcing*, ossia direttamente applicabili (art. 5) e una lista di obblighi suscettibili di ulteriori specificazioni da parte della Commissione (art. 6). Più nel dettaglio, al netto delle norme descritte brevemente nel precedente paragrafo, l'art. 5 della proposta di DMA introduce a carico dei *gatekeeper*:

a) Obblighi a vario titolo connessi alla materia di prezzi. Segnatamente, probabilmente in ragione del dibattito in tema di *parity clauses* (o *most favoured nation clauses*, MFN) – dibattito risoltosi nel senso di considerare tali clausole vietate[42] – i *gatekeeper* devono consentire agli utenti commerciali di offrire gli stessi prodotti o servizi agli utenti finali attraverso servizi di intermediazione online di terzi a prezzi o condizioni diverse da quelle offerte attraversoi servizi di intermediazione online del *gatekeeper* (art. 5, lett. b). La stessa logica, ovvero la volontà di vietare pratiche funzionali a strategie di prezzo che potrebbero comportare effetti collusivi, appare sottesa all'obbligo di fornire ad inserzionisti ed editori cui la piattaforma eroga servizi pubblicitari, su loro richiesta, informazioni relative al prezzo pagato dall'inserzionista o dall'editore, ed all'importo o alla remunerazione versati all'editore, per la pubblicazione di una determinata inserzione e per ciascuno

dal servizio di piattaforma di base per raggiungere gli utenti finali e il numero di utenti finali; c) le barriere all'ingresso derivanti da effetti di rete e vantaggi basati sui dati, in particolare in relazione all'accesso a dati personali o non personali e alla raccolta di tali dati da parte del fornitore o alle capacità di analisi di quest'ultimo; d) gli effetti di scala e in termini di portata di cui usufruisce il fornitore, anche per quanto riguarda i dati; e) il lock-in degli utenti commerciali o degli utenti finali; f) altre caratteristiche strutturali del mercato».

[41] Cfr. art. 15(4), Considerando 26 e 63.

[42] Le clausole di parità di prezzo, spesso usate dalle piattaforme operanti nel settore delle prenotazioni alberghiere, consentono alla piattaforma (*Online Travel Agency*, OTA) di richiedere che i fornitori non offrano prezzi più bassi o condizioni migliori su altre piattaforme o sui propri siti web. Il modo in cui le autorità di concorrenza hanno considerato la legalità di queste clausole è variato considerevolmente nel tempo e ha causato un certo grado di incertezza giuridica per le imprese. Si veda, ad esempio, l'istruttoria condotta dall'AGCM nei confronti di Booking ed Expedia (AGCM, Provvedimento n. 25422, 21 aprile 2015, I779 – *Mercato Dei Servizi Turistici-Prenotazioni Alberghiere On Line*). Sul tema si veda M. Colangelo, *Competition Law and Most Favoured Nation Clauses in Online Markets* in K. Mathis, A. Tor (eds.), *New Developments in Competition Law & Economics, Economic Analysis of Law in European Legal Scholarship Series*, Springer, 2019, https://ssrn.com/abstract=3293716.

dei pertinenti servizi pubblicitari forniti dal *gatekeeper* (art. 5, lett. g).

b) Divieti di strategie di tying e, in particolar modo, il dovere dei *gatekeeper* di consentire agli utenti commerciali di promuovere offerte agli utenti finali acquisiti attraverso il servizio di piattaforma di base e di stipulare contratti con tali utenti finali, a prescindere dal fatto che a tale fine essi si avvalgano o no dei servizi di piattaforma di base del *gatekeeper*. I *gatekeeper* sono altresì tenuti a consente agli utenti finali di accedere a contenuti, abbonamenti, componenti o altri elementi e di utilizzarli attraverso i servizi di piattaforma di base avvalendosi dell'applicazione software di un utente commerciale tramite la quale sono stati acquistati (art. 5, lett. c). Parimenti, la lett. f) dell'art. 5 dispone che la piattaforma debba astenersi dall'imporre agli utenti commerciali o agli utenti finali l'abbonamento o l'iscrizione a qualsiasi altro servizio di piattaforma di base quale condizione per accedere, registrarsi o iscriversi a uno dei suoi servizi. Complessivamente, dunque, il DMA pare essere volto a consentire che le piattaforme accolgano tanti e diversi servizi (strategie c.d. di *mix and match*) senza subordinare l'offerta di un servizio ad un altro servizio offerto dalla piattaforma; in altre parole, le disposizioni menzionate impongono un obbligo di compatibilità commerciale (tramite, appunto, il divieto di *tying*) che non può non essere accompagnato – a monte – da un obbligo di compatibilità tecnologica (ossia la capacità del servizio di accedere alla piattaforma), quale precondizione di fatto; ed, infatti, proprio in quest'ultimo senso paiono collocarsi gli obblighi di cui alle lettt. e) ed f) dell'art. 6, ai sensi dei quali, rispettivamente, la piattaforma si astiene dal limitare a livello tecnico la possibilità per gli utenti finali di passare e di abbonarsi a servizi e applicazioni software diversi, e deve consentire agli utenti commerciali (e ai fornitori di servizi ausiliari) l'accesso allo stesso sistema operativo e alle stesse componenti hardware o software disponibili o utilizzati nella fornitura di servizi ausiliari da parte del *gatekeeper* nonché l'interoperabilità con gli stessi.

c) L'obbligo di garantire il diritto degli utenti a fare azione contro i *gatekeeper*, ossia il dovere di astenersi dall'impedire agli utenti commerciali di sollevare presso qualsiasi autorità pubblica competente questioni relative alle pratiche dei *gatekeeper* o dal limitare tale possibilità (art. 5 lett. d); ed infine.

d) Il dovere di astenersi dall'imporre agli utenti commerciali che si avvalgono dei servizi di piattaforma di base del *gatekeeper* l'utilizzo o l'offerta di un servizio di identificazione[43] del *gatekeeper*, o l'interoperabilità

[43] Con l'espressione servizi di identificazione si intendono quei servizi che consentono qualsiasi

con lo stesso (art. 5, lett. e).

Inoltre, tra gli obblighi soggetti ad ulteriore specificazione rientrano, al di là delle disposizioni già richiamate e del generico divieto di discriminazione di cui alla lett. k) dell'art. 6, due obblighi che paiono richiamare i recenti casi *Google Android* e *Google Shopping*. Infatti, si impone ai *gatekeeper* di:

a) consentire agli utenti finali di disinstallare qualsiasi applicazione software preinstallata (che non sia essenziale per il funzionamento stesso del sistema operativo o del dispositivo) sul proprio servizio di piattaforma di base (art. 6, lett. b) ed una norma – ad esso complementare – che impone al *gatekeeper* di consentire l'installazione e l'uso effettivo di applicazioni software o di negozi di applicazioni software di terzi, avendo la piattaforma unicamente la facoltà di adottare misure proporzionate per garantire che le applicazioni software o i negozi di applicazioni software di terzi non presentino rischi ai fini dell'integrità dell'hardware o del sistema operativo fornito dal *gatekeeper* (art. 6, lett. c); ecco che allora, come si accennava, si tratta con tutta evidenza di disposizioni che traggono origine dal caso comunitario *Google Android*[44] e che rievocano i casi comunitari e statunitensi aperti contro Apple per la gestione del proprio *App Store*[45]*;*

tipo di verifica dell'identità degli utenti finali o degli utenti commerciali, indipendentemente dalla tecnologia utilizzata (art. 2, n. 15); sono considerati particolarmente importanti affinché gli utenti commerciali svolgano la propria attività poiché oltre a consentire loro di ottimizzare i propri servizi, possono, infondere fiducia nelle transazioni online (cfr., Considerando 40).

[44] CE, 18 luglio 2018, caso AT. 40099.

[45] Su denuncia di Spotify la Commissione ha aperto un procedimento contro Apple – tuttora in corso – volto ad accertare, tra l'altro, se Apple abbia usato il proprio *App Store* per ostacolare Spotify a vantaggio del proprio servizio di streaming musicale *Apple Music*; cfr. CE, *La Commissione apre indagini sulle regole dell'App Store di Apple*, 2020, Comunicato stampa IP/20/1073, https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1073. Inoltre, la Commissione europea ha aperto un'indagine antitrust riguardante i termini, le condizioni e le altre misure di Apple per l'integrazione di *Apple Pay* nelle app e nei siti web dei commercianti su iPhone e iPad, la limitazione da parte di Apple dell'accesso alla funzionalità *near-field communication* (*NFC*) sugli iPhone per i pagamenti nei negozi ("*tap and go*") e i presunti rifiuti di accesso ad *Apple Pay* per prodotti specifici di rivali su dispositivi mobili intelligenti di Apple; cfr. CE, *Commission opens investigation into Apple practices regarding Apple Pay*, 2020, Comunicato stampa IP/20/1075, https://ec.europa.eu/commissione/presscorner/detail/it/ip_20_1075. Negli Stati Uniti, la recente causa intentata da Epic Games contro Apple e Google assomiglia alle indagini europee. In particolare, nell'agosto 2020, Epic Games ha aggiunto un'opzione di pagamento diretto scontato per il videogioco di successo *Fortnite* accanto alle opzioni di pagamento di *iOS App Store* e *Google Play*, in violazione delle politiche di quei negozi e aggirando la tassa del 30% di Apple. Di conseguenza, *Fortnite* è stato rimosso da entrambe le piattaforme ed Epic ha intentato delle cause legali lamentando che Apple e Google si pongono come intermediari inevitabili per gli sviluppatori di *app* e in ogni transazione *in-app* e denunciando restrizioni nel mercato della distribuzione delle *app* e dell'elaborazione dei pagamenti *in-app*.

b) astenersi dal riservare un posizionamento più favorevole ai servizi e prodotti offerti dal *gatekeeper* stesso ovvero da terzi che appartengono alla stessa impresa rispetto a servizi o prodotti analoghi di terzi (art. 6, lett. d); e, come si preannunciava, e parimenti al divieto contenuto nella legge tedesca, anche questa disposizione pare richiamare il caso *Google Shopping*.

Complessivamente, dunque, il DMA include un ampio insieme di obblighi e divieti che sembrano riflettere il rischio che i *gatekeeper* rafforzino e meglio radichino il proprio potere *across markets*, attraverso strategie di *bundling commerciale e tecnologico*[46]. Come osservato in dottrina, poiché salvo poche eccezioni si tratta di obblighi e divieti destinati a trovare applicazione nei confronti dei *gatekeeper* a prescindere dalla natura del servizio offerto in concreto e al di là del *business model* adottato dalla piattaforma[47], non può escludersi che essi abbiano senso nei confronti di uno specifico servizio (ad esempio, un motore di ricerca), ma non di un altro (ad esempio, un *marketplace*) e viceversa[48]. Inoltre, il fatto che i menzionati obblighi e divieti catturino pratiche soggette a casi antitrust passati e in corso, conferma la volontà europea di rendere più veloce ed immediato l'*enforcement* delle norme e quindi il divieto di certe pratiche, rivolgendosi appunto alla regolazione in luogo del diritto antitrust. Il DMA, infatti, non permette alle imprese di presentare giustificazioni oggettive o difese connesse alle efficienze collegate alla pratica oggetto di obbligo e divieto[49].

6. *Conclusioni*

Il diritto antitrust mira a garantire il corretto funzionamento del mercato, perseguendo alcune condotte ma – di norma – senza modificare le c.d. dotazioni iniziali, ossia le risorse di cui inizialmente gode ogni impresa per svolgere la propria attività, si tratti di materie prime, competenze,

[46] P. Ibanèz Colomo, *The Draft Digital Markets Act: a legal and institutional analysis*'(2021) 3, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3790276.

[47] Solo alcuni obblighi dovrebbero, secondo il testo della proposta, applicarsi unicamente ad alcuni fornitori di servizi di base.

[48] Cfr. D. Geradin, *Should all digital gatekeepers be subject to the same obligations under the DMA proposal?, The Platform Law Blog*, 2021

[49] Le uniche eccezioni sono contemplate agli artt. 8 e 9 nei casi in cui gli obblighi mettano in pericolo la redditività economica del funzionamento della piattaforma ovvero per motivi imperativi di interesse pubblico.

capitali o altri input ancora.

Ebbene, nel contesto della quarta rivoluzione industriale, i big data si sono rivelati una risorsa particolarmente ambita, non solo perché da essi le imprese estraggono valore, ma anche perché essi sono fonte di potere ed, esattamente, di tre forme di potere: (*i*) il potere di manipolare l'opinione dei consumatori; (*ii*) il potere di rendere la propria offerta più appetibile; e (*iii*) il potere di intuire prima e meglio degli altri delle opportunità di business.

Mostrato come nessuna di queste forme di potere abbia a che fare con il potere di mercato, inteso come la capacità di peggiorare la propria offerta senza perdere un numero considerevole di clienti, il diritto antitrust: (*i*) può, ancorché a determinate condizioni, perseguire il potere di agire sulla domanda; (*ii*) può catturare, non perseguire, la capacità di una impresa di rafforzare la propria posizione di mercato migliorando i propri prodotti e servizi; ma (*iii*) non può impedire a un'impresa di ampliare il dominio della propria conoscenza.

Giacché le maggiori imprese statunitensi che operano nei mercati digitali sfruttano proprio questa capacità di intravedere nuove e potenziali opportunità di affari per creare i propri ecosistemi e giacché in molti temono che nel prossimo futuro le GAFAM finiranno con l'essere le uniche imprese capaci di competere "across markets", si pone l'esigenza di colmare il gap conoscitivo che separa dette GAFAM dalle altre imprese.

Ribadito che un eventuale obbligo a condividere i big data, ossia ad aprire gli insiemi di dati, sarebbe frutto di una scelta regolatoria, due sono in buona sostanza gli interrogativi che questo approccio alimenta. In primo luogo, pare ragionevole domandarsi se la volontà di impedire il successo delle GAFAM "*across markets*" rappresenti non un rimedio a fronte di una condotta anticompetitiva o di un fallimento del mercato, ma un chiaro e palese rifiuto dei risultati delle dinamiche competitive e, in ultima istanza, delle scelte dei consumatori. In secondo luogo, se fosse vero che il vantaggio competitivo delle GAFAM non risiede solo ed esclusivamente nei big data, ma anche in altri fattori come capitale e competenze, non è chiaro fin dove questo intento regolatorio finalizzato a garantire le pari opportunità possa spingersi.

Fortunatamente, almeno in Germania, queste forme di regolazione estrema si sono evitate, individuando una serie esaustiva di condotte che, poste in essere dalle imprese di "fondamentale importanza per la concorrenza *across markets*", potranno dirsi anticompetitive, senza bisogno di provarne gli effetti negativi sulla concorrenza, purché le imprese non

possano fornirne una giustificazione oggettiva.

Analogamente, anche nell'Unione Europea si è optato per una legislazione rivolta a una specifica categoria di soggetti, quasi univocamente identificati, e chiamata a individuare pratiche vietate e obblighi. Tuttavia, la scelta europea oltre ad introdurre alcune disposizioni volte a regolare la raccolta, l'uso e la condivisione di dati da parte dei *gatekeeper*, si colloca ben al di fuori del perimetro antitrust, perché prescinde da qualsiasi analisi caso per caso dell'impatto che i comportamenti considerati possono produrre sul mercato.

Vincenzo Zeno-Zencovich

Do "data markets" exist?

Abstract: Data are an infinite resource which is continuously produced in ever increasing amounts. Personal data shares with general data non-consumability and non-rivality, enabling individuals to use their data as an unlimited currency to buy a vast amount of digital services. How can traditional competition principles (whose basic cornerstones are scarcity and limits to power of expenditure) apply to such a new environment?
The article suggests that in order to pursue policy goals (consumer welfare, innovation, protection of individual fundamental rights) a holistic regulation, which takes into account the interests of the multiple stakeholders, and traditional consumer protection legislation are more appropriate, especially in the light of the general need for legal certainty.

1. *Introduction*

Presented, with typical boisterous journalistic style, as the "new oil", "Big Data" has rapidly revealed its profound differences: data are non-material entities, non-consumable and – to a certain extent – non-rivalrous.

This different qualification does not diminish the importance of Big Data. In the last decade it has increased and awareness by economists and lawyers has brought to a deeper knowledge of the phenomenon.

More recently Big Data have entered in the visual field of competition authorities who are concerned by possible restrictive consequences of detention of huge quantities of data by what are called "Big Tech" companies, under the two typical situations that require antitrust scrutiny: that of restrictive agreements or concerted practices; and that of abuse of a dominant position.

The obvious corollary of this approach is that of establishing if "data [or Big Data] markets" exist, and what is their nature.

It is a good intellectual practice, before claiming to have discovered some novelty, to look back and see what has happened in the past.

* This article was first published in *MediaLaws*, *Rivista di diritto dei media*, 2, 2019, pp. 1-17.

Information – *i.e*. structured and oriented data – has always existed. Especially in the financial and business sectors information was – and still is – collected to evaluate creditworthiness. Financial markets have always been informational markets, in which the value of a share or of a bond is dependent from the amount of information one possesses concerning a company and the context in which it operates[1]. At a very elementary level, before the Internet age, many companies were providing for a very small sum, information concerning the phone number or the whereabouts of a subscriber or of a business. Still now there is a flourishing market – especially in the medical sector and in the US – of personal data, which in certain cases can reach $50 per name[2].

Going back in time, in the famous *Associated Press v. US*[3] case the US Supreme Court established, with a clear pro-competitive approach, that access to news agency reports (again a case of structured data) could not be restricted to competing media companies[4].

[1] It is impossible to draft an exhaustive bibliography on the role of information in financial markets. For a few indications see D. Chambers-E. Dimson, *Financial Market History: Reflections on the Past for Investors Today*, CFA Institute Research Foundation, 2016; H. S. Houthakker-P.J. Williamson, *The Economics of Financial Markets*, Oxford, 1996; P.-J. Engelen, *Remedies to Informational Asymmetries in Stock Markets*, Cambridge, 2005.

[2] See the Report by the US Federal Trade Commission, *Data Brokers. A Call for Transparency and Accountability* (May 2014). For the economic analysis of a series of new information markets see D. Bergemann-A. Bonatti, *Markets for Information: An Introduction*, Cowles Foundation Discussion paper no. 2142 (August 2018).

[3] 326 US 1 (1945). In his usual assertive style Justice Black *per curiam* stated that «The First Amendment, far from providing an argument against application of the Sherman Act, here provides powerful reasons to the contrary. That Amendment rests on the assumption that the widest possible dissemination of information from diverse and antagonistic sources is essential to the welfare of the public, that a free press is a condition of a free society. Surely a command that the government itself shall not impede the free flow of ideas does not afford nongovernmental combinations a refuge if they impose restraints upon that constitutionally guaranteed freedom. Freedom to publish means freedom for all, and not for some. Freedom to publish is guaranteed by the Constitution, but freedom to combine to keep others from publishing is not. Freedom of the press from governmental interference under the First Amendment does not sanction repression of that freedom by private interests. The First Amendment affords not the slightest support for the contention that a combination to restrain trade in news and views has any constitutional immunity».

[4] Can information be considered an "essential facility"? According to the ECJ, in the *Magill* decision (C-241/91), yes. According to the US Supreme Court, in *Verizon v. Law Offices of Curtis Trinko*, 540 US 398 (2004), albeit in a somewhat different context, no («There is no duty to aid competitors. Antitrust analysis must always be attuned to the

What changes with Big Data? Many things, because the sheer size of data modifies their role, use and value.

What must be also considered is that the growth of telecommunication networks, their ubiquity, the fact that practically all objects are or will be connected, determines a constant production and flow of data which enable monitoring and decision making in real time.

Data have become an essential component – one might call them a raw material – of any business. In this context hundreds of businesses have developed making the collection, processing, sale and exploitation of data (or of their sub-products) their core business.

2. *Datafication*

If data are so important – if not essential – in contemporary economy and their importance will be ever-growing[5] it appears reasonable to try to define "data markets".

A few premises are necessary.

The term "data" which has been used recurrently is not altogether precise, and data scientists do not necessarily agree on its exact meaning and the difference with other terms[6].

This ambiguity is increased by the fact that practically any material object, any process, any event that exists or happens in this world – but also beyond this planet and in the infinity of outer-space – can be datafied.

The fact that everything – from the incredibly small to the incredibly big – can be and is actually digitalized has significant consequences on the aim of this paper. Data – whatever their precise, technical or related, meaning – are an infinite resource: they are not limited in time – one can datafy geological events that happened millions of years ago, as one can datafy explosions that happened in a remote galaxy distant from us

particular structure and circumstances of the industry at issue»).

[5] One can find sufficient elements in the vast report by the OECD, *Data-driven Innovation for Growth and Well-being* (October 2014), and see how much way has been made over the last four years.

[6] See N. Duch-Brown, B. Martens, F. Mueller-Langer, *The economics of ownership, access and trade in digital data*, JRC Digital Economy Working Paper 2017-01, 6 ss. At any rate, in this work I will be following the model which moves from data (any representation in digital form of something), to information (structured data with a discernible, for humans, meaning), to knowledge.

thousands of light-years.

And one can imagine that the production of data – quite differently from any natural resource – will never end, until we dispose of the means to collect and digitalize them[7].

This is something novel in economic theory – especially in its antitrust side – which generally contemplates scarce resources[8]. Furthermore, it would appear that in many cases the "production costs" of data are very low if not insignificant.

It is difficult to find equivalent situations. Numbers are infinite, and we find a market of numbers only when they are made scarce, as in the case of telephone numbers and there is a request for easy-to-remember numbers.

In the other cases the resources could be considered infinite (the sea, the sky), but public policy reasons (security, safety, environmental protection) restrict their use, exploitation and appropriation. But, setting aside specific regulation on "personal data", no such restrictions can be found with data.

3. *"Ownership" of data*

Quite commonly there is a great debate, also when tackling competition issues, on the ownership of data. Especially when the topic is examined by the economists (but often even by lawyers) there is a great confusion which requires to be dispelled. "Ownership" is not a notion which is engraved in some sacred tables. It is the result of centuries, millennia of theoretical, religious, political, social, economic evolution. And as the law can only be expressed in words "ownership" means what it means in English speaking jurisdictions. Once translated in a different language it means what it means in that jurisdiction. This is one of the main tasks of comparative law: trying to understand what similar terms mean in different legal systems; and how to find corresponding terms for similar legal institutions.

[7] Adapting the typical explanation of the infinity of numbers, if we imagine that data are finite, simply by processing such finite database or any of its elements we create one more (meta)datum which increases the number of existing data. This universe of data has been qualified as "datasphere": see J.-S. Bergé, S. Grumbach, V. Zeno-Zencovich, *The 'Datasphere', Data Flows beyond Control, and the Challenges for Law and Governance*, in *Eur. J. Comparative L.,* 5, 2018, 144 ss.

[8] Subsequently one tries to define the relevant market through the demand substitution test. The difficulty of applying this test to personal data is highlighted by D. S. Tucker, H. B. Wellford, *Big Mistakes Regarding Big Data*, in *The Antitrust Source*, December 2014, 5 s.

Admittedly these semantic problems escape rather coarse and one-size-fits-all economic theory, but at least lawyers should be aware of the pit-falls when they enter in the "ownership" debate. Ownership is a concept quite different from *proprieté* or from *Eigentum*.

Furthermore, one should add that trying to assert an "ownership" over one's personal data is an attempt that (in continental Europe) not only totally ignores over 150 years of debate on personality rights (von Gierke's and Kohler's contributions being the starting point)[9], but even forgets the roots of continental legal systems: "*Dominus membrorum suorum nemo videtur*"[10]. Again, this property-like approach can be understood – but not justified – when it comes from wannabee-lawyers who are unaware of the essential bearings of a legal system, but is unacceptable when it comes from academic lawyers: juggling and jumbling with the letters of the juridical alphabet does not produce a work of legal literature.

With these premises, one should point out that what law-makers, lawyers, economists and stakeholders are searching for is legal certainty. Once data is under the control of a business there can be no doubt that it has the right to use, not use and exploit such data, being well aware that as that data is non-rival it might be in the availability also of some other entity: the typical example is that of statistical data acquired from a public body. Whether one uses trade secret rules, or the *sui generis* protection for data banks[11] whoever lawfully holds the data is entitled – therefore the term "entitlement" appears much more appropriate than ownership[12] – to use them[13]. Only in some, very limited cases, there may be an obligation, set

[9] See G. Resta, *Personnalité, Persönlichkeit, Personality. Comparative Perspectives on the Protection of Identity in Private Law*, in *Eur. J. Comp. L.*, 3, 2014, 215 ss.

[10] Ulpian, L. 13 pr. D. 9, 2 («Nobody can own one's own limbs»).

[11] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information (as amended by Directive 2013/37/EU). In the Big Data/competition context see the analysis of the relevant CJEU case-law by I. Graef, *Market Definition and Market Power in Data: The Case of Online Platforms*, in *World Competition,* 38, 2015, 473 ss., at 481.

[12] "Ownership" of data is thoroughly investigated (and challenged) by F. Mezzanotte, *Access to Data: The Role of Consent and the Licensing Scheme*, in K.S. Lohsse, R. Schulze-D. Staudenmayer (eds.), *Trading Data in the Digital Economy: Legal Concepts and Tools,* Oxford, 2017. Similar critical views are expressed by S. van Erp, *Ownership of Digital Assets and the Numerus Clausus of Legal Object*s, Maastricht European Private Law Institute Working Paper No. 2017/6 (1 October 2017).

[13] There may also be criminal law provisions prohibiting from accessing and copying data held by a third party: see § 502 of the California Penal Code which was the object the *Facebook v. Power Ventures* decision (USDC N.D. California 25 September 2013).

by the law to disclose such data or make it accessible to third parties. This happens typically with businesses which are entrusted with public services and with some financial services.

At any rate what one sees is that with data – as with most digital entities (*e.g.* software programmes) – there is a significant shift from the "sales" paradigm (complete, unlimited in time and not reversable transfer of rights from vendor to buyer) to a licence model: licensor allows licensee to use a certain entity, for a certain scope, for a certain time. There is no transfer but agreed access to a database, and eventually to data analytics facilities, with no right to further disseminate the data[14].

Even more radical is the model by which the entity that holds the data does not disclose it but simply allows – under consideration and stringent contractual terms and conditions – a third party to use its data analytics tools. This model is common in market research[15].

The final result is that "trade in data" tends to be rather limited (and generally not allowed in Europe owing to GDPR constraints) and does not include "big data"[16].

4. *"Data markets" or "data services"?*

One could conclude that a generic "data market" does not exist. What

[14] This tendency is widely examined in A. Perzanowski-J.M. Schultz, *The End of Ownership. Personal Property in the Digital Economy*, MIT Press, 2016. See also H. Zech, *Data as a Tradeable Commodity. Implications for Contract Law* (available at SSRN – Nov. 2017) (especially § 4).

[15] For these reasons the Arrow Information Paradox (see N. Duch-Brown, B. Martens, F. Mueller-Langer, *The economics of ownership, access and trade,* cit., at 46), appears to be of limited importance because there is no, or it is extremely controlled, release of information. One should also consider that the value of data – or what one can extract from it – is highly dependent on its timeliness. Obsolescence is often a matter of days if not of hours: I need to know here and now if the train or the plane is on time or is late. D. L. Rubinfeld, M.S. Gal, *Access Barriers to Big Data*, in *Ariz. L. Rev.*, 59, 2017, 339 ss. suggest (at 370) that unavailability of past data might not necessarily be a competition concern as firms «might also invest more resources in creating better analytical tools rather than in gathering more data». See also D. Bergemann, A. Bonatti, *Markets for Data*, Society for Economic Dynamics 2012 Meeting Papers 538.

[16] H. Zech, *Data as a Tradeable Commodity. Implications for Contract Law*, cit., points out (para. 5) that the service paradigm limits – as a default rule in German law, but in any case in standard form contracts in use in the trade – the possibility of transferring to a third party the right to take advantage of the service.

one can, and should, do is a careful process of distinguishing between the many sectors in which data are an essential element. So what one should be concerned with are not data per se, but rather the multifaceted services which require data for their functioning[17].

Therefore, one should consider that search engines are different from repositories which are different from social media which are different from travel and accommodation intermediaries which are different from etc. etc[18].

Looking at the different services one can understand if there are competitive constraints and if there are barriers to entry[19].

How relevant are data in this kind of examination? Again, it is important to distinguish. There are services whose main business is to collect data from users, data which subsequently are processed, aggregated, analyzed and then sold to third parties[20]. There are other services in which, together with the collection of data, the business "sells" its users to advertisers, commonly through banners, but most profitably by allowing the insertion of cookies, which allow the third parties to monitor user preferences and promote their goods and services.

There are also services for which data are a source of collateral revenue in respect of the core business – generally intermediation – which ensures

[17] This appears to be the approach taken by the EU Commission in the *IBM Italia/ UBIS* merger (COMP/M 6921, 19.6.2013) which was granted after analysis of the different services, and not of the databases they were using. The following *Facebook/ WhatsApp* merger decision (COMP/M 7217) focused more on the databases held by the two companies but concluded that there was no evidence of a dominance and that there was no down-stream data market. It is significant that the Italian Competition Authority sanctioned Facebook with a Euro 3 mln fine for this merger not for antitrust violations but for misleading consumer practices, as it had not notified WhatsApp users that their data would have transferred to (and processed by) Facebook (Decision PS10601, 11.5.2017). See also H. Zech, *Data as a Tradeable Commodity. Implications for Contract Law*, cit.

[18] Appropriately I. Graef, *Market Definition and Market Power in Data,* cit., at 479, suggests the need of careful distinguishing between the different kinds of data, the use that is made of them and the procedures used to analyse them.

[19] A competition issue arises when a company holds exclusive data banks as in the *Dun & Bradstreet/Quality Education Data* merger which brought to a FTC order of divestment which was accepted by D&B (see the 10 September 2010 Decision and Order, available online). D. L. Rubinfeld, M.S. Gal, *Access Barriers to Big Data*, *Ariz. L. Rev*., 59, 2017, 339 ss., point out competition concerns when a business controls up-stream production and provision of data in a certain sector, preventing competitors from creating a similar database. This brought the FTC in the *Nielsen/Arbitron* merger to issue an order (24 February 2014) to divest certain activities and licence access to certain data.

[20] « Data however are mostly intermediary goods that are used in production processes by other parties»: N. Duch-Brown, B.Martens, F. Mueller-Langer, *The economics of ownership, access and trade*, cit., at 28.

very high commissions. In order to understand the role of data one should therefore try to understand to what extent data are the source of revenue of the business and when availability of certain data gives the business a specific market power[21].

This appears to be – at least at a very first glance – the approach of the German Competition Authority (Bundeskartellamt*)* in its very recent *Facebook* decision according o which the company has a dominant position in the German market for social networks and therefore is subject to special obligations under competition law[22]. The decision, having taken what

[21] See the *Microsoft/LinkedIn* merger (Case M.8124, 6.12.2016) at § 179: «Assuming such data combination is allowed under the applicable data protection legislation, there are two main ways in which a merger may raise horizontal issues as a result of the combination under the ownership of the merged entity of two datasets previously held by two independent firms. First, the combination of two datasets post-merger may increase the merged entity's market power in a hypothetical market for the supply of this data or increase barriers to entry/expansion in the market for actual or potential competitors, which may need this data to operate on this market. Competitors may indeed be required to collect a larger dataset in order to compete effectively with the merged entity than absent the merger. Second, even if there is no intention or technical possibility to combine the two datasets, it may be that pre-merger the two companies were competing with each other on the basis of the data they controlled, and this competition would be eliminated by the merger». The Commission however concluded that in the specific case «the combination of their respective datasets does not appear to result in raising the barriers to entry/expansion for other players in this space, as there will continue to be a large amount of internet user data that are valuable for advertising purposes and that are not within Microsoft's exclusive control» (§ 180). *Distingue frequenter* is the caveat of I. Graef, *Market Definition and Market Power in Data,* cit., at 505: «A hypothetical or potential market for data can be defined by looking at the substitutability of different types of data and in particular at the functionality which can be offered with a specific set of data as input. In this way, separate relevant markets can possibly be defined for offline and online data and, as further subsegmentations within the latter market, for search, social network and e-commerce data».

[22] See the Bundeskartellamt decision of 7 February 2019 in case B6-22/16 (an English case summary is available online; see also the press release with comments by the Chairman of the Competition Authority). The decision cannot be commented in length in this paper. The doubts it raises are that the "social media" market is substantially tailored on Facebook, in the sense that the term "social media" is simply a synonym of Facebook. Further there is a debatable overlapping of rather different and distinct set of rules when the decision states that «The violation of data protection requirements found is a manifestation of Facebook's market power» and when it qualifies Facebook's terms and conditions as unfair [which they surely are, but this is the role consumer protection authorities]. Finally, when the decision states that Facebook has «gained a competitive edge over its competitors in an unlawful way and increased market entry barriers» it begs the question if, once have designed such a tailored market, there can be "competitors" and why data give Facebook a dominance. For an answer to these doubts one can refer to the proposal of doing without the definition of a relevant market and looking at competition "across-markets" analysed in depth by M.

might be called the "privacy short-cut", has prohibited Facebook from implementing its data processing policies.

A further classification is however necessary. Most of the data is generated by the users or – the difference is relevant – by the use of the service in itself which creates so-called meta-data[23]. This provision of data requires to be better analyzed. In some cases, the provision of data by users is ancillary to the service offered, such as in the case of travel and accommodation intermediaries[24].

In other cases, instead, the exchange is quite clear: the business is offering a service without monetary payment but receives as consideration the data of the users[25]. It took considerable time before law-makers realized that behind the so-called "free" provision of digital services there was a very elementary economic operation : the operator attracts users with its services and collects from them micro-data which the users consider of no economic interest[26], but once they are aggregated they allow extremely valuable profiling of the user and the creation of homogenous groups for marketing purposes.

This is a development of the commercial TV model, in which broadcasters bought/created programmes to attract viewers who were then "sold" to advertisers. Again, it took a few decades before law-makers understood the dynamics of so-called two-sided markets.

The difference is that with TV programmes there is no exchange between broadcasters and viewers (the latter can switch to a different channel or turn off the set when commercials are broadcast).

Maggiolino, *I big data e il diritto antitrust*, Milan, 2018, 264 ss. See also A. Pezzoli, *Big data e antitrust: una occasione per tornare ad occuparci di struttura?*, in V. Falce, G. Ghidini, G. Olivieri (eds.) *Informazione e big data fra innovazione e concorrenza*, Milan, 2018, at 253. But see *contra* the conclusions of M. Gambaro, *Big data, mercato e mercati rilevanti*, in V. Falce, G. Ghidini, G. Olivieri (eds.) *Informazione e big data fra innovazione e concorrenza*, Milan, 2018, at 208.

[23] I. Graef, *Market Definition and Market Power in Data,* cit., at 475.

[24] Aptly J. Drexl, *Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy*, MPI Research Paper no. 18-23, at 27 points out that there are several cases in which individuals pay a monetary consideration for receiving data-driven services (*e.g.* automobiles, sports wearable devices) and therefore the data they provide is not the counter-performance.

[25] « The collection of personal data consequently operates as an indispensable currency used to compensate the providers for the delivery of their services to users»: I. Graef, *Market Definition and Market Power in Data,* cit., at 477.

[26] See however the field research by S. Spiekermann, J. Korunovska, C. Bauer, *Psychology of Ownership and Asset Defense: Why People Value Their Personal Information Beyond Privacy*, 2012 (available at SSRN).

In the case of digital services, instead, in order to take advantage of the services offered, users must be constantly connected and therefore are paying the service with their data: pay-as-you-go.

As it is not a monetary exchange, one can see the transaction in two specular ways: the user's data are the *quid-pro-quo* for the services; and the services are the *quid-pro-quo* for the data. This last aspect is not adequately considered[27]. A data company in order to acquire its raw materials must buy them on the market. It generally does so by inducing users to use their services. Competition – and antitrust scrutiny – therefore is on, and between, the latter[28].

To present the economic reality more precisely: data companies manage to collect more data from users because they offer them more efficient and attractive services[29]. Users prefer one provider rather than another because for the price they pay (their data) they receive services which they value more. There does not appear to be a significant difference – from the point of view of the user – between the data which are provided (*e.g.* only general identification data; or data on preferences and localization, etc.). The price therefore is, subjectively, always the same[30]. And one should add that as data are non-consumable, non-rivalrous and continuously produced there is no limit to the expenditure of the user[31].

[27] See A. Metzger, *Data as Counter-Performance: What Rights and Duties Do Parties Have?*, 8 *JIPITEC*, 8(2), 2017, 1 ss.

[28] D. Auer, N. Petit, *Two-Sided Markets and the Challenge of Turning Economic Theory into Antitrust Policy,* in *Antitrust Bulletin*, 60, 2015, 426 ss. point out that in these cases «applying a [SSNIP test or a] "small but significant decrease in content quality" test would certainly prove a daunting task».

[29] I would beg to differ from the concern expressed by D. L. Rubinfeld, M.S. Gal, *Access Barriers to Big Data*, in *Ariz. L. Rev.*, 59, 2017, 339 ss. that «consumers may enjoy lower-priced and higher quality products that are intended to "lure them" to use particular online services» (at 375). This case appears to be a typical example of unchallengeable competition on the merits". Or should one envisage some sort of "predatory services"?

[30] «Contrary to usual economic transactions, users as suppliers of data cannot determine the amount and type of information they want to supply and do not have any influence on what they will get in return»: I. Graef, *Market Definition and Market Power in Data,* cit., p. 490.

[31] This economic approach is strongly countered by European data protection authorities which claim that personal data, being a fundamental right, cannot be used as valid consideration for the provision of digital online services. See the EU Article 29 Data Protection Working Party, *Guidelines on consent under Regulation 2016/679* (28 November 2017 – 10 April 2018): «As data protection law is aiming at the protection of fundamental rights, an individual's control over their personal data is essential and there is a strong presumption that consent to the processing of personal data

One should therefore carefully distinguish between services which are paid with a monetary remuneration, for which consumers/users must necessarily choose among many in accordance with their budget. And the services-against-data exchange where hypothetically the user/consumer can buy an unlimited amount with the same data[32].

Looking at things from the perspective of the enterprise not only data are an infinite resource but also, there is no limit to the expenditure capacity of users who, metaphorically, are all carefree billionaires[33].

If one considers the hundreds of popular Apps, one can see how, very practically, this market works. As consumers have not exhausted their spending resources, the barriers to entry on the market do not appear to be on the demand side.

The other – parallel, not alternative – system to acquire fresh and precious data is that of increasing the production of data exploiting new "data-mines". The internet-of-things (IoT) phenomena is a typical example: data are no longer produced by humans, but by objects which

that is unnecessary, cannot be seen as a mandatory consideration in exchange for the performance of a contract or the provision of a service» (§ 3.1.2) The argument however is not convincing: there are many fundamental rights which are commonly traded with the limit of their not complete forfeiture (may I refer to V. Zeno-Zencovich, *Limitazioni contrattuali alla manifestazione del pensiero*, in *Diritto dell'informazione e dell'informatica,* 1995, 991 ss., on contractual limitations to freedom of expression; and to Id., *Profili negoziali degli attributi della personalità, ibidem,* 1993, 545 ss., on the commodification of aspects of personality, typically image, name, privacy). The opinion of the Article 29 WG supersedes the more cautious preliminary opinion of the European Data Protection Supervisor on "Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy" (March 2014).

[32] «While natural persons have a right to reject cookies and other tools to collect their personal data in web browsing environments, for example in search engines, very few make use of that right and simply accept cookies because it is the lowest cost solution that enables them to benefit from access to online information sources»: N. Duch-Brown, B. Martens, F. Mueller-Langer, *The economics of ownership, access and trade,* cit., p. 31.

[33] For these reasons the traditional arguments on pricing (and on monopolistic surcharge) have little sense in personal data markets. In *LiveUniverse v. MySpace* (USDC C.D. California – 4 June 2007) Matz J. stated that «Indeed, market share can be measured by figures other than just sales or revenue» and suggests as a parameter to measure market power «advertising revenue generated from the number of visitors to the personal profiles and networks of friends generated with and contained within the social networking web platform». The decision however rejected the antitrust claim by LiveUniverse stating that there was no evidence of harm to consumers: «The content they created is still available, and readily accessible. Internet aficionados easily move from one website to another in seconds». Which is exactly what happened "one click away", having Facebook supplanted MySpace.

are connected to machines. One can reasonably expect that in the near future such form of production will be dominant, with significant changes in the market[34].

Services will move from what are now the most common devices (personal computers, tablets, handsets) to widespread consumer objects: in the first place, automobiles[35], then refrigerators, household appliances, and progressively all consumer goods including so-called wearables. The "smart houses" technologies are typically data-driven, in a circular process: data are necessary to provide new services; and new services generate more data.

5. *Two-sided markets*

This again suggests careful distinguishing when trying to define "data markets". Although it took a while before law-makers – especially in the field of competition – understood what two-sided markets were and how they disrupted acquired mental habits, now it is finally accepted that provision of digital services on the Internet generally generates a two-sided market. The provider collects data from users and sells targeted advertisement services to business that want to reach certain groups.

However, this is not always true, *i.e.* not all data markets are two-sided. In fact, the term appears to be used as a catch-name in improper contexts[36].

It is sufficient to go back to brick-and-mortar economic models: supermarkets, considered as physical platforms, are not the actors of a two-sided-market which sees on the one side producers of goods and on the other side consumers. A distributor is part of a vertical economic process that starts with the production of the raw materials and goes on with all the

[34] I. Graef, *Market Definition and Market Power in Data,* cit., at 486, points out the "S-curve" value of data. Not always more amounts of data create more value, while diversity and specificity may be more valuable.

[35] See the complex issue of collection of data by automobile producers in J. Drexl, *Legal Challenges,* cit., at 14 ss. and Regulation (EC) No 715/2007 which allows garages to access to vehicle repair and maintenance information.

[36] Quite correctly D. Auer, N. Petit, *Two-Sided Markets,* cit., point out the «the myriad of labels that have been tagged on 'two-sided markets' in subsequent [to Rochet & Tirole's seminal article] scholarship, possibly with the intention of better capturing the dynamics of those markets: "multi-sided platforms", "two-sided networks", "informational intermediation", or "two-sided strategies"» (p. 434); and that «The literature today displays a jungle of competing two-sided market models» (p. 460).

intermediate steps until the final product reaches the consumer[37].

One can therefore doubt that a digital distribution platform is always part of a two-sided market, especially when there are other, parallel, forms of distribution and when the consumer is using the platform simply to obtain more easily the product or the service he/she is seeking. Or in cases in which there is simply an exchange of data against services, and the service provider, subsequently, uses the data to provide separate and unrelated services to third parties. A typical example can be the "Street View" service offered by Google, which creates it sending vehicles with fish-eye lenses around a town and subsequently offers it to its clients who pay for it through their data.

And as digital platforms may be multi-service providers, not every service gives rise to a two-sided market. Clearly this should be considered when trying to assess what a data market" actually – and not in a pre-fabricated model – is.

One could venture the idea that simply because users pay the services they receive with data does not make this a "data market". The fact that ordinarily we buy products and services paying a monetary counterpart does not turn every market in a "money market".

[37] The statement made here does not ignore the significant debate on whether supermarkets are (Rochet & Tirole) or are not (Rysman) part of a two-sided market (for an extensive examination see D. Auer-N. Petit, *Two-Sided Markets,* cit., at 436 ss.). Setting aside complex and debatable theoretical analysis on so-called Coasian bargaining (Ronald Coase is a giant, not a God), Rochet & Tirole's definition ends up rendering any form of not vertically integrated distribution organization a two-sided market (whether the grocer's shop on the corner or the huge supermarket). If you buy wholesale you sell retail, and the price structure is quite easily set in a competitive environment both up-stream and down-stream. Shopping malls, instead, are quite different: the owner of the premise does not buy any product to resell it. He builds a facility that is rented to retailers, creating a physical market which attracts customers just as any market square does since the Middle-Ages. What digital service providers do is create a platform where they sell digital space and software programming to vendors who want to attract buyers (for hotel rooms, airlines tickets, any sort of product). From this point of view, one could distinguish the role of Amazon when it resells books that it has bought (operating as a bookstore) and when it enables the sale of products it does not hold (operating as an intermediary). See A. Hagiu, J. Wright, *Marketplace or reseller?,* Harvard Business School, WP 13-092, 31 January 2014. Quite appropriately Auer and Petit point out (at 438) that «The lack of semantic homogeneity in economic discourse may also be an explanatory factor» [of the differences]. Incidentally one can note that, at end of the day, the EU, with the PSD2 Directive, cut the gordian knot of one of the oldest "two-sided markets", that of credit cards, by setting the cost of intermediation that can be charged on the merchant.

6. *Legislative and regulatory constraints*

Although quite recent, "data markets" are not at all some kind of new world that awaits only to be conquered by economic forces for then being ordered though *ex post* competition rules.

What is objectionable in this approach is that notwithstanding the fact Ronald Coase's theories on institutional economics have been about for 80 years surprisingly – if not annoyingly, at least in Europe – quite often analysis of markets begins in an entirely theoretical vacuum and only after are adjusted accordingly to dreary reality.

Without delving too much in this topic, the former remarks suggest that "data markets" are, from the beginning, different according to the political, social, legal context in which they exist. A US "data market"[38] is different from a European "data market" which is different from a Chinese "data market". Surely there are some common features, but it is precisely looking at these features that it is possible to detect the institutional elements that differentiate the outcome.

It is therefore preferable to look first at what the context is, in order to understand to what extent the market conforms to it. This is ever more true considering that common wisdom tells us that new digital technologies have been so disruptive and economically profitable because they by-passed the regulatory framework set for analog technologies.

From a EU perspective it would be advisable to consider in first instance the already extremely complex interaction between IP rules, data protection regulations, public sector information, sectorial regulation (telecommunications, financial markets, services of general interest). Once one has mapped the normative scenario it is possible to investigate to what extent can market forces act and the role of competition rules.

The opposite approach is likely to arrive to the conclusion that regulatory exceptions have swallowed the free market rule.

7. *Intellectual property rights*

The first element that should be considered are intellectual property (and related) rights. All digital technologies are shielded by an IPR thicket

[38] For some very practical cases and scenarios see D. B. Hoffman, *Antitrust in the Financial Sector* (Fordham University speech, 2 May 2018).

made of patents, software and semi-conductor protection laws, trade secrets. This is not the place to discuss the merits (many) and demerits (in the same amount) of this situation. Clearly this is situation generated by heavy lobbying by industries – all industries – with little consideration by law-makers towards public and general interests. At any rate IPR and competition law have been engaging for a few decades a never-ending duel. It is not very realistic to expect it will end in data markets, while it appears easier to predict that non-circumventable IPRs will generate new forms of business that want to avoid being blocked from the outstart by judicial challenge.

More concretely if we are looking at Big Data one can reasonably say that they are part of the assets of the business that has collected them, and they are protected by general rules on ownership and trade secret on a firm's intangible patrimony. Data analytics operate on the basis of software protected by Sw laws. Algorithms are, expressly, not protected but it is doubtful one can legally force an entity to disclose them. If they were covered by patent law one could imagine compulsory licences, but they are not, which closes, from a legal point of view, *de lege lata* the discussion.

The best example – from a European perspective – of the contradictory trends in this field and of the difficulty – if not impossibility – of finding a balance between monopolistic pressures and pro-competition policies is given by the EU know-how and trade secret directive (2016/943). Its first recital is self-explanatory: «Businesses and non-commercial research institutions invest in acquiring, developing and applying know-how and information which is the currency of the knowledge economy and provides a competitive advantage. This investment in generating and applying intellectual capital is a determining factor as regards their competitiveness and innovation-related performance in the market and therefore their returns on investment, which is the underlying motivation for business research and development». There is an obvious, lip-service, reference (recital 38) to the general application of competition rules set out in Articles 101 and 102 TFEU[39].

[39] J. Drexl, *Designing Competitive Markets for Industrial Data – Between Propertisation and Access*, MPI Research Paper no. 16-13 (p. 67) points out that EU competition law «shows considerable shortcomings as regards the data economy: first, the requirement of market dominance in Article 102 TFEU considerably limits the scope of application of this rule and requires an often burdensome assessment. Second, it is quite uncertain to what extent Article 102 TFEU can be applied in cases in which, as will be frequently be the case, the data holder is not competing with potential customers in downstream data-related markets. Of course, Article 102 TFEU can also be relied upon to remedy

Among the protected trade secrets are «commercial data such as information on customers and suppliers, business plans, and market research and strategies» (recital 2), and this data may be processed by the trade secret holder in compliance with general data protection rules (recital 35).

From a very practical point of view this means that the database held by a business not only is protected by a trade secret but furthermore it may not be disclosed to third parties because of data protection limitations. The result is that any business, even one holding a dominant position, has a double defence against allegations of exclusionary practices concerning the data it holds[40].

8. *Personal data protection*

If IPRs strengthen the position of data companies, one must also consider that they operate in an, indirectly, highly regulated sector.

The first and most obvious – and nightmarish – constraint is set out by the General Data Protection Regulation (GDPR). If one looks at it not with the usual rhetoric of fundamental rights, but from an economic perspective, the GDPR tells us that:

Personal data (whatever this means: realistically any kind of data

excessive pricing. However, competition law enforcers can hardly be expected to act as price regulators in the data economy, which is characterised by information problems and huge uncertainties regarding the value of data».

[40] J. Drexl, *Legal Challenges,* cit., at 16 ss. stresses the role that freedom of information and free flow of information should have in regulating data markets. It should be noted, however, that – notwithstanding highfalutin proclamations by EU institutions – the law in action points in a significantly different direction: see *e.g.* the CJEU *Verlag Esterbauer* decision (C-490/14) asserting an exclusive right of the Land of Bavaria on its topographic maps and preventing the use of them for maps for cyclists; or the *Renckhoff* decision (C-161/17) stating an exclusive right of a photographer when his photo was used on a school presentation and put online. And even more significantly the Digital Single Market Directive (which should be finalized in early 2019) which creates litigation-prone regulation on text and data mining and confers upon press publishers the right to levy a remuneration for the further dissemination by information society providers. For these reasons advocating "data sharing" on the basis of FRAND principles (see H. Richter, P. R. Slowinski, *The Data Sharing Economy: On the Emergence of New Intermediaries*, in *IIC,* 50, 2019, 4 ss.; and G. Colangelo, *Accesso ai* Data *e e condizioni di licenza F/RAND*, in V. Falce, G. Ghidini, G. Olivieri (eds.), *Informazione e big data fra innovazione e concorrenza*, Milan, 2018, 135 ss.) appears to be wishful thinking.

remotely related to a physical person[41]) cannot be freely appropriated by data companies[42].

Personal data in order to be collected by data companies require the consent of the person to whom the data are referred.

The nature of this consent can be seen in various theoretical ways. From an economic point of view, it is the basis of a transaction "services against data".

The contractual approach has its advantages, but also its drawbacks, in the first place because the contract must conform to the GDPR; in the second place because a contract between a data company and a natural person is qualified, practically always (at least in Europe), as a consumer contract. Therefore, the data company must comply with the GDPR and with the over-arching (and subject to expansive interpretation) consumer protection regulation, in particular the part concerning unfair contractual terms and unfair commercial practices[43].

At any rate, whatever the legal quibbles over the meaning and scope of the GDPR, economic reality tells us that through the theory of consent (whether express or tacit) vast amounts of personal data are lawfully made available to the data collector[44].

[41] See J. Drexl, *Legal Challenges,* cit., at 3. See the CJEU decision in the *Breyer v. Germany* case (C-582/14, decided on 19 October 2016) where a dynamic IP address is considered "personal data".

[42] This specific aspect was considered by the EU Commission when granting the *Microsoft/LinkedIn* merger (Case M.8124, 6.12.2016 §§ 177-178).

[43] It is sufficient to peruse the general terms and conditions tucked away in an inconspicuous link at the bottom the home page of the main service providers to verify that they are a fair of unfair terms. From a competition point of view the most relevant are those that – directly or indirectly – determine a lock-in effect for users preventing them from transferring or even cancelling the data held by the provider (e.g. e-mail messages; texts, photos and videos posted on a repository). May I refer to V. Zeno-Zencovich, G. Giannone Codiglione, *Ten legal perspectives on the "Big Data revolution"*, in *Concorrenza e Mercato,* 23, 2016, 29 ss., p. 40 ss.

[44] See N. Duch-Brown, B.Martens, F. Mueller-Langer, *The economics of ownership, access and trade*, cit., p. 17: «The GDPR *de facto* (but not *de jure*) assigns property rights on personal data to the data collector, however limited they may be due to his fiduciary role. In reality, data subjects exchange their personal data in online markets, for example when they access "free" online services in return for letting the service provider or data controller collect some personal data. In these cases, the data subject retains the specific rights on his data as defined in the GDPR; the service provider acquires the residual rights». Incidentally one should point out that the GDPR determines a significant fragmentation of nominal entitlement on the same data. But as its aim is mostly ideological, this issue is generally ignored (for an analysis of personal data as form of commons may I refer to V. Zeno-Zencovich, *La 'comunione' di dati personali. Un contributo al sistema* dei *diritti*

The GDPR is a typical regulatory barrier to trade: the big tech data companies – practically all US – who want to do business in Europe must comply with EU laws and regulations, and most important of all, may not – at least from a legal point of view – export the data they hold in other countries[45].

One must consider that GDPR is only the first of a wide regulatory move to regulate "data markets"[46]. The data-protectionist approach will be enhanced when the socalled e-privacy regulation will be voted, as it is aimed specifically to data-companies who collect data over digital networks. This reminds us that markets work under the combined action of the business involved and of their clients, upstream and downstream, but the playing field is drawn by other actors, among which the legislature and regulators are the most important[47]. Paraphrasing Kirchmann's quote dating back to the mid-19th century: «Three lines from Parliament and entire markets go to the scrap-house»[48].

della personalità, in *Diritto dell'informazione e dell'informatica* 2009, 5 ss.).

[45] The 1995 GATS Treaty clearly could not envisage trade in data. There have been some attempts to overcome limitations, at least in an EU/US perspective: see the "European Union-United States Trade Principles for Information and Communication Technology Services" (4 April 2011) and the aborted TTIP Treaty (specifically the chapter on e-commerce and ITC services). But also, in the Canada-Europe Trade Agreement (CETA) the CJEU in its Opinion 1/15 found that data protection concerns had not been sufficiently taken into account. And attempts by Facebook to find a "convenient" regulator in the EU (the Irish Data Protection Commissioner) were rejected by the CJEU in its *Datenschutz Schleswig-Holstein* decision (C-210/16, decided on 5 June 2018).

[46] «Policy makers walk a thin line between enhancing privacy protection and not losing the social welfare benefits of data aggregation and overcoming anti-commons in data use» (see N. Duch, Brown-B. Martens-F. Mueller-Langer, *The economics of ownership, access and trade,* cit., at 34).

[47] For this reason, I would express some reservations on the notion of "Economics of Privacy" (see A. Acquisti, C. Taylor, L. Wagman, in *J. Economic Literature,* 54, 2016, 442 ss.). "Privacy" is an entirely legal institution. Its nature and its content depend on the will and the whim of legislatures and – in common law jurisdictions – of hundreds of courts. One can use the term as shorthand for "personal data" but then, necessarily, one has to delve in the intricacies of data regulation. In any case the use of the term "privacy" tends to perpetuate a 19th century notion (*à la* Warren & Brandeis) while the issue of data protection has rendered notions such as "seclusion" and «*la vie privée doit etre murée*» (Royer-Collard, 1819) marginal, and focuses on issues such as control, steering, and manipulation of society by public and private entities (see J. Drexl, *Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy*, MPI Research Paper no. 18-23, at 5: «mere economic criteria can no longer suffice to provide a policy framework for markets where privacy interests are particularly important»).

[48] «*Drei berichtigende Worte des Gesetzgebers und ganze Bibliotheken werden zu Makulatur*»

9. *Level playing fields?*

Personal data laws are not the only provision one must take into account. Data companies have, from a certain point of view, a privileged position in respect of telecommunication operators who are expressly prohibited from collecting, processing and reusing the traffic data they receive from their users. Article 6 of Directive 2002/58 is very clear: «Traffic data relating to subscribers and users processed and stored by the provider of a public communications network or publicly available electronic communications service must be erased or made anonymous when it is no longer needed for the purpose of the transmission of a communication without prejudice to paragraphs 2, 3 and 5 of this Article and Article 15(1)».

From a policy perspective the provision is very clear. From an economic point of view, it is puzzling. The only justification one can find is that it was set in an era when data-companies were still embryonal.

The sense (or the non-sense) of the provision is that the first generators of the data – all digital communications pass through telecommunication networks – may not extract informational value from such traffic, which instead is the wealth of the so-called over-the-top companies, to which the prohibition contained in Article 6 does not apply. From a practical point of view a phone-call with one's handset through a telecom operator is subject to the restrictive rule. The same conversation, through WhatsApp, is not. Which is something that makes no sense from a legal, regulatory and policy point of view.

Therefore, the first – and extremely relevant – barrier to access data-markets is set by the EU legislation which creates an uneven playing-field.

If the policy aim is that of reducing assumed market power of certain players, and encourage European enterprises to become data companies, surely the first move should be to abolish this barrier. This does not seem to be envisaged by the very recent European Electronic Communications Code which recasts previous legislative texts[49] or by the new Privacy and Electronic Communications Regulation. But following this rather blinkering approach, it is difficult to ask competition law (and authorities) to mend one's asymmetries. And speaking of asymmetries, a further –

(J.H. von Kirchmann entitled his 1847 lecture *Die Wertlosigkeit der Jusrisprudenz als Wissenschaft*' [The fallacy of law as science]. A re-edition of the lecture is published by Manutius Verlag, Heidelberg, 2000).

[49] Directive (EU) 2018/1972/EU of the European Parliament and of the Councilof 11 December 2018 establishing the European Electronic Communications Code (composed of "only" 326 recitals, 127 Articles and 12 Annexes).

indirect – one can be detected in the second Payment services Directive (PSD2), where it imposes upon financial institutions a duty to give access to financial data of their clients to payment providers who are not traditional financial institutions.

It is notorious that many data-companies will be providing payment services and therefore will be able to combine, inter-relate and cross-analyze the huge amount of data they possess with the extremely valuable data on financial transactions of their clients.

10. *Some conclusions*

"Data markets" are in magmatic phase, especially because it is doubtful that we dispose of adequate intellectual and methodological tools to describe them. One must therefore limit oneself to some very cursory conclusions[50]:

Data are an infinite resource, which is commodified and acquired by data companies through many economic models.

Users generate an infinite quantity of data and, as data are non-consumable and non-rivalrous, when data are the counter-performance for digital services they have no limits to their expenditure capacity.

Data resources are moving from user-generated production to object-generated production with a massive increase of the latter that will over-

[50] For quite the opposite conclusions see M.E. Stucke-A.P. Grunes, *Debunking the Mith over Big Data and Antitrust*, in *CPI Antitrust Chronicle*, May 2015 (2) (and more in depth in their book *Big data and competition policy*, Oxford, 2016). They are countered (always from a US perspective) by D.D. Sokol-R. Comerford, *Antitrust and Regulating Big Data*, in *Geo. Mason L. Rev.,* 23(5), 2016, 1129 ss., at 1161: «Antitrust law is ill-suited to police Big Data and its use by online firms. The empirical case for regulating Big Data as an antitrust concern is still lacking. Further, from a theoretical perspective, not enough work has yet been done to thoughtfully study and analyze how antitrust could, or should, be applied to specific issues involving Big Data. In fact, the lack of empirical evidence, robust theories, or, indeed, legal precedent suggests that there is no cause for concern in this arena with regard to antitrust law and Big Data. All that is available at present are general theories of exclusion applied to this new area. Until antitrust authorities can match theories of harm with specific factual circumstances and show negative competitive harm to consumers, the antitrust case against Big Data is a weak one». But one has seen in para. 4 how in the *Facebook* decision the German competition authority has leap-frogged these doubts by, substantially, qualifying violations of data protection rules as evidence of dominance and of market abuse.

shadow the former[51].

Data markets are extremely diverse in their structure and functioning. One of the most common features, from the production side, is that data companies enter into some kind of agreement, offering services or benefits in exchange in order to funnel towards them the product (data).

Far from being an unregulated market, data collection already has to take into account an extremely complex legal environment, where the most relevant rules are set by IP rights and by data protection laws.

The playing field for data markets is rather uneven and presents significant asymmetries among the various actors.

In this scenario, rather than a case for *ex post* competition remedies, it would appear that a holistic approach – that looks at the forest and not at the single tree – could be much more beneficial for fostering policy goals such as access, inclusion and innovation. More specifically, consumer welfare – inasmuch as it is encroached by massive appropriation of personal data by enterprises – appears to be more efficiently protected and pursued by pro-consumer *ex ante* and *erga omnes* regulations, such as injunctions and sanctions against unfair, deceptive and aggressive commercial practices[52].

[51] « The dynamics of the digital economy can hardly be measured with the traditional tools of of competition law. (…) the analysis is often a static snapshot analysis» (R. Podszun-S. Kreifelds, *Data and competition law*, in V. Mak, E. Tjong Tjin Tai, A. Berlee (eds.), *Research Handbook in Data Science and Law*, Cheltenham-Northampton, 2018, p. 195).

[52] «Case law does not support the contention that data collection is an antitrust problem. The nature of the relationship between platform users and data collectors is more likely to fall within the realm of consumer protection law (including privacy and data protection law) than competition law. Online data have generated unprecedented consumer benefits in terms of free online services, improved quality of services and rapid innovation. The ability to offer free services via monetization of data sales and advertising is mostly seen as a pro-competitive effect and not harmful from a competition perspective. The absence of monetization would reduce the volume and increase the cost of online services and reduce competition in product markets» (N. Duch-Brown, B.Martens, F. Mueller-Langer, *The economics of ownership, access and trade,* cit., p. 21). Similar, but distinguishing, opinions are expressed by M. Botta, K. Wiedemann, *EU Competition Law Enforcement vis-à-vis Exploitative Conducts in the Data Economy. Exploring the Terra Incognita*, MPI Working paper no. 18-08, p. 67: «The unilateral imposition of unfair contractual terms seems to be the most likely kind of exploitative conduct to be successfully prosecuted by an N[ational] C[ompetition] A[uthorities] in the near future».

ROMA TRE
UNIVERSITÀ DEGLI STUDI

These two volumes collect twenty five articles and papers published within the "Governance of/through Big Data" research project financed by the Italian Ministry of Universities. The research project, which was promoted by Roma Tre University, as project lead, and saw the participation of professors and reseachers from Bocconi University in Milan; LUMSA University in Rome; Salento University in Lecce and Turin Polytechnic, covers multiple issues which are here presented in five sections: Algorithms and artificial intelligence; Antitrust, artificial intelligence and data; Big Data; Data governance; Data protection and privacy.

Giorgio Resta and **Vincenzo Zeno-Zencovich** are full professors of comparative law in the Roma Tre University.

Roma TrE-Press
2023