QUADERNI DELLA RIVISTA

RIVISTA DELLA CORTE DEI CONTI

CORTE DEI CONTI

Intelligenza artificiale

Impiego e implicazioni sotto il profilo tecnologico, etico e giuridico

Quaderno n. 2/2024

Estratto da

La *Rivista della Corte dei conti* è a cura del Servizio Massimario e Rivista

La rivista è consultabile anche in:
www.rivistacorteconti.it

RIVISTA DELLA CORTE DEI CONTI

QUADERNI

Quaderno n. 2/2024

Direttore responsabile
Pres. Tommaso Miele

SOMMARIO

PER UNA GENEALOGIA CRITICA DELL'INTELLIGENZA ARTIFICIALE

di *Teresa Numerico* (*)

Abstract: L'articolo ha l'obiettivo di effettuare una genealogia dell'attuale contesto di successi dell'Intelligenza artificiale. Il testo offre una descrizione di tutte le alternative concettuali che ne hanno animato la storia lunga settant'anni. Dalla versione dell'IA di impostazione logicista, alla cibernetica, dal connessionismo, all'integrazione con la macchina, fino ad arrivare ai risultati più promettenti che si sono succeduti negli ultimi dieci anni che derivano da alcuni metodi che si basano sull'addestramento offerto da ingenti corpora di dati, con la finalità di estrarre regolarità e schemi noti come pattern da proiettare sulle serie future dei medesimi dati per prevederne l'orientamento e favorire l'intervento. Queste strategie note come machine learning e deep learning hanno ottenuto notevoli risultati, ma le loro applicazioni necessitano di una chiara funzione obiettivo. Senza di essa non abbiamo modo di valutarne gli effetti a causa dell'impossibilità di controllare i processi interpretativi e l'organizzazione del calcolo delle probabilità, dal momento che i dati sono troppi per controllarli a mano. Inoltre, questi metodi basati sull'apprendimento e l'addestramento attraverso i dati conducono a dei risultati che tendono alla standardizzazione e alla normalizzazione con la tendenza ad amplificare gli stereotipi già presenti nei corpora. Tale effetto è particolarmente eclatante nell'ambito dell'IA generativa. La produzione artificiale di contenuti non può nemmeno garantire la loro veridicità perché ciò implicherebbe una curatela dei dati di addestramento, impossibile considerata la loro enorme quantità. Chi usa questi strumenti è costretto a controllare i risultati, verificando le risposte per evitare le allucinazioni, o meglio gli errori che sistematicamente possono verificarsi, in modo imprevisto, ma costante. Questo dovrebbe farci riflettere sulle scelte politiche relative a quali compiti possano essere svolti da questi sistemi in modo autonomo e quali richiedano invece l'integrazione stretta con le capacità umane e quali debbano essere assunti esclusivamente da esseri umani, anche in relazione alle problematiche di responsabilità delle decisioni esercitate.

The article aims to perform a genealogy of the current context of success surrounding Artificial Intelligence. It offers a description of all the conceptual alternatives that animated its history over the past seventy years. From the logicist version of AI to cybernetics, from connectionism to integration with machines, it explores the most promising results that emerged over the last decade, derived from methods based on training with large data corpora, aimed at extracting patterns to project onto future series of the same data for prediction and intervention purposes. These strategies, known as machine learning and deep learning, have achieved significant results; however, their applications require a clear objective function. Without it, we cannot evaluate their effects due to the impossibility of controlling interpretative processes and the organization of probability calculations, given that the data volume is too large for manual control. Moreover, these data-driven learning and training methods lead to outcomes that tend toward standardization and normalization, amplifying existing stereotypes within the corpora. This effect is particularly pronounced in the realm of generative AI. The artificial production of content cannot guarantee its truthfulness, as this would imply curating the training data an impossible task, given their vast quantity. Users of these tools are compelled to verify results, checking responses to avoid hallucinations or systematic errors that can occur unexpectedly but consistently. This should prompt us to reflect on political choices regarding which tasks can be autonomously performed by these systems, which require close integration with human capabilities, and which should be exclusively undertaken by humans, especially in cases in which there are decision-making processes, that require a clear accountability structure.

(*) T. Numerico è professoressa associata di Logica e Filosofia della Scienza all'Università di Roma Tre.

Alcune parti di questo articolo sono una rielaborazione del primo capitolo del volume Big Data e algoritmi, Carocci, Roma, 2021.

1. *Introduzione*

Ad attribuire l'intelligenza alla macchina tra i pionieri dell'informatica aveva pensato Alan Turing (1). Il suo suggerimento, perché i dispositivi fossero accreditati come intelligenti, era che fingessero di maneggiare il linguaggio. Era l'obiettivo esplicitamente assunto dal suo famoso Test, o gioco dell'imitazione. Dalla sua attenzione per il linguaggio e per la sua difficile governabilità da parte della macchina si passò a una visione riduzionista dell'IA secondo cui le macchine sarebbero in grado di sostituire la mente attraverso un sistema fisico-simbolico di impostazione logicista. Contemporaneamente a questo paradigma della simulazione della mente, si sviluppa la cibernetica, un sapere transdisciplinare, il cui obiettivo è abbattere la differenza tra organico e inorganico attraverso l'analisi dell'attività degli agenti, considerati nelle loro prerogative di comunicazione e controllo. Il modello di integrazione tra vivente e macchina conduce a immaginare l'intelligenza meccanica come l'esito di un processo di progressivo ampliamento delle capacità umane, attraverso strategie di comunicazione e controllo.

La prospettiva dell'interazione uomo-macchina si è evoluta negli ultimi anni in una tendenza alla registrazione in forma di dati dell'azione umana, possibile grazie alla diffusione della rete Internet che ha consentito una registrazione di tutte le azioni compiute nell'ambito della comunicazione digitale. L'intelligenza artificiale si è trasformata, così, nella costruzione di algoritmi di apprendimento in grado di interpretare i dati ai fini di prevedere, anticipare e orientare il comportamento umano, dopo averlo discretizzato in un flusso di dati misurabili. Non si tratta più di una simulazione dell'attività umana, ma di una sua riproduzione, dopo averla reificata nella forma delle sue tracce online. La replica digitale serve per anticipare abitudini e preferenze delle persone, e per valutare e orientare le loro attività. La riproduzione che gli algoritmi di *machine learning* e *deep learning* prospettano riguarda sia l'essere umano come oggetto di analisi, sia l'essere umano in quanto decisore e interprete delle informazioni disponibili.

Il computer, infatti, a queste condizioni, svolge il ruolo per il quale è più adatto: quello di una macchina per eseguire calcoli. Dopo aver cercato, quindi, invano di esorcizzare il corpo chiudendolo tra parentesi, nell'ipotesi di riuscire a costruire sistemi simbolici del tutto privi di riferimento fisico, attraverso la mediazione della cibernetica, l'intelligenza artificiale si è riorientata verso l'interazione comunicativa tra esseri umani e macchine. Tale interazione in un primo tempo si è manifestata come integrazione in cui la macchina era sotto il controllo dell'operatore umano e poi, con l'avvento dei dati prodotti dalle azioni umane registrate, ha sussunto sotto di sé la capacità di prendere decisioni e quella di generare autonomamente contenuti, sia pure dipendendo, per ora, dall'operatore umano per la costruzione degli algoritmi per interpretare i dati. Tale evoluzione propone una serie di questioni di natura epistemologica, etica, sociale e politica che vale la pena analizzare.

La prima questione riguarda la presa di decisione. Non è facile, infatti, sottoporre al controllo di operatori umani le decisioni algoritmiche. Come suggeriva preoccupato Wiener, gli esseri umani rischiano di essere lasciati fuori dal *feedback* perché troppo lenti (2).

La seconda questione è che la macchina, come segnalava Turing (3), deve solo fingere di essere intelligente ed essere capace di ingannare giudici non esperti di tecnologia. Ma se non si può controllare quello che il dispositivo fa – e il meccanismo è programmato per ingannare gli umani – come possiamo fidarci delle sue decisioni?

2. *Quando è nata l'intelligenza artificiale?*

Il termine Intelligenza Artificiale è ormai diventato imprescindibile. Qualsiasi attività sembra poter essere aumentata, migliorata o sostituita utilizzando i suoi sistemi. Quando l'espressione fu coniata – come titolo del progetto del 1956 per un *Summer workshop* del Dartmouth College – il concetto non era ancora famoso (4). Il gruppo che firmò la proposta fu coordinato da John McCarthy, che nel 1955 richiese il finanziamento di un seminario ristretto a una decina di partecipanti alla Rockfeller Foundation. La scelta del termine avvenne proprio perché non si riferiva precisamente a nessuna delle ricerche in voga in quegli anni tra la cibernetica e la teoria degli automi. Diversamente dall'espressione meno impegnativa usata precedentemente da Alan Turing nei suoi articoli

(1) Cfr. Turing A. (1950), *Computing machinery and intelligence*, "Mind", vol. 59, 433-460; ristampato in Copeland (2004), *The essential Turing*, Clarendon Press, Oxford, 441-464 (trad.it. *Macchine calcolatrici e intelligenza*, in Lolli G. (a cura di) *Intelligenza Meccanica*, Bollati Boringhieri, Torino, 1994, 121-57).

(2) Cfr. Wiener N. (1950/1954), *The Human Use of Human Beings*, Houghton Mifflin, Boston (trad. it. i*ntroduzione alla cibernetica: l'uso umano degli esseri umani*, Bollati Boringhieri, Torino, 1966) e Wiener N. (1960), *Some Moral and Technical Consequences of Automation*, in "Science", New Series, vol.131(3410) (May 6, 1960), pp. 1355-8.

(3) Cfr. Turing A.M. (1948/2004), *Intelligent machinery*, in Copeland J. (Ed.) (2004), *The essential Turing*, Clarendon Press, Oxford, pp.410-432 (trad.it. *Macchine intelligenti*, in Lolli G. (a cura di) *Intelligenza Meccanica*, Bollati Boringhieri, Torino, 1994, pp.88-120) e Turing A.M. (1952/2004), *Can Automatic calculating machines be said to think?*, discussione registrata alla BBC il 10 gennaio 1952 tra Turing, Max Newman, Richard Braithwaite e Geoffrey Jefferson, pubblicata in Copeland J. (Ed.) (2004), *The essential Turing*, Clarendon Press, Oxford, pp.494-506.

(4) McCarthy J., Minsky M. Rochester N. Shannon C. (1955), *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, in "AI Magazine", vol.27 Number 4 (2006), 12-4; consultato il 29 settembre 2024) (trad. it. *Proposta di un Progetto di ricerca estivo sull'intelligenza artificiale presso il Dartmouth College*, in *Sistemi intelligenti*, XVIII, dic. 2006, 413-428).

sul tema (5) e cioè macchine intelligenti (*intelligent machinery*), *artificial intelligence* si riferisce alla possibilità che queste macchine di nuova creazione possano simulare l'intelligenza umana, non solo eseguire procedure considerate intelligenti.

Il progetto di McCarthy era ambizioso, ma inclusivo teneva insieme diverse attività di ricerca, come la capacità di simulazione di compiti intelligenti, le regole per la programmazione di un linguaggio in grado di eseguire manipolazioni simboliche, la creazione delle cosiddette reti neurali, che propendono per modellare il funzionamento delle attività cerebrali, piuttosto che dell'intelligenza *tout court*, i metodi di autoapprendimento delle macchine, la capacità di astrazione e creatività, oltre che i metodi per misurare la complessità di un calcolo.

Nonostante la fase embrionale nella quale le ricerche si trovavano, nel progetto iniziale emergevano già alcune delle grandi dicotomie che hanno caratterizzato la storia burrascosa di questa disciplina che – pur essendo giovane (non sono ancora 70 anni da quando il termine è stato coniato) – ha ricevuto moltissime attenzioni scientifiche e mediatiche, attraversando varie 'estati' e vari 'inverni'.

3. *L'intelligenza artificiale ha molte anime*

La prima contrapposizione riguarda l'oggetto della simulazione, cioè se le macchine debbano simulare il funzionamento delle capacità cognitive di alto livello, come la capacità di manipolare simboli e ottenere risultati nell'ambito della matematica, del *problem solving* o di altre discipline, formali e astratte, o se invece debba simulare o modellare il funzionamento cerebrale, molto più complesso, ma indipendente da meccanismi di manipolazione simbolica comprensibili.

La seconda opposizione metodica e contenutistica riguarda il ragionamento opposto all'apprendimento (*reasoning vs learning*). Se si debbano concentrare le forze a simulare il ragionamento, ovvero la capacità di trarre conclusioni affidabili da premesse prestabilite in base a regole di inferenza definite prima di cominciare a usarle, o se sia più produttivo e vantaggioso partire dai dati disponibili e cercare metodi per estrarre regolarità e modelli dinamici di funzionamento del fenomeno in esame. In questo secondo caso la macchina lavorerebbe su strategie di apprendimento, costruite dal basso verso l'alto, cioè a partire dai dati oggetto di investigazione, senza avere in anticipo una rappresentazione simbolica del problema da risolvere. Ci sono vantaggi e svantaggi in ognuna delle soluzioni. Il ragionamento permette di spiegare i risultati ottenuti dalla macchina, perché le sue regole sono esplicite, ma non riesce ad andare oltre il programma, e rischia facilmente di incorrere in una serie di potenziali esplosioni computazionali che rendono impossibile ottenere un risultato per problemi del mondo reale. L'apprendimento invece può fornire sempre risultati utili, ma non è sempre affidabile nel dare conto delle ragioni di una certa conclusione. Possiamo, quindi, ottenere risultati inattesi e sorprendenti, senza essere sicuri della loro correttezza e spiegabilità. Inoltre, i dati usati per addestrare i programmi di apprendimento sono parte integrante della loro capacità di trarre conclusioni, ma conservano una certa contingenza o, peggio, una traccia dei meccanismi rappresentativi iniqui che caratterizzano le nostre società, specialmente se si riferiscono alle soggettività umane e alle loro fragilità.

La terza dicotomia riguarda gli obiettivi da attribuire alle prestazioni delle macchine eventualmente intelligenti, l'alternativa allude alla possibilità di usare le loro attività per acquisire nuove conoscenze o per esercitare abilità creative, sorprendenti ma non sempre affidabili rispetto alla descrizione del mondo a cui si riferiscono.

La quarta alternativa riguarda la scelta di usare le macchine per sostituire gli umani, o per amplificarne le capacità, integrandole con quelle macchiniche. Se vogliamo sostituire l'attività umana ci immaginiamo dispositivi, dotati di autonomia organizzativa, ai quali delegare tutto il processo cognitivo; se invece pensiamo a strumenti di amplificazione e integrazione, immaginiamo delle interfacce che ci permettano di comunicare agevolmente, ma sulle quali non si possa scaricare la responsabilità delle decisioni operative e che funzionino solo in relazione a richieste e input governati dagli esseri umani. La premessa e la rilevanza di questi dispositivi è connessa con le eventuali capacità comunicative e protesiche rispetto alle attività da gestire e il loro design presuppone il controllo umano dell'azione.

In questo articolo ci concentreremo soprattutto sulla dicotomia tra *reasoning* e *learning*, oltre che sulla distinzione tra integrazione e sostituzione che ha attraversato il campo delle macchine intelligenti o delle macchine per l'amplificazione dell'intelligenza, per poi descrivere alcune delle strategie di maggior successo dell'Intelligenza artificiale basata sul *Machine Learning*, che sono al centro del grande successo che ha caratterizzato l'ultimo decennio.

4. *La cibernetica e la comunicazione con la macchina*

Il progetto dell'intelligenza artificiale era affiancato fin dal principio da un altro modello di sviluppo delle macchine, quello della cibernetica, un campo di ricerche transdisciplinare, inventato da Norbert Wiener. La cibernetica considerava le macchine come dispositivi indistinguibili da altri agenti organici. Il titolo del testo del 1948 che diede vita a questa transdisciplina era: *Cybernetics or Control and Communication in the Animal and*

(5) Cfr. Turing, A. *Computing Machinery*, op.cit.

the Machine (6). Il suo obiettivo era analizzare i dispositivi sotto il profilo della comunicazione e del controllo. Vi si rappresentava il controllo come un tipo particolare di comunicazione in cui era necessario non solo trasmettere un messaggio informativo, ma anche essere sicuri che l'altro agente avesse ricevuto ed eseguito l'ordine. Secondo tale rappresentazione esseri viventi e dispositivi meccanici potevano essere analizzati attraverso lo stesso modello teorico di comportamento.

Il modello si basava sul *feedback* che permetteva all'agente che riceveva il messaggio di rispondere adattando il proprio comportamento al messaggio ricevuto dall'altro agente o dall'ambiente stesso. Vista da questa prospettiva, la capacità degli esseri viventi di adeguarsi all'ambiente per sopravvivere poteva essere valutata in modo simile a quella delle macchine di obbedire agli ordini impartiti o, meglio, di comportarsi in maniera teleologica, essendo in grado di portare a termine lo scopo che veniva loro prefissato.

Il principale modello di riferimento della cibernetica era la recente esperienza bellica nella quale radar, missili balistici e altri dispositivi meccanici funzionavano attraverso la fissazione di un obiettivo balistico, mostrando così di possedere una loro finalità nel raggiungere lo scopo che era loro attribuito, qualunque fosse: dal riconoscere un velivolo nemico sul radar, a inseguire un obiettivo in movimento per colpirlo. Si trattava dell'anticipazione di missili, droni e bombe intelligenti che hanno in quel frangente la loro origine.

Wiener came to see the predictor as a prototype not only of the mind of an inaccessible Axis opponent but of the Allied antiaircraft gunner as well, and then even more widely to include the vast array of human proprioceptive and electro-physiological feedback systems. The model then expanded to become a new science known after the war as "cybernetics," a science that would embrace intentionality, learning, and much else within the human mind (7).

La tesi di Peter Galison – uno studioso di cibernetica membro della *Stanford School* (8)– è che la guerra avesse trasformato ogni nemico in un nemico meccanico:

On the mechanized battlefield, the enemy was neither invisible nor irrational; this was an enemy at home in the world of strategy, tactics, and maneuver, all the while thoroughly inaccessible to us, separated by a gulf of distance, speed, and metal. It was a vision in which the enemy pilot was so merged with machinery that (his) human-nonhuman status was blurred. In fighting this cybernetic enemy, Wiener and his team began to conceive of the Allied antiaircraft operators as resembling the foe, and it was a short step from this elision of the human and the nonhuman in the ally to a blurring of the human-machine boundary in general (9).

Se il nemico non era un altro essere umano, ma un agente strettamente connesso con una macchina allora, in un certo senso, anche le forze aeree alleate erano simili al nemico. Sul campo di battaglia avveniva questa elisione, questa confusione del confine tra essere umano e macchina in generale.

Tale sconfinamento della macchina nell'ambito del vivente era stato già suggerito da un articolo scritto da Wiener con altri due suoi colleghi, Arturo Rosenblueth e Julian Bigelow, dal titolo *Behavior, purpose, teleology* (10), proprio durante la Seconda guerra mondiale.

A further comparison of living organisms and machines leads to the following inferences. The methods of study for the two groups are at present similar. Whether they should always be the same may depend on whether or not there are one or more qualitatively distinct, unique characteristics present in one group and absent in the other. Such qualitative differences have not appeared so far. The broad classes of behavior are the same in machines and in living organisms (Rosenblueth, Wiener, Bigelow 1943, 21).

Nell'articolo si mettevano le basi per affermare che ci fossero differenze di rilievo nei metodi di studio relativi a esseri viventi e alle macchine dal punto di vista della comunicazione e del controllo. Tale conclusione, quindi, prometteva un abbattimento sostanziale della barriera che separava organico e inorganico. La prospettiva di un possibile studio congiunto di esseri viventi e macchine rispetto alla comunicazione – di cui il controllo era solo un caso particolare – consentiva di analizzare i dispositivi meccanici come casi analoghi a creature organiche, ma prospettava soprattutto la possibilità di un incontro, di una relazione, di un'integrazione tra viventi – in particolare umani – e macchine.

(6) Cfr. Wiener N. (1948/1961), C*ybernetics: or control and communication in the animal and the machine*, Hermann & Cie, Paris, MIT Press, Cambridge Mass (trad.it. *La Cibernetica*, Mondadori, Milano, 1968).

(7) Galison P. (1994), *The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision*, in "Critical Inquiry", vol. 21, No. 1 (Autumn), 228-66, cit. p. 229.

(8) La *Stanford School* è una scuola di filosofia della scienza di impostazione pluralista, cioè contro l'idea della scienza come un tutto unitario, i cui membri originari erano passati per la Stanford University tra gli anni Ottanta e Novanta. Le tesi di questo gruppo di studiosi si concentravano sulla natura plurale dei modelli scientifici e su un approccio post-positivistico basato sulle pratiche della conoscenza scientifica.

(9) Cfr. Galison P. *The ontology of the enemy*, *op. cit.* p. 233.

(10) Rosenblueth A., Wiener N., Bigelow J. (1943), *Behavior, Purpose and Teleology*, in "Philosophy of Science", vol. 10 (1) (Jan), 18-24, (trad. it Comportamento, scopo, teleologia, in Somenzi V., Cordeschi R., (a cura di) *La filosofia degli automi*, Bollati Boringhieri, Torino, 1994, 78-85).

Se le macchine potevano essere integrate negli esseri umani come una loro parte – come suggerivano le ricerche della cibernetica – il problema della radicale separazione tra loro sarebbe stato superato. Sarebbe stata, cioè, scavalcata la principale mancanza dell'intelligenza artificiale, quella legata alle capacità del corpo, a cui si sarebbe potuto sopperire attraverso una integrazione con il corpo dell'utente implicato nella macchina attraverso un'interazione con lei.

Tale progetto di integrazione tra esseri umani e macchine e della conseguente loro collaborazione basata sul principio della comunicazione e del controllo fu molto proficuo, più di quello della tradizione dell'intelligenza artificiale almeno fino agli anni Duemila. Dall'interazione uomo-macchina derivarono alcuni dei maggiori successi dell'informatica: dal sistema operativo che si propone come *user-friendly,* basato sulla programmazione orientata agli oggetti (*object-oriented programming*), che introduce icone, finestre, *groupware,* e la capacità intuitiva di cliccare con il mouse o col dito sulle icone. Tutte queste innovazioni derivano dall'intuizione della possibilità di una simbiosi umano meccanica suggerita da Licklider (11). L'integrazione con la macchina era alternativa alla sostituzione meccanica dell'attività umana nello svolgere compiti intelligenti, prospettata dall'IA. L'integrazione, infatti, prevedeva una collaborazione nella quale ciascun agente avrebbe mantenuto le proprie caratteristiche più qualificanti integrandosi con l'altro:

To think in interaction with a computer in the same way that you think with a colleague whose competence supplements your own will require much tighter coupling between man and machine (12).

La simbiosi con la macchina derivava dalla cibernetica come disciplina che si occupava della comunicazione e del controllo. Il contributo rivoluzionario all'IA della cibernetica consiste nella riflessione sulla centralità della comunicazione tra esseri umani e macchine, e dalle conseguenze di questa impostazione derivino i maggiori successi della disciplina in generale, e l'entrata in una nuova agenda di ricerca. Comunemente invece si ritiene che, mentre l'IA fosse concentrata sulla riproduzione della mente, la cibernetica fosse interessata alla riproduzione del cervello. Sebbene questi due orientamenti distinti fossero in atto, erano solo modelli astratti che la macchina si incaricava di simulare senza avere molto a che fare né con la mente, né con il cervello umano. Erano solo diverse impostazioni su come costruire un modello efficiente dell'attività umana (13). L'aspetto della riproduzione dell'attività neurale sarà poi oggetto del progetto di ricerca del connessionismo (14), che spesso viene associato allo sviluppo delle macchine intelligenti della cibernetica. Eppure, l'articolazione della *human-machine interaction* è stata l'eredità della cibernetica che ha avuto il maggior impatto sull'IA, perché consentì di lanciare un metodo per la cattura dei modelli cognitivi umani necessari per l'evoluzione delle macchine e la loro progressiva autonomizzazione.

Fu Licklider, inoltre, che pure aveva partecipato alle fasi progettuali della cibernetica (insieme a Robert Taylor, a promuovere concretamente l'idea del computer come strumento di comunicazione (15). I due sollevarono l'esigenza di un dispositivo in grado di connettere macchine differenti, capaci sia di interagire tra loro, sia di mediare con gli esseri umani. L'obiettivo delle macchine interconnesse era aperto a ogni possibile uso, anche quelli imprevisti o non ancora ipotizzati:

These new computer systems we are describing differ from other computer systems advertised with the same labels: interactive, time-sharing, multiaccess. They differ by having a greater degree of open-endedness, by rendering more services, and above all by providing facilities that foster a working sense of community among their users (16).

Tale articolo fu l'anticipazione della nascita di Arpanet, l'antesignana della rete Internet che si realizzò solo un anno dopo, sotto la guida di Taylor nell'ambito di un ufficio di ARPA (successivamente divenuta DARPA) chiamato Ipto (*Information processing technique office*).

A very important part of each man's interaction with his on-line community will be mediated by his OLIVER […]. At your command, your OLIVER will take notes (or refrain from taking notes) on what you do, what you read, what you buy and where you buy it. It will know who your friends are, your mere acquaintances. It will know your value structure, who is prestigious in your eyes, for whom you will do what with what priority, and

(11) Licklider J.C.R, (1960), *Man-Computer Symbiosis*, in *IRE Transactions on Human Factors in Electronics*, vol. HFE-1, 4-11; consultato il 20 settembre 2024).

(12) Cfr. *ivi*, 4.

(13) Vedi Cordeschi, R., (2002), *The discovery of the artificial*, Springer Netherland, Dordrecht, per una descrizione precisa dei diversi modelli di intelligenza in gioco all'epoca: dall'emulazione della mente attraverso un sistema fisico simbolico al connessionismo che invece si proponeva di emulare una rete neurale in cui i singoli nodi erano unità elementari che agivano in parallelo con tutti gli altri nodi collegati da una rete che si attivava a seconda della quantità di stimoli ricevuti.

(14) Vedi infra per maggiori dettagli.

(15) Cfr. Licklider J.C.R., Taylor R.W. (1968), *The computer as a communication device*, in "Science and Technology: For the Technical Men in Management", No 76, April, 21-31.

(16) *Ivi*, 31.

who can have access to which of your personal files. It will know your organization's rules pertaining to proprietary information and the government's rules relating to security classification (17)

La lungimiranza di Licklider e Taylor non solo anticipò e indirizzò la nascita della rete Arpanet, ma si spinse anche a introdurre la possibilità di ottenere dati e correlazioni sui comportamenti delle persone ai fini di anticiparne e orientarne i comportamenti. OLIVER, che rappresenta una specie di assistente personale onnisciente che i due studiosi prevedono, è l'antesignano di tutta la nuova serie di servizi e progetti legati all'analisi dei Big Data nell'ambito della ricerca sociale, oltre a incarnare un primo esempio di assistente personale digitale, e anche in un certo senso dei vari dispositivi di Intelligenza artificiale generativi del tipo di ChatGPT.

5. *Tra* reasoning e learning*: diverse strategie al confronto*

Per entrare nel merito della distinzione tra *reasoning* e *learning* che ha costituito la storia dell'IA fin dalle origini dobbiamo partire da quel filone che si riprometteva di simulare le reti neurali del cervello e le loro capacità computazionali sub-simboliche. A questa tendenza apparteneva Frank Rosenblatt che nel 1957 costruì il Perceptron (1958), seguendo la tradizione di ricerca che aveva mosso i primi passi con l'articolo di McCulloch e Pitts del 1943. L'intuizione del Perceptrone era costruire una rete neurale in cui ogni nodo era una versione molto limitata di un neurone che apprendeva attraverso un processo di prova ed errore (Somenzi e Cordeschi 1994, passim).

Il sistema, tuttavia, non funzionava molto bene, sia perché era troppo elementare, sia per mancanza di dispositivi hardware sufficientemente potenti da sostenere la richiesta di capacità di calcolo di cui la macchina avrebbe avuto bisogno per i risultati. Il progetto venne interrotto da un famoso libro di grande influenza scritto da Minsky e Papert nel 1969, dal titolo *Perceptrons*. In questo libro si dimostrava che l'uso di una rete a due livelli come era quella di Rosenblatt non avrebbe garantito di apprendere come categorizzare degli aspetti che si proponeva di riconoscere. Nel libro si sosteneva che, sebbene fosse possibile che una rete a più livelli potesse ottenere dei risultati interessanti, non si poteva mai assicurare che sarebbe stata in grado di addestrarsi per garantire risultati soddisfacenti. Il successo del libro fu complice di un cambio di direzione delle ricerche del settore che abbandonarono le reti neurali. Trascorsi alcuni anni, però, l'interesse per questi strumenti aumentò. Da un lato si rianimò il connessionismo – per il quale ci fu un primo revival negli anni Ottanta con la pubblicazione di *Parallel distributed processing* (18) (19) – dall'altro si attivarono gli studi sul *machine learning*, sia pure in contesti minoritari e con scarsi finanziamenti (20).

Le ricerche prevalenti nell'ambito dell'IA tra la fine degli anni Sessanta e gli anni Ottanta erano invece, fondate sulla conoscenza simbolica (21). Nell'ambito del *machine learning*, si usava l'apprendimento basato sulla conoscenza, che prevedeva, in assenza della grande quantità di dati, ora disponibile, di costruire basi dati di informazioni da inserire nel dispositivo, magari utilizzando un linguaggio di programmazione di tipo dichiarativo. Tale strumento consentiva di trarre inferenze per nuove scoperte, partendo dalle conoscenze inserite in memoria. Questo sistema, però, era lento, rigido richiedeva un grande costo computazionale, eppure fino agli anni Novanta era ancora molto in voga nell'ambito dell'IA. Veniva a volte chiamato, in modo un po' dispregiativo GOFAI (Good Old Fashioned AI), cioè Intelligenza artificiale Classica o all'antica (22). Fino a che non erano disponibili i Big Data, però, non c'erano alternative, perché il sistema delle reti neurali non funzionava bene, a causa dell'assenza dei dati per addestrare le reti e di quelli per testare l'addestramento. Con la disponibilità delle banche dati, rese accessibili attraverso la rete, e con lo sviluppo delle GPU (graphic processing unit), che si rivelarono più adeguati dei normali processori a gestire grandi capacità di calcolo, l'interesse per il *learning* aumentò. Un manipolo di studiosi che non avevano mai smesso di credere alle reti neurali divenne improvvisamente *mainstream*. Si trattava di Geoff Hinton, Yoshua Bengio, Yann LeCun i quali, non a caso, nel 2018 hanno ricevuto insieme il Turing Award, considerato il Nobel per l'informatica (23).

(17) *Ivi*, 38-39.

(18) Rumelhart, D.E., Hinton, G.E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, MIT Press, Cambridge, MA.

(19) Vale la pena segnalare che tra gli autori di questo volume c'è propri Geoffrey Hinton a cui nell'ottobre del 2024 è stato conferito il premio Nobel per la fisica, rendendolo il primo studioso ad aver avuto sia il Nobel sia il Turing Award per l'informatica. Hinton ha inventanto delle reti neurali artificiali sul modello di alcuni fenomeni della termodinamica, chiamate macchine di Helmholtz, oltre ad aver partecipato alla diffusione della tecnica della Backpropagation, uno degli algoritmi più utilizzati per addestrare una rete non supervisionata, ed essere a capo del progetto che nel 2012 costruì una rete non supervisionata Alexnet che aveva vinto la sfida di classificare le immagini correttamente promossa dalla banca dati ImageNet di cui parleremo più avanti.

(20) Cfr.Marcus G., Davis E. (2019), *Rebooting AI: building artificial intelligence we can trust*, Pantheon, New York, cap. 3.

(21) Cfr. Numerico T. (2021) *Big Data e Algoritmi*, Carocci, Roma, cap.1.

(22) Cfr. Boden, M.A. (2018). *Artificial intelligence: A very short introduction*. Oxford University Press, Oxford, Trad. it *Intelligenza Artificiale*, Il Mulino, Bologna, 2019.

(23) Cfr. Marcus, Davis Rebooting Ai, *op. cit.*, cap. 3, *passim*.

Uno dei motivi di grande successo del *machine learning* basato sulle reti neurali è che a fronte dei tanti dati disponibili per l'addestramento, la macchina non ha bisogno di troppo lavoro ingegneristico esperto perché viene poi affinata attraverso la grande quantità di tentativi che conduce per dare senso ai dati. Questo processo è molto diverso da quello in cui tutte le conoscenze dovevano essere inserite nella macchina sotto forma di sistemi esperti, come accadeva nel vecchio tipo di apprendimento in cui la base dati doveva essere esplicita.

Tuttavia, il risultato di Minsky e Papert non è mai stato confutato. Attraverso i meccanismi di apprendimento basati sulle reti neurali non possiamo essere sicuri dei risultati che otteniamo perché potrebbero non essere del tutto corretti, sebbene si possano ottenere risultati interessanti (24). Significa anche che se utilizzeremo questi sistemi di apprendimento in contesti delicati come cure mediche, o decisioni di giustizia sociale non potremmo rendere conto del perché di certi esiti e non potremmo essere sicuri che non contengano falsi positivi o falsi negativi, cioè situazioni nelle quali la valutazione è scorretta o perché considera la presenza di alcune caratteristiche nei soggetti sbagliati (falso positivo), o perché non si accorge che alcuni soggetti hanno una certa caratteristica (falso negativo).

6. *Machine learning*

La rilevanza delle tecniche del *learning*, e in particolare del *deep learning* hanno diverse ragioni, tra le quali spicca lo sviluppo di processori più performanti che in passato e la grande disponibilità dei corpora di addestramento, oltre agli algoritmi di apprendimento sviluppati nel corso di diversi decenni. Gli strumenti attuali sono gli eredi delle tecniche di reti neurali che simulavano il funzionamento cerebrale. Lo sviluppo degli algoritmi di apprendimento usati per addestrare i sistemi era già avvenuto negli anni zero del duemila, ma il 2012 costituisce un punto di svolta. Il sistema AlexNet (25) una rete di *deep learning*, vinse la famosa sfida per il riconoscimento di immagini (26), distaccando notevolmente il secondo classificato.

I sistemi di *machine learning* e specialmente quelli di *deep learning*, che hanno avuto così tanto successo hanno tre principali obiettivi: azioni di riconoscimento, attività di classificazione con l'obiettivo della previsione e del contestuale supporto alla presa di decisione o *automated decision making*, compiti generativi, come la produzione di contenuto testuale o multimediale generato a partire da un prompt di ricerca.

Si tratta di attività differenti che, però, hanno in comune alcune caratteristiche. Affinché questi sistemi abbiano successo è necessaria la disponibilità di grandi quantità di dati pertinenti all'ambito a cui il sistema si applica. Per il riconoscimento di volti o immagini è necessario disporre di un database di immagini adeguatamente etichettato, per il supporto alla presa di decisione è necessario un database di informazioni rilevanti sulle quali addestrare il sistema a estrarre le regolarità e proiettarle sul futuro dopo aver creato adeguati criteri per classificare i dati. E abbiamo bisogno di un grande *corpus* testuale e di immagini e di file multimediali per estrarre nuovi prodotti multimediali o i nuovi testi richiesti.

7. *La centralità dei dati*

Le diverse applicazioni, quindi, condividono la richiesta di una disponibilità adeguata di informazioni pertinenti sulle quali esercitare gli algoritmi per scegliere come classificarli per estrarne le caratteristiche utili al riconoscimento, alla decisione o alla produzione. Tali dati contribuiscono, talvolta in modo estremamente rilevante a orientare i risultati di questi sistemi, e sono parte integrante del processo cosiddetto 'intelligente'.

La prima questione critica di queste procedure riguarda la disponibilità dei dati. Sappiamo che la rete internet e i suoi servizi e specialmente l'avvento delle piattaforme hanno consentito la raccolta sistematica di grandi quantità di dati in formato digitale. Tale opportunità ha offerto ai possessori delle piattaforme, e a coloro che da loro si rifornivano, di costruire un nuovo spazio di appropriazione nel quale vigesse una completa assenza di regole di tutela. Di chi sono, infatti, i dati personali disponibili in formato digitale sulle piattaforme proprietarie? Nei contesti scientifici raccogliere dati è costoso, complicato e obbedisce a precise regole molto stringenti, ma i dati sperimentali garantiscono che chi li usa abbia anche contribuito a crearli, abbia seguito degli standard disciplinari per la loro raccolta e li metta a disposizione della collettività dei propri pari per ulteriori verifiche e per replicare gli esperimenti. I dati digitali non seguono queste regole perché la raccolta avviene considerando appropriabili i contenuti resi digitali e non si riferisce a nessuna pratica ufficiale per la costruzione di campioni effettivamente rappresentativi dei fenomeni che sarebbero oggetto della ricerca. Abbiamo quindi un ribaltamento delle regole per la raccolta dei dati e osserviamo un processo di appropriazione e di conseguente estrazione di valore senza

(24) *Ibidem*.

(25) Cfr. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. (2012) Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems* 25, pubblicato ufficialmente in *Communications of the ACM*, *60*(6), 84-90, 2017. Si tratta di un articolo cruciale sia per il risultato dirompente nella sfida del riconoscimento di immagini lanciata nel concorso di ImageNet voluto dal Lin Fei Fei, sia perché si usava il sistema dei grafical processing unit, i processori grafici che diventeranno da allora in poi lo standard per il funzionamento dei sistemi di *deep learning*. Erano, cioè, due innovazioni in una gli algoritmi di convolutional neural network usati per il riconoscimento delle immagini e l'infrastruttura tecnica per il processamento dei grandi dati necessari allo scopo.

(26) Il nome della sfida è: *ImageNet Large Scale Visual Recognition Challenge*.

giustificare il tipo d'uso che viene effettuato, né rispetto a chi li ha prodotti, né in relazione al rispetto di standard scientifici (27).

Il ruolo centrale svolto dai dati per l'addestramento nel machine learning nell'esercitare le valutazioni impedisce la spiegabilità dei risultati ottenuti, soprattutto quando l'obiettivo del processo di apprendimento non può essere chiaramente esplicitato, come invece accade quando la funzione target di un processo algoritmico di addestramento riguarda la vittoria nei giochi come scacchi e go. Il sistema algoritmico del *deep learning* è fatto in modo da non esercitare un controllo diretto sugli strati interni della rete, conseguentemente la relazione tra processo di addestramento e suoi esiti si limita a considerare solo le prestazioni dell'output. Le procedure del *deep learning* non rispettano una delle caratteristiche definitorie degli algoritmi: la riconoscibilità del rapporto tra un passo e i suoi effetti sulla procedura.

Un altro aspetto su cui vale la pena riflettere criticamente è la relazione che intercorre tra conoscenza e creatività nell'uso degli strumenti digitali. Già Turing nel 1950 (28) affermò che, se vogliamo che le macchine abbiano comportamenti intelligenti, non possiamo impedire loro di commettere degli errori.

Nelle attività di riconoscimento la questione dell'errore si pone per il fatto che i processi algoritmici preposti a questo profilo, oltre a dipendere dall'accuratezza della base dati, funzionano come procedure di discriminazione che selezionano una certa forma discriminandola dalle altre in base alla differenza. Il processo umano di riconoscimento è diverso perché funziona per incremento di dettagli e non per selezione di forme a partire dalle differenze (29). Tale caratteristica implica necessariamente la possibilità di giudizi discriminatori che eliminano soggettività o aumentano gli errori in certi ambiti in cui le forme non si possono costituire per assenza di dettagli disponibili. È esemplare, a questo proposito, la difficoltà dei software di riconoscimento facciale nel distinguere le donne afroamericane, per le quali forse non ci sono abbastanza dettagli (30).

L'altro problema molto rilevante riguarda la capacità di questi sistemi di prevedere il futuro e intervenire a valutare o a manipolare le attività dei singoli soggetti di cui si detengono i dati come conseguenza della previsione.

Sappiamo che l'uso dei sistemi di supporto alla presa di decisione si basa sull'esercizio della statistica e sul principio di induzione che nulla possono dire sul caso singolo, e non possono garantire la verifica delle predizioni future, soprattutto se hanno per oggetto le persone singole, come nel caso in cui siano implicati meccanismi sociali come l'accesso al welfare o al credito, la definizione del premio assicurativo, la valutazione del rischio di pericolosità, il reclutamento del personale.

8. *Criteri di validazione della conoscenza*

Il futuro è incerto e, soprattutto rispetto ai singoli soggetti, nulla valgono le previsioni collettive, eppure tali previsioni esercitano un enorme potere a causa della loro provenienza da fonti accreditate che le trasformano in prescrizioni che cambiano la vita delle persone sulle quali vengono esercitati i giudizi (31).

Nella discussione tra conoscenza e creatività dobbiamo ammettere che questi sistemi non possano essere accreditati come strumenti conoscitivi, almeno rispetto all'idea che ne abbiamo in relazione ai criteri di validazione adottati a partire dallo sviluppo della scienza moderna. Nonostante l'instabilità e l'incertezza o, peggio, il potenziale discriminatorio di queste previsioni, tendiamo a pensare di poter delegare a questi sistemi di supporto alla presa di decisione la responsabilità di scelte dolorose e intrusive relative a diritti e equità delle persone che vivono nei contesti democratici. Oltre al problema che queste decisioni vengono prese da una scatola nera, cioè da un sistema algoritmico opaco e inaccessibile, abbiamo la questione della legittimità di proiettare il passato sul futuro nel caso di situazioni contingenti e imprevedibili, laddove l'addestramento basato sui dati concreti potrebbe naturalizzare differenze e marginalità che sono storicamente il frutto di passate ingiustizie.

Diverso sarebbe il discorso se questi strumenti di supporto decisionale fossero solo affiancati a decisori umani, costruendo percorsi di interazione esplicita e fornendo le chiavi di comprensione dei processi inferenziali che sono al lavoro nei sistemi algoritmici usati per l'apprendimento delle regolarità dei dati, detto *pattern recognition*.

9. *Intelligenza artificiale generativa: la creatività dei modelli*

Gli ultimi clamorosi risultati dell'intelligenza meccanica riguardano i cosiddetti *foundation models*, cioè quegli strumenti che, basandosi su grandi corpora di immagini, suoni e testi sono in grado di restituire contenuti creativi come output di un prompt di richiesta più o meno dettagliato. Il successo degli ultimi due anni è dovuto allo sviluppo di una tecnologia chiamata transformer che è il frutto di un articolo citatissimo dal titolo *Attention is all*

(27) Beaulieu A., Leonelli S. (2022), *Data and Society*, Sage, London.

(28) Turing A. 1950, *Intelligent Machinery, op. cit.*

(29) Chun, W.H.K. (2021). *Discriminating data: Correlation, neighborhoods, and the new politics of recognition*. MIT press, Cambridge (Mass.).

(30) Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification, in *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.

(31) Hildebrandt, M. (2021). *The issue of bias. The framing powers of machine learning* (dec. 2019). In Pelillo M., Scantaburlo T. (eds.) *Machine we trust. Perspectives on dependable AI*, Mit Press, Cambridge (Mass.).

you need (32). In questo testo si sottolinea l'importanza dell'interazione con la richiesta, il prompt, a cui si deve dare una speciale attenzione.

In questo caso siamo di fronte a un dispositivo capace di invenzione, privo di velleità di tipo conoscitivo, che si nutre ancora una volta della disponibilità di contenuti digitali, come dimostra la scelta di tanti giornali come il New York Times o il Manifesto di negare l'accesso ai propri dati per l'addestramento. Il grande successo di questi strumenti come ChatGPT che ha avuto in pochi giorni più di 100 milioni di utenti nel novembre 2022 riposa anche sulle capacità mimetiche di questi sistemi che spesso finiscono per restituire direttamente parti dei corpora di addestramento. Il funzionamento di questi metodi per ottenere la creatività dipende dalla previsione statistica di una rosa delle risposte più frequenti in relazione al prompt di ricerca. Si tratta, quindi, di riprodurre un'aspettativa il più possibile prevedibile. Alcuni studiosi si spingono a ritenere che la produzione di testi e altri contenuti prefiguri l'attivazione di una vera autonomia soggettiva da parte del sistema (33), ma le frequenti allucinazioni, la mancanza di conoscenze di tipo comune, l'assenza di rapporto con il mondo esterno rendono il dispositivo produttivo, ma difficilmente capace di una vera intelligenza. In relazione a questa produzione di contenuti originali, entro certe definizioni di intelligenza, potrebbe risultare plausibile includere anche gli agenti sociotecnici implicati nella produzione di contenuti nella lista di quelli intelligenti, sebbene questo concetto non si possa precisare in modo univoco.

Questi strumenti aprono però nuovi scenari critici in quanto implicano l'appropriazione dei corpora e lo sfruttamento dell'esercito di persone coinvolte nella loro annotazione e nel processo di *Reinforcement learning by human feedback* (RLHF) che include il lavoro di alto livello di coloro che si esercitano a migliorare la qualità delle risposte fornite. Ma soprattutto attivano una nuova difficoltà per chi ne fruisce: quella di riconoscere se immagini, voci e descrizioni riguardano situazioni concrete di cui sono la testimonianza o introducono contenuti sintetici che nulla hanno a che fare con eventi e situazioni reali. Se non riusciremo a contenere quest'esplosione di ambivalenza, potremo trovarci nella situazione di non avere più una rappresentazione del mondo condivisa, sia pure criticamente, potremmo invece assistere all'esplosione di una molteplicità di rappresentazioni del mondo simili a quelle causate dalle teorie del complotto e dai terrapiattisti che per adesso si limitano a identificare minoranze prive di presa concreta sul mondo. Difficile dire come validare i contenuti per distinguere quelli immaginari e quelli oggetto della descrizione concreta del mondo, soprattutto per la potenziale esplosione quantitativa di immagini, suoni e testi, che risulterebbero impossibili da controllare attraverso la capacità cognitiva umana.

Possiamo anche riconoscere, nelle scelte inferenziali dei sistemi generativi, esiti discriminatori e diffusori di stereotipi. Osserviamo, come è segnalato in alcuni articoli del gruppo di Sasha Luccioni (34) di *Hugging Face*, che ad alcune richieste del *prompt*, relativamente alle immagini di persone che ricoprono certe professioni, corrisponda una rappresentazione pregiudiziale, etnicamente orientata che esibisce delle esplicite distorsioni della realtà a favore di standard, concentrati su una visione tradizionalista. La ricerca valuta il risultato di *Text to image*, cioè risposte a *prompt* testuali con immagini dei sistemi commerciali di IA generativa più noti, in relazione a diverse professioni rispetto a genere e appartenenza etnica delle persone che vengono riprodotte nelle immagini richieste, osservando il carattere profondamente discriminatorio e l'esclusione di identità marginalizzate in relazione a certe professioni, considerate più rispettabili o di successo. Tali risultati sono ancora più penalizzanti rispetto alle soggettività più fragili, tanto da superare, nella rappresentazione standard, la fenomenologia concreta della presenza reale etnica e di genere di professioni di alto livello come giudici, medici, manager ecc. La distorsione rappresentativa dei dati sintetici generati dalle richieste degli utenti amplifica la visione tradizionale del mondo polarizzandone i contorni.

Gli stessi progettisti dei sistemi generativi devono segnalare il carattere allucinatorio di alcune risposte, proponendosi di migliorare le prestazioni con le versioni successive dei *software* (35). Ha colpito la fantasia collettiva l'*exploit* del *chatbot* di Google, Gemini, che suggeriva di aggiungere la colla alla salsa della pizza (36), tuttavia, usare il termine "allucinazioni" potrebbe essere fuorviante. Il significato di questo termine, infatti, si riferisce a errori di carattere precettivo dei dispositivi che potrebbero contribuire ad amplificare la retorica della quasi umanità dei sistemi generativi. Forse varrebbe la pena chiamare i risultati erronei con il loro nome come provano a

(32) Vaswani, A. Shazeer, N, Parmar, N. U. et al *Attention is all you need*, Advances in Neural Information Processing Systems. 30. Curran Associates, Inc.

(33) Cfr. Roncaglia G. (2023) *L'architetto e l'oracolo*, Laterza, Roma-Bari.

(34) Cfr. Luccioni, A.S., Akiki, C., Mitchell, M., & Jernite, Y. (2023). Stable bias: Analyzing societal representations in diffusion models. arXiv preprint arXiv:2303.11408. Per un articolo più divulgativo sull'argomento dallo stesso gruppo di ricerca, rimando al blog dell'azienda francese Hugging face, per la quale lavora Sasha Luccioni, che è la principal investigator di questo progetto su IA generativa e pregiudizi strutturali, questo il post sulla newsletter aziendale: Sasha Luccioni et al. (26/06/2023) Ethics and Society Newsletter #4: Bias in Text-to-Image Models.

(35) Cfr. Prabhakar Raghavan Gemini image generation got it wrong. We'll do better, Google Blog 23/02/2024, ma secondo i commentatori ancora a maggio 2024 le cose per il generatore di immagini Gemini di Google non sono migliorate, vedi Kyle Wiggers, *Google still hasn't fixed Gemini's biased image generator*, TechCrunch, 15/05/2024, e più in generale Alex Cranz, *We have to stop ignoring AI's hallucination problem*, The Verge, !5/05/2024.

(36) Cfr. Elizabeth Lopatto, *Google still recommends glue for your pizza*, The Verge, 12/06/2024.

fare un gruppo di studiosi che titolano il loro articolo *ChatGPT is bullshit*. Secondo questo lavoro – ma non sono gli unici o i primi a notarlo – attribuire all'IA generativa delle allucinazioni è mistificatorio e scorretto. "Riteniamo che queste falsità, la generale attività dei *large language models*, si possa comprendere meglio come *bullshit* […] i modelli sono, in modo rilevante, indifferenti alla verità dei loro risultati" (37). Varrebbe la pena riflettere sul perché tutto questo clamore per dei sistemi non interessati o impossibilitati a rispondere correttamente alle nostre domande, che richiedono una costante e attenta capacità di *editing* e di correzione attiva, oltre che orientati dai dati e dai vincoli loro imposti a veicolare una rappresentazione della società e delle sue relazioni che restituisce le preferenze dei più forti, amplificandole rispetto alle soggettività e alle loro condizioni di privilegio o di marginalità.

10. *La politica dei sistemi intelligenti*

Le trasformazioni promosse dall'uso degli strumenti di intelligenza artificiale hanno un carattere fortemente politico. La scelta di delegare decisioni, soprattutto quelle che riguardano il futuro, come nel caso della giustizia predittiva non può essere risolto da meccanismi di automazione. Ci troviamo in contesti incerti e contingenti. Cedere il passo a oracoli instabili e incomprensibili non riguarda tanto una loro presunta superiorità, ma un eventuale spinta, più o meno consapevole, a evitare di assumere la responsabilità delle scelte da esercitare in contesti imprevedibili.

Un aspetto cruciale da considerare è la rilevanza dell'addestramento sui dati disponibili che contengono le disparità e le iniquità che sono attualmente vigenti nel mondo, per questo avere sistemi che inferiscono proiezioni future a partire da questi non è troppo rassicurante come mostrano molti studiosi (38).

Quanto alla volontà di affidarsi a sistemi generativi per la produzione dei contenuti, porterà molti cambiamenti nelle attività intellettuali, ma non è detto che rappresenterà un miglioramento delle condizioni di vita delle persone. Il processo di raccolta dei dati personali di miliardi di persone costituisce la condizione di possibilità di questi strumenti, ma si tratta di meccanismi di appropriazione di risorse comuni. Verrà un momento in cui questo passaggio incongruo diverrà chiaro per tutti. La possibilità di uso di certi strumenti non dà di per sé il diritto di esercitarla. Ci sono regolazioni che impediscono di usare i dispositivi alla loro massima potenza, come il codice stradale, o le regole sul porto d'armi. Sappiamo bene che non basta che una cosa sia possibile per essere autorizzata. Sebbene il contesto della raccolta dei dati sia complesso per il fatto che l'appropriazione si svolge in luoghi lontani dalla giurisdizione dove valgono le conseguenze legate all'estrazione di valore da quei contenuti.

Come suggeriscono acutamente Acemoglu e Johnson (39) non tutte le trasformazioni tecnologiche portano con sé un progresso per la collettività. Dobbiamo compiere delle scelte politiche ed epistemologiche per orientare queste innovazioni affinché possano migliorare le condizioni di vita degli umani e in generale dei viventi.

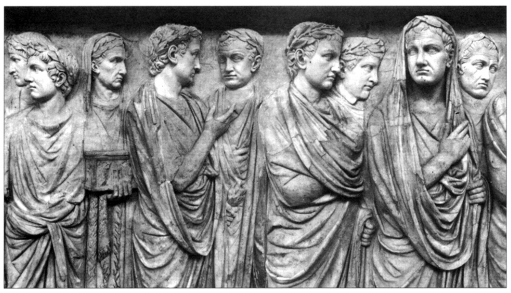
*** * ***

(37) Hicks, M.T., Humphries, J. & Slater, J. ChatGPT is bullshit. Ethics Inf Technol 26, 38 (2024).

(38) Qui ci limitiamo a citare alcuni testi estremamente espliciti a questo riguardo, Couldry N., Mejias U.A. (2019), *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*, Stanford University Press, Stanford (ca), trad.it. *Il Prezzo della connessione*, Il Mulino, Bologna, 2022, D'ignazio, C., & Klein, L.F. (2023). *Data feminism*. MIT press, Cambridge, Crawford K. (2021) *Atlas of AI*, Yale University Press, New Haven e London, Trad.it. *Nè intelligente, né artificiale*, Il Mulino, Bologna, 2021, Osoba O.A., Boudreaux B., Saunders J. et al. (2019), *Algorithmic Equity: A Framework for Social Applications*, RAND Corporation, Santa Monica, CA.

(39) Acemoglu D., Johnson S. (2023) *Power and progress, PublicAffairs*, New York, Trad, it. *Potere e progresso*, Il Saggiatore, Milano, 2023.

Direzione e redazione
Via Antonio Baiamonti, 25 - 00195 Roma - tel. 0638762191 - E-mail: massimario.rivista@corteconti.it



ISBN 978-88-498-5743-6

9 788849 857436

I.S.S.N. 0392-5358
(c.m. 30-443900120101)