



A closer look at ChatGPT's role in concept map generation for education

Daniele Schicchi, Carla Limongelli, Vito Monteleone & Davide Taibi

To cite this article: Daniele Schicchi, Carla Limongelli, Vito Monteleone & Davide Taibi (30 Apr 2025): A closer look at ChatGPT's role in concept map generation for education, Interactive Learning Environments, DOI: [10.1080/10494820.2025.2497110](https://doi.org/10.1080/10494820.2025.2497110)

To link to this article: <https://doi.org/10.1080/10494820.2025.2497110>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 30 Apr 2025.



Submit your article to this journal [↗](#)



Article views: 311



View related articles [↗](#)



View Crossmark data [↗](#)

A closer look at ChatGPT's role in concept map generation for education

Daniele Schicchi ^a, Carla Limongelli ^b, Vito Monteleone ^a and Davide Taibi ^a

^aInstitute for Educational Technology, National Research Council of Italy, Palermo, Italy; ^bEngineering Department, Roma Tre University, Rome, Italy

ABSTRACT

Rapid progress in natural language processing (NLP) has ushered in a new era of artificial intelligence (AI) models. Large language models (LLMs), powered by deep learning algorithms, excel in understanding and generating coherent human language responses, revolutionizing various fields. In education, these models promise to enrich the learning experience by providing personalized support to students and facilitating communication between students and teachers. In particular, Generative Pre-trained Transformer (GPT) technology provides instant and contextually relevant responses, and it has become an increasingly interesting topic in education. This article explores the potential of ChatGPT to enhance learning activities in secondary schools, with a specific focus on concept maps, a valuable tool for fostering meaningful learning. Concept maps are widely recognized for their educational benefits but existing tools for automatic concept map generation lack flexibility and semantic understanding. In contrast, ChatGPT's adaptability and responsiveness make it a compelling candidate for generating concept maps tailored to individual needs. This paper sheds light on ChatGPT's potential in education and offers insights into its effectiveness in supporting learning activities through concept map generation. We present a study involving 83 secondary school students to test their perceptions of ChatGPT-generated concept maps. We compared concept maps created by both ChatGPT and teachers across six different topics. Using PlantUML, we standardized the concept maps, and then used various questionnaires to assess their quality, effectiveness, and impact on student performance. The statistical analysis revealed that the concept maps generated by ChatGPT were comparable in quality to those produced by teachers. This highlights the usefulness of ChatGPT in supporting structured learning activities while also significantly reducing the time and effort required to create concept maps.

ARTICLE HISTORY

Received 13 January 2024

Accepted 7 April 2025

KEYWORDS

Large language models; concept maps; educational technologies

1. Introduction

The progress made in the field of Natural Language Processing (NLP) has opened up new avenues for creating high-performance Artificial Intelligence (AI) models. Large Language Models (LLMs) leverage deep learning algorithms to effectively interpret and grasp human language and subsequently generate logical and well-connected textual responses in accordance with user inquiries (Devlin et al., 2019). These models have significantly enhanced the capabilities of machines to comprehend

CONTACT Davide Taibi  davide.taibi@itd.cnr.it

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

human-generated text, exhibiting impressive conversational proficiency and their generative prowess in extracting information from large amounts of data to offer valuable insights to the users. Their potential has caught worldwide attention impacting several areas by supporting the creation of new solutions to real problems such as legal and financial analysis (Lakkaraju et al., 2023), customer support, content generation, industry 4.0 (Javaid et al., 2023) and many others (Liu et al., 2023). LLMs have also the potential to enhance various aspects of education significantly. With their assistance, teachers can fine-tune their lesson preparation, making learning activities more inclusive and engaging for students from diverse backgrounds, and they can facilitate communication with students and their families, fostering a more supportive learning environment. Overall, LLMs are a valuable tool that can aid educators in providing a comprehensive and effective learning experience for students (Trust et al., 2023).

Recently, ChatGPT¹ has been revealed as one of the most important breakthroughs by setting the record of the fastest-growing consumer application in history and thus drawing worldwide attention toward the LLMs models². It is an advanced chatbot that employs the latest Generative Pre-trained Transformer (GPT) technology (Radford et al., 2018) to produce highly relevant and accurate digital content instantly, generating text-based responses to user queries and providing a seamless and efficient communication experience. Its use in the education field is a research topic with the aim of disclosing its relevance for both students and teachers. The potential benefits of ChatGPT in promoting teaching and learning have been discussed in various research papers, including Tili et al. (2023), Lo (2023), Theophilou et al. (2023), Schicchi et al. (2023), and Chiu (2023). Moreover, it can support learners with special needs, as recently underlined by the UNESCO Guidance for generative AI in education and research (Fengchun & Wayne, 2023). This guidance reports that generative AI models have the potential to assist people with hearing or visual impairments through features like subtitles, audio descriptions, and text-to-speech conversions. Additionally, generative models aid learners who use minority languages in collaborating with peers from diverse linguistic backgrounds. However, it is essential to recognize that it also comes with potential risks of use such as plagiarism, and over-reliance. Therefore, it is crucial to carefully balance the benefits and drawbacks of a such new technology and use it appropriately to ensure the enhancement of the learning experience while mitigating potential risks³.

The aim of this paper is to investigate the effectiveness of using ChatGPT for supporting learning activities in high school. Specifically, we prompted ChatGPT to generate concept maps (CMs), a well-known learning tool that stimulates students to achieve meaningful learning by representing information, knowledge, or facts in a graphical form Novak and Cañas (2006).

Concept maps, Novak (1990), have been widely used as a tool for teaching and learning in educational settings. These graphical representations of knowledge structures can help students organize their understanding of complex concepts, identify relationships between different ideas, and communicate their knowledge effectively. Over the years, research has shown that using concept maps in education can lead to improved learning outcomes, increased engagement, and enhanced critical thinking skills (Harris & Zha, 2013). Moreover, technological advances have made creating and sharing CMs easier, allowing for more collaborative and interactive learning experiences. Despite the potential benefits of CMs, however, all educators have not fully embraced their use in education and there are still challenges to overcome, such as how to integrate CMs into curriculum design and assessment effectively. The adoption of concept maps has been limited due to their complex and time-consuming creation process. Users must report concepts graphically, resulting in a significant investment of personal effort. Additionally, the reuse of concept maps proves challenging, given that most are in image format, which cannot be edited.

Recently there has been a large production of tools to automatically generate CMs from texts (e.g.: AlgorEducation⁴), but they lack the possibility to adjust them on a personal necessity like, for example, to have a synthetic or detailed map or to have only the key concepts expressed. Conversely, the maps produced by ChatGPT may vary depending on the type of in-depth inquiry or specific request made to the chatbot.

We experimented to examine the potential of ChatGPT in supporting the creation of concept maps, which can play an important role in the learning process but can present difficulties for educators and students. Our study involved the comparison of concept maps generated by ChatGPT 3.5 (hereafter referred to as ChatGPT) with those manually generated by teachers. By utilizing version 3.5 of ChatGPT, the only freely accessible version at the time of the experiment, we wanted to promote inclusion ensuring that a wide range of students can benefit from its features. Inclusion is a fundamental educational issue that must be tackled as described in Sustainable Development Goal of United Nations⁵. It is essential to make sure that AI can be used by all students, irrespective of their gender, ethnicity, special educational needs, and socio-economic status.

In our analysis, we shed light on the capabilities of ChatGPT in extracting concepts, creating concept-to-concept links, and using an effective nomenclature to name concepts analyzing a source text. The study was carried out by analyzing the students' opinions regarding the maps' structure and evaluating the usefulness of the created maps, considering the students' learning outcomes via questionnaires. In particular, we answer the following research questions:

RQ1 : Can ChatGPT generate complete CMs starting from a text?

RQ2 : Do students achieve improved performance using ChatGPT-generated CMs?

RQ3 : Can students recognize CMs automatically generated?

The paper is organized as follows: In Section 2, we summarize the most important related works that are relevant to using concept maps and Large Language Models. Section 3 explains the methodology we have employed for collecting data to achieve our research objective. In Section 4, we present the final outcomes of the analysis conducted on the collected data, and Section 5 discusses these results. Finally, in Section 6, we provide conclusions highlighting future research directions.

2. Literature review

2.1. Concept map

Concept maps (CMs) have diverse applications in education (J. Novak, 2010), fostering critical thinking and collaborative knowledge development (Novak, 1990) and also (Adesope & Nesbit, 2009; Harris & Zha, 2013; McClure et al., 1999; Shih & Chang, 2020; Wang et al., 2017). They support the development of learning and instructional strategies, promote learner autonomy, and facilitate interaction and peer learning (Van Boxtel et al., 2002). This contributes to increasing the success rate and making learning activities easier, even for students with special needs (J. D. Novak, 2010). CMs are also useful for assessing comprehension, interpreting student responses, and predicting performance (Anohina-Naumeca et al., 2011; Maharjan & Rus, 2019; Omer et al., 2020). They empower students, aid in organizing class materials, and reduce reliance on rote memorization (Safdar et al., 2012). CMs can be created collaboratively using platforms like 'graphed' (Ionas & Geana, 2021), and they assist in ontology acquisition (Starr & De Oliveira, 2013).

However, the usefulness of CMs is often not fully exploited, as the work involved in constructing them can be time-consuming and can hold teachers or students back from using them. The effectiveness of CM is recognized mainly in the field of special needs such as dyslexia (Lami & Locatelli, 2008), but it is equally useful and effective for the learning of able-bodied people, especially for collaborative learning (Adesope & Nesbit, 2009).

Over the years, concept maps have been studied extensively in educational research, and several articles have emerged as key contributions to this field. In particular, "The Theory Underlying Concept Maps and How to Construct and Use Them" by Novak and Cañas (2006) provides a comprehensive overview of the theoretical underpinnings of concept maps, as well as empirical evidence of their effectiveness as a tool for promoting student learning and understanding of complex concepts.

A more recent work by Chun-Chun Chang and Tu (2022) provides an overview of how concept maps can support education, highlighting their impact on knowledge construction, learning design, and various subject areas.

Recognizing the extreme usefulness of concept maps (CMs) for learning, several software tools have been developed for generating concept maps, including those that automatically create maps from textual documents. CmapTools⁶, Miro⁷, Gitmind⁸, Algoreducation⁹, Mimind¹⁰, are some of the most popular software.

2.2. Large language models and ChatGPT

A Large Language Model, with ChatGPT being the most popular example, is a form of Artificial Intelligence that utilizes deep learning algorithms to process and analyze vast amounts of textual data. These models are designed to comprehend human language at a profound level, enabling them to generate logical and well-connected responses in line with user inquiries. One of the primary benefits of LLMs is their capability to understand and interpret the subtleties of human language. This allows LLMs to generate responses that are not only grammatically accurate but also contextually appropriate and relevant. LLM represents a significant breakthrough in AI and has the potential to revolutionize the education sector.

The usage of ChatGPT has gained popularity among people, and educational institutions are now promoting academic integrity by creating clear policies and guidelines (Adeshola & Adepoju, 2023). Recently, several studies have shown the advantages of using ChatGPT in education. For instance, Farrokhnia et al. (2023) provides a detailed SWOT analysis outlining the system's strengths, such as its capacity to generate text customized to suit varying learning styles and levels. In their study, Surameery and Shakor (2023) explored the effects of an intelligent tutoring system for programming instruction that used ChatGPT technology. The results showed improved student performance and engagement. Similarly, Sallam (2023) investigated the use of ChatGPT in medical education and found that it successfully enhanced students' knowledge and comprehension.

In general, LLMs can benefit students in several ways. They can help students acquire information for completing assignments and projects by answering specific questions about a topic (Kasneji et al., 2023). LLMs can also suggest books and articles for research purposes (Alafnan et al., 2023). LLMs can act as tutors and correct students' language mistakes in real-time, which can help improve their language skills (Cotton et al., 2023). They can also explain solutions to assignments and prevent students from wasting time on futile research. Additionally, LLMs can provide additional services like organizing class schedules, tracking assignment deadlines, creating to-do lists, and supporting students' motivation to enhance their efficiency and effectiveness (Kasneji et al., 2023). From a teacher's perspective, LLMs have been a valuable tool for creating educational content. LLMs can generate educational material such as summaries, quizzes, and assignments, which can significantly reduce the involvement of teachers in creating such content (Kianian et al., 2023). This enables educators to invest their time and energy in other crucial aspects of teaching, such as classroom management and personalized learning (Sinha et al., 2023).

Despite its exceptional language analysis and synthesis capabilities, the literature offers few contributions on using ChatGPT for creating concept maps. Recently, ChatGPT has been utilized by Tao and Xu (2023) to create mind maps. The authors highlight the benefits of ChatGPT in improving efficiency and creativity in map production. Tools that use ChatGPT for creating concept maps have been developed by the Edraw Content Team¹¹ and by Xmind¹², they also offer a rough analysis of its behavior in such a context.

2.3. Comparing ChatGPT and other AI techniques for building concept maps

Our research investigates the role of ChatGPT in generating dynamic concept maps with existing literature on the application of artificial intelligence in educational settings. This approach aligns with

and challenges the conventional educational methodologies by incorporating advanced language model capabilities for interactive learning.

Studies like those exploring interactive concept maps tailored for special education needs emphasize the necessity for adaptable educational tools that cater to diverse learner needs (Galletti et al., 2022). Our use of ChatGPT extends this adaptability by allowing for real-time interactions, thereby enhancing personalized learning experiences. Unlike the structured methods using affinity propagation for concept map generation from texts (Qasim et al., 2013), ChatGPT introduces flexibility and responsiveness, supporting the evolution of educational content to meet student requirements dynamically.

Parallel to the dynamic updating mechanisms in educational content driven by big data (Liu, 2024), ChatGPT leverages real-time data to continuously adapt and personalize learning materials. This reflects a broader trend towards agile and data-driven educational practices, where content is not static but evolves according to the educational environment and student interactions.

In contrast to methods that convert textbook content into concept maps (Olney et al., 2011), our approach promotes an inquiry-driven learning model. ChatGPT not only facilitates access to beyond-textbook knowledge but also encourages students to actively engage with learning materials. This fosters critical thinking and a deeper understanding of the subject matter, moving away from rote memorization to a more exploratory and engaging educational experience.

In summary, while supporting current trends in educational technology by integrating AI to facilitate learning, our research challenges traditional models by advocating for a participant-driven approach to knowledge acquisition. The use of ChatGPT in creating concept maps represents a significant advancement, positioning our work at the forefront of AI applications in education.

3. Materials and methods

To achieve the specific goal of our research, we followed a rigorous and systematic approach. This involved conducting experiments using carefully controlled variables and procedures to obtain reliable and valid results.

We enlisted the help of students to understand the utility of ChatGPT in creating useful concept maps evaluated in several aspects and compared with the ones created by teachers.

The framework employed in this study for testing the efficacy of generating concept maps using ChatGPT is guided by UNESCO's guidelines on the ethical use of generative AI in education and research (Holmes & Miao, 2023). The approach prioritizes human-centered AI usage, ensuring that the integration of AI tools like ChatGPT into educational processes enhances learning while adhering to ethical principles such as inclusivity, equity, and the preservation of human agency. Central to this method is a commitment to promoting accessibility and fairness across diverse learners, regardless of linguistic, cultural, or socio-economic backgrounds.

A key aspect of this approach is the protection of human agency. AI-generated content is not considered authoritative or final, rather, it is treated as a supplementary tool that stimulates critical thinking and discussion. Throughout the concept map generation process, students engage in an active review of ChatGPT's outputs. This ensures that human judgment remains central, as any inaccuracies or incomplete associations presented by the AI are identified and reported through human intervention.

In terms of implementation, the input preparation process involved curating data from curriculum guidelines and class materials. These inputs were designed to align with the educational goals of the study and to ensure the maps generated by ChatGPT reflected the intended learning outcomes.

While ChatGPT facilitated the extraction of concepts and relationships between them, students remained involved in refining and enhancing the output, ensuring that the maps were contextually relevant. The validation phase was a critical part of the framework. Concept maps created by ChatGPT were reviewed by students to ensure their accuracy and completeness, reinforcing the

importance of human oversight and ethical considerations in AI deployment within educational settings. This process highlighted the role of learners not just as passive recipients of AI-generated content but as active participants in refining and applying the outputs meaningfully.

Produced concept maps served as effective tools for fostering inquiry-based learning in order to investigate the potential for generative AI to assist in educational activities.

The following will provide more details about each experimentation stage, including the tools and materials used.

3.1. Participants

Our research involved 83 students from four classes at a technical secondary school in Italy. The sample consisted of 62 males and 20 females, while one person preferred not to specify the gender. The students attended three different k-12 classes and were aged between 15 and 20 years old.

Each student was assigned a unique code to ensure anonymous participation. The students evaluated several CMs generated by teachers and ChatGPT, as described below.

3.2. Materials

3.2.1. Concept maps

Through a collaborative effort with school teachers, we have carefully identified six distinct topics coherent with the students' curriculum, in both the technical and humanistic areas. The chosen topics concern computer science and history subjects, the complete set of topics is listed in [Table 1](#) in which the symbol T_i , with $1 \leq i \leq 6$, represents the corresponding textual content provided by teachers.

Teachers were also asked to create a concept map for each topic. For the sake of comprehension, we identify each map created by the school teacher as a Domain Expert (DE) with the notation C_i^{DE} , with $1 \leq i \leq 6$. Hereinafter the terms domain expert and teacher are used interchangeably in this paper.

The same set of texts T_i were utilized to request ChatGPT to generate concept maps associated with them. Since ChatGPT version 3.5 can only generate textual output and not graphical representations, we requested ChatGPT to generate the PlantUML map version. PlantUML¹³ is a highly versatile software component that facilitates the swift and efficient creation of diagrams in a streamlined and user-friendly manner. It offers a proprietary language allowing the user to draft a wide range of intricate diagrams easily and conveniently.

We leveraged the prompt *Could you create a concept map of the following text by using PlantUML format?* to query ChatGPT. Then, the prompt was enriched with the text T_i of the topic. A detailed example is shown in [Appendix A.2](#).

The concept maps created by ChatGPT are identified with the notation C_i^{GPT} , with $1 \leq i \leq 6$.

For consistency and aesthetic appeal, the maps made by teachers have been standardized and visually represented through PlantUML. The dataset used for our experiment consists of maps created by both ChatGPT and domain experts.

Table 1. The list of topics and subjects covered in our experiment.

Text id	Topic	Subject
T_1	Web Applications	Computer Science
T_2	Classes and Objects in PHP	Computer Science
T_3	Project Lifecycle	Computer Science
T_4	Napoleon's Reforms	History
T_5	First World War	History
T_6	The First Post-War Period in Italy	History

This approach granted us to obtain a set M of twelve concept maps, six generated by ChatGPT and six generated by domain experts.

A graphical representation of the entire concept map generation process is shown in [Figure 1](#).

3.2.2. Questionnaires

We defined four online questionnaires during the different stages of the experimental session. The questionnaires are detailed below:

- (1) **Demographics:** We collected participants' data related to age, gender, and school year.
- (2) **Concept Map Structural questionnaire:** with this questionnaire, we solicited participants' evaluation of concept maps. Participants were presented with multiple-choice questions to assess the concept maps, both created by ChatGPT and the domain expert. This questionnaire aims to evaluate the general appreciation of the concept map as well as the three dimensions of representation, usefulness, and defectiveness detailed below.

General Appreciation measures the students' perception of the map's quality expressed via five different degrees by using a 5-point Likert scale and answering the question *Please express a general evaluation of the map*. The objective of such a question is to have a global measure of the map's quality that involves several characteristics chosen intrinsically by the students.

Representation aims at measuring the author's capability of both including and arranging concepts according to the source text. It involves the extraction of concepts from the text, creating concept-to-concept links, and using an effective nomenclature to name concepts. To measure the representation dimension, students have been asked to answer the following question via a vote on a 5 points Likert scale: *Does the map represents correctly the concepts in the text?*, where 1 means *the map does not express the concept in the text at all*, and 5 means *the map express the concept in the text totally*.

Usefulness aims at stating the student's perception of the map's function to study the subject topic. Students were asked to answer the question *Do you believe that the current map is a helpful tool for studying the subject topic?* via a 5-point Likert scale. Students can state that *the map is useless* voting 1, or stating that *the map is useful* voting 5.

Defectiveness expresses the map's completeness regarding concepts. It aims at measuring the author's precision during the map creation and it is based on the presence/absence of concepts and the position where they are located in the map. To measure the map's defectiveness, students have to answer two questions: *Are there any missing concepts in the map?* and *Are there concepts in the wrong position on the map?* answering either yes or no. In case of yes we further ask to motivate the answer.

- (3) **Topic-related questionnaire:** for each map, a multiple-choice questionnaire was created. These questions are centered on the themes and content covered in the respective maps' associated texts. Importantly, participants were presented solely with the conceptual maps and were not provided with the accompanying text content. We emphasize that these questionnaires were created by teachers different from those who created the conceptual map, to avoid bias in the expected responses.

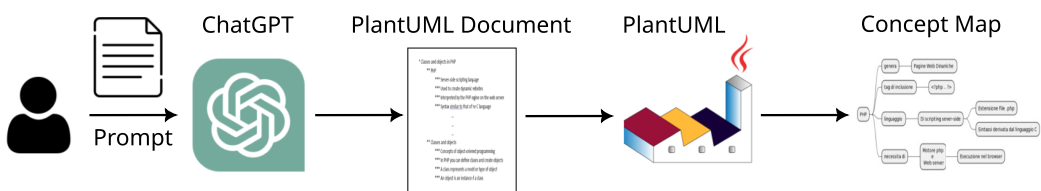


Figure 1. Pipeline of the concept map creation process through ChatGPT and PlantUML.

- (4) **“Reveal the AI” questionnaire:** We presented participants with the six pairs of maps, displaying both the concept map generated by the domain expert and the corresponding one generated by ChatGPT, and asked which one is generated by ChatGPT. In this survey, we challenge the participants to identify the concept map generated by the AI model.

For our experimental setup, we created surveys by using the *LimeSurvey* platform. *LimeSurvey*¹⁴ is an online survey software that allows users to quickly create and administer surveys, questionnaires, and polls. It provides a user-friendly interface for designing surveys, collecting participant responses, and analyzing the data.

3.3. Experimental setup

Our experiment consists of three stages.

- *Stage 1:* students were asked to evaluate the concept maps by the Concept Map Structural questionnaire. In this stage each student evaluates all the 12 maps; demographic data were also collected.
- *Stage 2:* each student uses 6 concept maps, three generated by ChatGPT and three generated by the teacher, to answer the Topic-related questionnaires.
- *Stage 3:* students were asked to detect which concept maps were generated by ChatGPT using the ‘Reveal the AI’ questionnaire.

At each stage of the experiment, students do not know which map was created by ChatGPT and which by the teacher. This helped to obtain valuable feedback from students on the quality of maps.

Appendix reports a complete example for the topic related to T_2 .

During *Stage 1*, students were given a set of 12 concept maps to be evaluated by using the Concept Map Structural questionnaire. In the next section, we report the results in detail.

Stage 2 aims to comprehensively and impartially evaluate the effectiveness of AI-generated and expert-crafted concept maps. The students were split into two groups and tasked with answering four multiple choice questions related to each map’s subject (see Appendix A.4). The aim of *Stage 2* was to evaluate the effectiveness of the maps in helping students understand core concepts and develop their competencies. By analyzing the student’s performance, we can gain valuable insights into the benefits of using this tool to enhance the learning process.

In particular, each group received two sets of 6 concept maps: one set comprised three maps generated by ChatGPT, denoted as C_i^{GPT} , and the other set included three maps crafted by domain experts, labeled as C_i^{DE} .

As shown in Table 2, to guarantee a well-rounded evaluation, we used an alternating pattern to assign maps to the different groups. We ensured that each group used a map created by a different author (ChatGPT or domain expert) for the same topic. For example, if *Group 1* was assigned the map C_1^{GPT} , *Group 2* would use the corresponding C_1^{DE} map.

In *Stage 3*, a couple of maps are shown to students, and they are asked to identify which one was created by ChatGPT. The experiment in this stage aims to understand the students’ capabilities to recognize concept maps drafted by a generative artificial intelligence model.

4. Results

4.1. Stage 1 – completeness of AI-generated concept maps

Table 3 gathers results from the first stage of the experience, showing, beyond the general appreciation, the students’ expressions of representation, usefulness, and effectiveness. Votes of the first three dimensions have been aggregated by computing both the mean and the standard deviation.

Table 2. Map distribution among students' groups.

Topic	Group 1	Group 2
T_1 Web Applications	C_1^{GPT}	C_1^{DE}
T_2 Classes and Objects in PHP	C_2^{DE}	C_2^{GPT}
T_3 Project Lifecycle	C_3^{GPT}	C_3^{DE}
T_4 Napoleon's reforms	C_4^{DE}	C_4^{GPT}
T_5 First World War	C_5^{GPT}	C_5^{DE}
T_6 The First Post-War Period in Italy	C_6^{DE}	C_6^{GPT}

To validate the results, we used the signed-rank test by Wilcoxon (Woolson, 2007), a non-parametric statistical test that is used to compare two groups of paired data points. The choice of such a test is required because of the non-normality distribution of the data, tested by the Shapiro-Wilk test (Shapiro & Wilk, 1965), and the need to compare the paired students' votes that express the preferences concerning the maps created respectively by ChatGPT and Teachers.

Let x_i be the distribution of the votes related to C_i^{GPT} (the maps generated by ChatGPT) and y_i be the distribution of the votes related to C_i^{DE} (the maps created by teachers). The Wilcoxon test assumes no specific distribution of the data and supposes as null hypothesis that the difference between the paired observations $d_i = x_i - y_i$ with $1 \leq i \leq 6$ in the population is zero. We chose as the alternative hypothesis to test if the distribution underlying d is stochastically greater than a distribution symmetric about zero.

We computed the p -value via the python SciPy library (Virtanen et al., 2020), a widely used library that provides algorithms for optimization, integration, interpolation, eigenvalue problems, algebraic equations, differential equations, and statistic.

Defectiveness is measured by assessing the frequency of errors in the maps. This involves counting the number of maps the students perceive as incomplete, with missing concepts, or with concepts positioned incorrectly. By analyzing these factors, we can gain insights into the accuracy and completeness of the conceptual maps. These values can then be divided by the total number of students' votes (i.e. 12 maps for each student with yes/no response) to evaluate the percentage of incomplete maps. Table 4 expresses for each map C_i created both by ChatGPT and domain experts, the number of students who assume that C_i misses concepts or concepts are placed in the wrong position. Table 4 also contains the overall percentage of votes related to the Missing Concept and the Wrong Concept Position that have been expressed by students for the maps created by ChatGPT and domain experts.

Under the dimension of general appreciation, metrics in Table 3 highlight ChatGPT's better performance, which has an average score of 3.99, versus the teachers', whose maps have been rated 3.7811. The representation score shows that ChatGPT performs better than domain experts in extracting concepts from the text, establishing concept-to-concept links, and using an effective nomenclature to name concepts. In the same way, students have stated that the maps created by ChatGPT are more useful for studying the subject topic, rating them via a score of 3.8594, which overcomes the one achieved by the domain experts' maps. Defectiveness is evaluated considering the lack of concepts and concepts put in the wrong position within maps. In this case, students consider the maps created by ChatGPT more complete than the others, stating only in 8.23% of cases that

Table 3. Measures of the maps students evaluation expressed by mean, standard deviation related to General Appreciation, Representation, and Usefulness.

Dimension	ChatGPT		Domain expert		p -value
	Mean	Std	Mean	Std	
General appreciation	3.9900	1.0089	3.7811	1.0032	0.0001**
Representation	4.0221	0.9279	3.7791	1.0274	0.0000**
Usefulness	3.8594	1.0298	3.7129	1.0277	0.0047**

** p -value ≤ 0.01 ; * p -value ≤ 0.05

Table 4. The first table highlights the number of students who believe that one or more concepts are missing from the maps, while the second table shows the number of students who believe that concepts are placed in the wrong position. The tables also display the overall percentage of votes related to Missing Concepts and Wrong Concept Position.

<i>Maps with Missing Concepts</i>	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	Tot	%
ChatGPT	15	10	4	2	5	5	41	8.23%
Domain expert	39	38	27	8	12	13	137	27.5%
<i>Maps with Wrong Concept Position</i>	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	Tot	%
ChatGPT	11	10	1	0	1	5	28	5.6%
Domain expert	15	14	3	3	7	2	44	8.83%

such maps miss concepts. Moreover, concepts in maps created by ChatGPT are arranged better than the ones in the maps created by domain experts.

4.2. Stage 2 – students’ performance when AI generated map are used

In the second stage, ChatGPT’s concept maps are compared with those created by domain experts to evaluate their effectiveness in helping students answering topic-related questions. The experimental setup allows us to analyze such an aspect by comparing the maps created by ChatGPT and the ones created by the teachers one-to-one, following the order of maps shown in Table 2. Table 5 includes the per-topic results achieved by students via the usage of either ChatGPT maps or domain experts’ maps. It is expressed as the mean of the mean values of the questions correctly answered by each student. The table also highlights the *p*-value, representative of the statistical validation carried out via a T-test for the means of two independent samples relying on the python library SciPy.

According to the results, students performed well in five out of six topics using maps created by both authors. Specifically, maps created by domain experts enable students to score better in three topics out of five, while maps created by ChatGPT outperform the ones created by domain experts in the remaining two topics. The most successful cases involve using ChatGPT to answer questions about *Web Applications* and using the domain experts’ map to answer questions about *Napoleon’s reforms*. However, students have demonstrated limited performance in answering questions about the *First World War* using maps created by both authors. In the latter case, the results do not show any statistical significance.

4.3. Stage 3 – students’ capability in recognizing AI generated concept maps

In Stage 3, we investigate the students’ capabilities to identify whether a concept map is generated by an Artificial Intelligence model or created by humans. The collected data is arranged in Table 6, showing the complete set of performances achieved by students. Table 6 shows that 51.80% of students have selected the right map at least four times out of six, 34.93% of students spotted the maps created by ChatGPT at least five times, and only 7.22% of students were capable of spotting the entire set of maps created by ChatGPT.

Table 5. Mean of the means students’ score for each topic grouped by author. $|\Delta|$ indicates the absolute value of the difference between the scores achieved by ChatGPT’s maps and those achieved by teachers’ maps.

Topic	ChatGPT	Domain expert	$ \Delta $	<i>p</i> -value
Web Applications	0.8929	0.7188	0.1741	0**
Napoleon’s reforms	0.6667	0.8143	0.1476	0.0134*
First World War	0.3929	0.3698	0.024	0.6007
Classes and Objects in PHP	0.375	0.5143	0.1393	0.0104*
Project Lifecycle	0.4929	0.401	0.0919	0.0394*
The first post-war period in Italy	0.3698	0.5714	0.2016	0.0007**

***P*-value $\leq .01$; **P*-value $\leq .05$

Table 6. Students' performance in identifying maps created by ChatGPT.

Rate	#Stud	%Stud	%Stud (Cumulative)
6/6	6	7.22	7.22
5/6	23	27.71	34.93
4/6	14	16.87	51.80
3/6	16	19.28	71.08
2/6	15	18.07	89.15
1/6	6	7.23	96.38
0/6	3	3.61	100.00

After the questionnaire, the students were invited to make written observations on the map traits utilized to discern the maps produced by ChatGPT. We have gathered 27 student opinions and analyzed them manually, including the most significant ones in [Table 7](#). Recurrent arguments are the length of the text included in the nodes, the organization and completeness of the maps, and the map's structure.

5. Discussion

During the initial stage of the experiment, we analyze the structure of concept maps based on students' opinions regarding four dimensions: general appreciation, representation, effectiveness, and usefulness. Such investigation helps us to answer RQ1, which aims at clarifying the capabilities of ChatGPT to generate concept maps from a free text. ChatGPT has been challenged to extract concepts from the source text and relations among them. Students who have compared the maps created by teachers and ChatGPT without knowing the maps' author have shown a clear preference for the outcomes of ChatGPT. As indicated in [Table 3](#), scores of ChatGPT's maps surpass on average the ones achieved by the maps created by teachers. Our analysis of the first three dimensions reveals that ChatGPT can generate high-quality maps that students generally appreciate. The analysis of last dimension, defectiveness, confirms the attitude of language models to perform well in the concept extraction task (Günel & Aşlıyan, 2010), which is a fundamental ability for creating concept maps. Moreover, students' opinions confirm that ChatGPT can extract semantic relationships from a free text, linking concepts by meeting their structure's hierarchy described in the source text (Wadhwa et al., 2023). This is a new trend in current research, where LLMs are used to solve the relation extraction task, a common problem in the natural language processing field that aims to identify and classify relationships between entities mentioned in a text (Carta et al., 2023; Zhang et al., 2023).

Concept map creation requires rigorous tasks, including extracting relevant information to identify pertinent concepts within a specific field, establishing connections between those concepts, and organizing them into a hierarchical structure (Novak & Cañas, 2006). The comprehensive analysis

Table 7. Insights of students' opinions from *Stage3*.

	Students' Opinions
1	<i>maps shorter and more concise</i> were made by the teacher while the <i>maps rich in information and more specific</i> were made by artificial intelligence
2	differences between the concept map drafted by teachers and the ones created by the artificial intelligence were the <i>organization and completeness</i> of the text and the <i>amount of information</i> provided
3	I recognized the concept map created by artificial intelligence noticing the <i>length of the text</i>
4	sentences written by an AI are more <i>articulate</i>
5	the maps made by AI are very <i>complex and articulate</i> , they are done in a <i>very detailed</i> way compared to those made by the teacher which are <i>more concise</i> by summarizing the topic in a simpler way
6	AI-created maps are much <i>more detailed and reasoned</i>
7	Artificial Intelligence maps are <i>very verbose</i> by providing more information than needed
8	the teacher writes in a <i>more grouped and non-schematic</i> way
9	maps created by artificial intelligence seem to be <i>more reasoned</i> than what a teacher might make
10	I thought the most <i>concise and clear</i> maps might have been created by a professor
11	Artificial intelligence gives clearer insights by creating <i>longer maps</i>

indicates that ChatGPT can efficiently address these tasks, making it a useful tool for both teachers and students to create concept maps. Its usage results in significant time savings, which could prove to be invaluable in an academic or professional context.

As part of the experiment, the second stage aims to collect students' feedback regarding the usefulness of concept maps created by ChatGPT in improving students' educational performance and competencies.

Students were asked to answer questionnaires on technical and historical subjects using the provided concept maps. The analysis of the collected data details the ChatGPT skills in promoting meaningful learning by creating clear maps that make it easy to memorize new concepts, understand relations between concepts to create new knowledge, and recover the information to answer a set of specific questions correctly. We conducted a performance analysis of students who utilized maps created by ChatGPT and those created by teachers. This enabled us to answer RQ2.

The results presented in [Table 5](#) indicate that both teacher-created maps and ChatGPT maps are effective tools in helping students answer questions related to specific topics. However, there is a difference in the performance of students who use maps created by teachers compared to those who use ChatGPT maps. The data shows that students who use maps created by teachers outperform their peers who use ChatGPT maps in three topics out of five. This finding highlights the importance of teacher expertise and the value of human input in creating educational resources (Alshehri, 2023; Marcelo & Baptista, 2023). Nevertheless, ChatGPT appears to be a valuable tool that can assist in teaching and learning. The data indicates that students who utilize ChatGPT maps outperform their peers in the remaining two topics, highlighting the potential of ChatGPT in producing high-quality maps. From the teacher's point of view, ChatGPT can significantly reduce the effort teachers spend creating concept maps for students (Çakmak, 2010) since the produced maps require only a supervising process and, eventually, minor adjustments. On the other hand, ChatGPT can assist students in their learning process by providing maps that can help them acquire new knowledge and competencies. This is evident from the analysis of the average difference between the scores achieved by the maps created by ChatGPT and those created by teachers. Upon examining solely the statistically significant outcomes, we have determined that the variation in performance, as shown in [Table 5](#) is 0.1509, which is not considered a substantial variance.

From recent findings in the LLMs field (Lee et al., 2023; Wu et al., 2023), it is possible to hypothesize that a knowledge refinement process through the retraining of ChatGPT, or one/few-shot learning, can enhance its capability of creating concept maps. This advancement can be utilized in contexts where students have special needs, promoting inclusivity and accessibility in education.

In the third phase of the research, we investigated students' ability to discern maps generated by ChatGPT from the ones created by teachers. The results revealed that around fifty percent of the students demonstrated the aptitude to identify the characteristics of ChatGPT-based maps for the most part, without any prior exposure to this concept mapping technique. Based on the analysis of opinions of students who answered at least four questions correctly, it has been found that the maps created by ChatGPT contain a significant amount of text (see opinions 1, 2, 8, 11 in [Table 7](#)). Some students see this trait as an indicator of completeness, while others feel that ChatGPT cannot summarize the core concepts and instead uses parts of the original text as map nodes. On the other hand, teachers meet the original definition of concept maps, which focuses on highlighting concepts and avoiding unnecessary text. A few students have mentioned that the text structure of ChatGPT's maps is more verbose and articulate than others (see opinions 4, 7 in [Table 7](#)).

In conclusion, our study enables us to advance the discourse on the utilization of ChatGPT in generating concepts, whose literature shows it has served as a supportive tool rather than a fully automated solution for concept map creation (Omar, 2023). ChatGPT has the potential to offer a range of benefits for both students and teachers. It can help students to create concept maps that can enhance their educational experience, achieving meaningful learning. ChatGPT can reduce teachers' workload by automating the manual process of creating concept maps. This allows them to provide

an enhanced teaching activity, for instance, by personalizing their teaching materials for each student, leading to more effective learning outcomes.

5.1. Study generalizability

The findings of this study can be generalized to broader educational settings due to the adaptability and flexibility of ChatGPT or the other large language models. These models are capable of tailoring their content to accommodate diverse learning requirements, making them suitable for a wide range of learners. For example, ChatGPT can simplify complex texts, making the information more accessible to students at varying comprehension levels (Schicchi & Taibi, 2024). This ability to dynamically adjust content can also extend to the generation of simplified concept maps, ensuring that even students with limited prior knowledge can benefit from the learning activity. Additionally, LLMs like ChatGPT have been successfully employed to assist individuals with specific learning challenges, such as dyslexia (D'Urso & Sciarrone, 2024), by presenting information in more digestible and structured formats. These features highlight the broad applicability of LLMs in education, making them valuable tools for inclusive and adaptable learning environments across different educational contexts.

Moreover, this study was conducted with secondary school students, but the approach can be modified to support concept map generation in elementary, middle school, and higher education contexts, thanks to the versatility of ChatGPT, capable of adapting the process of creating concept maps on the basis of the specifics of the final users by modifying the prompt accordingly. Educators can customize the prompts and structure of the maps to emphasize key concepts relevant to diverse subjects. This involves tailoring the input text and map structure to match each group's cognitive and academic levels.

Although the study focused on specific computer science and history topics, the same methodology could be used for various subjects, including STEM (science, technology, engineering, mathematics), humanities, and arts. For example, in science courses, ChatGPT can be exploited to generate concept maps to visualize complex systems such as ecosystems or chemical reactions. Similarly, concept maps can help illustrate relationships between characters, themes, and historical events in literature or social studies.

5.2. Study reproducibility

The study used a standardized approach that can be applied in different settings, making it fully reproducible through the provided documentation of source texts, submitted questionnaires, and a detailed explanation in the Methodology section. Additionally, all the instruments are free to use. Researchers who wish to expand on this study can utilize the same prompt (see Section 3.2) and PlantUML specifications to ensure consistency in creating maps. This standardized approach would be beneficial for testing different versions of ChatGPT (e.g. GPT-4) to observe how changes in the model impact the results.

In this study, concept maps were created using structured and consistent source texts for various topics, which were evaluated by humans. To replicate these results, future researchers should ensure that the input texts are comparable in terms of length, complexity, and content structure. By standardizing the format of the text provided to ChatGPT, the influence of input variability on the quality of generated concept maps can be minimized, resulting in a reproducible outcome.

The study used structured questionnaires to assess students' comprehension and opinions of the concept maps created. To ensure reproducibility, future research should employ the same assessment tools or create similar ones that match the learning goals of the subject under investigation to ensure standardization.

The study's methodology can be replicated effectively and extended to different educational contexts by addressing these factors. This document contains all the necessary details to allow

open access to the study's methodology, including the datasets, prompt scripts, and evaluation criteria, in order to ensure reproducibility. In addition, supplementary materials have been published, such as the exact text inputs used, sample concept maps, and evaluation rubrics, to enable other researchers to precisely replicate the study's setup¹⁵.

Ensuring reproducibility will not only validate the robustness of the research findings but also pave the way for wider adoption of AI-generated concept maps as effective educational tools.

5.3. Study limitations

This study has limitations that must be acknowledged to contextualize the findings and identify areas for future research.

First, the experiment was conducted exclusively with secondary school students from a single institution, all residing in the same region of Italy. This limited sample reduces the generalizability of the results, as the study does not include a broader range of students in terms of educational levels (e.g. elementary, middle, and higher education) or geographical diversity. Further research should include participants from multiple regions and educational stages to evaluate whether the observed outcomes can be replicated in different contexts and educational levels. Moreover, the study was conducted in a single experimental session, limiting its ability to capture long-term learning outcomes. A longitudinal study design could be implemented to evaluate the lasting effects of using ChatGPT-generated concept maps on students' knowledge retention and academic performance over time.

Second, the study focused on only two academic subjects, Computer Science and History, using a limited number of concept maps (six in total). This narrow scope may not reflect ChatGPT's potential across a broader array of subjects, such as Mathematics, Science, or Literature. Future studies should extend the experimentation to other subjects and incorporate a larger number of concept maps to provide a more comprehensive evaluation.

Another limitation concerns the version of the AI model used. The research utilized ChatGPT version 3.5, which might have certain shortcomings in its understanding and generation of complex educational content compared to newer models like ChatGPT 4.0 or other emerging Large Language Models, such as Claude, Google Gemini, and LLama 3.2. Newer models have been developed since the time of this study; therefore, their impact on the quality of concept maps should be considered in future research.

The quality analysis of the concept maps was primarily based on general appreciation, representation, usefulness, and defectiveness. However, other dimensions, such as readability, clarity of visual representation, and cognitive load, should be thoroughly investigated. Future research could incorporate these dimensions to provide a more nuanced assessment of concept map quality.

These limitations highlight the need for further research to validate and expand our findings, ensuring that the potential of AI-generated educational tools is fully understood and appropriately utilized.

6. Conclusions

In this study, we examined the role of ChatGPT in assisting teachers and students in creating concept maps. Previous literature has demonstrated the potential of ChatGPT in education (Okulu & Muslu, 2024; Songkram et al., 2024; Urhan et al., 2024). Our findings contribute to a better understanding of how ChatGPT can be used in the development of educational tools. The experimentation reveals that, in certain cases, concept maps generated by the teacher, as domain expert, demonstrate better outcomes than those generated by ChatGPT, while in other instances, the maps produced by ChatGPT outperform those created by teachers. While the difference may not be substantial, we can still be satisfied with the performance of ChatGPT. In fact, automatic map generation (at equal quality) provides considerable advantages for teachers, primarily regarding the speed at which maps can be generated. This offers a ready-made starting point upon which teachers can build, making modifications as needed.

Furthermore, students can derive several benefits from the use of concept maps generated by ChatGPT. Firstly, the visual representation of information aids in better comprehension and retention of complex concepts. The maps offer a structured overview that can serve as a study aid, facilitating review and reinforcement of key ideas. Additionally, the accessibility and availability of pre-generated maps can save students time, allowing them to focus on deeper engagement with the material.

In summary, the advantages of automatic map generation extend beyond the speed of map creation for teachers. They encompass enhanced adaptability, diverse perspectives, and the encouragement of creativity. Students benefit from improved comprehension, time efficiency, and opportunities for collaborative learning.

Concept maps are powerful tools that enhance understanding and inspire innovation in teaching methodologies. They can have a profound impact in fostering meaningful learning experiences, reinforcing our commitment to further understanding and harnessing the potential of these valuable teaching aids. This study is only a first step in confirming the quality (comparable with teachers) of the concept maps generated by ChatGPT.

There are numerous possibilities for improvement and extensive use of chatGPT in creating concept maps. It can be used to facilitate the creation of collaborative concept maps. The recent findings in Liu et al. (2023) highlight the importance of structured instructional strategies to enhance collaborative learning, suggesting methods such as the 'thoughts-statement' approach to improve students' engagement and understanding in group discussions.

Concept maps can be customized to suit various learning styles. Drawing from Hyerle's work in Hyerle (2014), where specific concept maps called Thinking Maps are considered, one can create maps that include more descriptive texts for verbal learners or interactive maps (audio/video) for kinesthetic learners. Additionally, it is possible to create maps of varying complexity using appropriate language based on the cognitive and academic levels of students. Taking inspiration from research on adapting thinking maps to Bloom's taxonomy (such as in Bunt et al., 2022), the use of tools like ChatGPT makes this approach straightforward to implement. The potential of this tool makes it possible to create compensatory tools for students with different learning deficits. Implementing these learning techniques can offer significant potential for development and innovation in the educational field, facilitating a more personalized and engaging learning experience for all students.

Notes

1. <https://chat.openai.com/>
2. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/> (accessed 2024/10/23)
3. <https://platform.openai.com/docs/chatgpt-education> (accessed 2024/10/23)
4. <https://app.algoreducation.com/maps>
5. <https://sdgs.un.org/topics/education>
6. <https://cmap.ihmc.us>
7. <https://miro.com>
8. <https://gitmind.com>
9. <https://www.algoreducation.com>
10. <https://mimind.cryptobees.com>
11. <https://www.edrawmind.com/article/mind-map-with-chatgpt.html> (accessed 2024/10/23)
12. <https://xmind.app/blog/chatgpt-and-xmind-how-to-create-a-mind-map-with-chatgpt/> (accessed 2024/10/23)
13. <https://plantuml.com/>
14. <https://www.limesurvey.org/>
15. <https://zenodo.org/doi/10.5281/zenodo.12207324> (accessed 2024/10/23)

Acknowledgments

The publication was realized with the co-financing of the European Union – ERDF or ESF, PON Research and Innovation 2014–2020–DM 1062/2021.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Daniele Schicchi is a research fellow at the Institute for Educational Technologies of the National Research Council of Italy (CNR-ITD). He obtained a Ph.D. in Information Communication Technology at the Doctoral School in Computer Science offered by the Consortium of Catania, Messina, and Palermo universities (Italy). He is involved in the study of artificial intelligence. His research interests cover the application of computer science methodologies and innovative technologies mainly to education and natural-language-related fields. A special interest concerns Learning Analytics, Educational Data Mining, and Text Simplification, primarily under the Machine Learning perspective.

Carla Limongelli got her PhD in Computer Science, in 1993, from the Sapienza University of Rome, Italy. She is currently Associate Professor at Roma Tre University in Rome. Her research activity mainly focuses on Artificial Intelligence in Education, mainly on Intelligent Adaptive Learning Environments, User Modeling and User-Adapted Interaction, Retrieval of didactic materials, Similarity criteria for concept maps comparison, Educational RS; Indoor RS, and Deep Learning Techniques for emotion detection. She has organized some international workshops, congress tracks, conferences, and special issues in the area of Artificial Intelligence in Education.

Vito Monteleone obtained his PhD in 2019 at the University of Palermo in Technological Innovation Engineering. His primary research interest is centered on the application of Artificial Intelligence (AI), with a particular focus on Change Point Detection, Event Detection, and Temporal Analysis. In addition to his core areas of expertise, he has recently expanded his academic pursuits to include the field of Education and Social Media Analysis. Within these domains, he investigates the intricate intersection of cutting-edge technologies, such as Deep Learning, Semantic Web, NLP and Generative Artificial Intelligence, with a focus on the design of educational processes. Currently, his primary emphasis is on discerning how these technologies can be applied effectively within the school environment.

Davide Taibi is a senior researcher of the Institute for Education Technology, National Research Council of Italy. He is also a part-time lecturer with Computer Science Department, University of Palermo and contract professor with the Libera Università Maria Ss. Assunta University in Palermo. Since more than 20 years, he has been working in the educational technology research area, and his main research interests concern pedagogical applications to smart environments, learning analytics, artificial intelligence and enriched reality in education. He coordinated two European funded projects on the development of data literacy competencies for university and business sectors. He has also collaborated as an associated partner in the projects mEducator best practice network and LinkedUp (Linking Web Data for Education).

ORCID

Daniele Schicchi  <http://orcid.org/0000-0003-0154-2736>

Carla Limongelli  <http://orcid.org/0000-0002-0323-7738>

Vito Monteleone  <http://orcid.org/0000-0001-7600-1251>

Davide Taibi  <http://orcid.org/0000-0002-0785-6771>

References

- Adehola, I., & Adepoju, A. P. (2023). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, 32(10), 1–14.
- Adesope, O., & Nesbit, J. (2009). A systematic review of research on collaborative learning with concept maps. In *Handbook of research on collaborative learning using concept mapping* (pp. 238–255). IGI Global.
- Alafnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*, 3(2), 60–68.
- Alshehri, B. (2023, October). Pedagogical paradigms in the AI era: Insights from Saudi educators on the long-term implications of AI integration in classroom teaching. *International Journal of Educational Sciences and Arts*, 2(8), 159–180. <https://doi.org/10.59992/IJESA.ISSN.2976-7237>
- Anohina-Naumeca, A., Grundspenkis, J., & Strautmane, M. (2011). The concept map-based assessment system: Functional capabilities, evolution, and experimental results. *International Journal of Continuing Engineering Education and Life Long Learning*, 21(4), 308–327. <https://doi.org/10.1504/IJCELL.2011.042790>
- Bunt, B. J., Grosser, M., & van Tonder, D. (2022). *A novel proposal to use thinking maps to embed blooms' taxonomy within teaching, learning, and assessment*. Springer.

- Çakmak, M. (2010). An examination of concept maps created by prospective teachers on teacher roles. *Procedia – Social and Behavioral Sciences*, 2(2), 2464–2468. <https://doi.org/10.1016/j.sbspro.2010.03.354>
- Carta, S., Giuliani, A., Piano, L., Podda, A. S., Pompianu, L., & Tiddia, S. G. (2023). Iterative zero-shot LLM prompting for knowledge graph construction. *arXiv preprint arXiv:2307.01128*.
- Chiu, T. K. F. (2023). The impact of generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 32(7), 1–17. <https://doi.org/10.1080/10494820.2023.2172044>
- Chun-Chun Chang, G. J. H., & Tu, Y. F. (2022). Roles, applications, and trends of concept map-supported learning: A systematic review and bibliometric analysis of publications from 1992 to 2020 in selected educational technology journals. *Interactive Learning Environments*, 31(9), 1–22.
- Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 1–12.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Human language technologies*, Vol. 1 (Long and Short Papers) (pp. 4171–4186). Association for Computational Linguistics.
- D’Urso, S., Sciarrone, F. (2024). AI4LA: An intelligent chatbot for supporting students with dyslexia, based on generative AI. In A. Sifaleras, & F. Lin (Eds.), *Generative intelligence and intelligent tutoring systems*. ITS 2024. Lecture Notes in Computer Science, Vol. 14798. Springer. https://doi.org/10.1007/978-3-031-63028-6_31
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 31(9), 1–15.
- Fengchun, M., & Wayne, H. (2023). *Guidance for generative AI in education and research*. UNESCO.
- Galletti, M., Anslow, M., Bianchi, F., Calanca, M., Tomaiuoli, D., Vedovelli, D., & Pasqua, E. (2022). Interactive concept-map based summaries for SEND children. In *Proceedings of the thirty-first International Joint Conference on Artificial Intelligence. AI for Good-Projects* (pp. 5236–5243). IJCAI. <https://doi.org/10.24963/ijcai.2022/727>
- Günel, K., & Aşlıyan, R. (2010). Extracting learning concepts from educational texts in intelligent tutoring systems automatically. *Expert Systems with Applications*, 37(7), 5017–5022. <https://doi.org/10.1016/j.eswa.2009.12.011>
- Harris, C., & Zha, S. (2013). Concept mapping: A critical thinking technique. *Education*, 134(2), 207–211.
- Holmes, W., & Miao, F. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing.
- Hyerle, D. (2014). Thinking Maps® : A Visual Language for Learning. In A. Okada, S. J. Buckingham Shumand, & T. Sherborne (Eds.), *Knowledge cartography: Software tools and mapping techniques* (pp. 73–87). Springer.
- Ionas, I., & Geana, M.. (2021). *Graphed: A web-based concept mapping application for instruction and research*. 7th International Conference on Higher Education Advances (HEAD’21). Editorial Universitat Politècnica de València.
- Javaid, M., Haleem, A., & Singh, R. P. (2023). A study on ChatGPT for Industry 4.0: Background, potentials, challenges, and eventualities. *Journal of Economy and Technology*, 1, 127–143. <https://doi.org/10.1016/j.ject.2023.08.001>
- Kasneçi, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneçi, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kianian, R., Sun, D., Crowell, E. L., & Tsui, E. (2023). The use of large language models to generate education materials about uveitis. *Ophthalmology Retina*, 8(2), 195–201.
- Lakkaraju, K., Jones, S. E., Vuruma, S. K. R., Pallagani, V., Muppasani, B. C., & Srivastava, B. (2023). LLMs for financial advice: A fairness and efficacy study in personal decision making. In *Proceedings of the Fourth ACM International Conference on AI in Finance* (pp. 100–107). Association for Computing Machinery.
- Lami, G., & Locatelli, M. (2008). Dyslexia and concept maps: An indispensable tool for learning. In *3rd International Conference on Concept Mapping “Concept Mapping–Connecting Educators”* (pp. 152–154). University of Estonia.
- Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., Kim, H. (2023). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in english education. *Education and Information Technologies*, 29(9), 1–33.
- Liu, W. (2024). Dynamic updating mechanism of teaching content driven by both artificial intelligence and big data within an efficient classroom. *Applied Mathematics and Nonlinear Sciences*, 9(1), 5995–6016.
- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology*, 1(2), 100017. <https://doi.org/10.1016/j.metrad.2023.100017>
- Liu, S., Kang, L., Liu, Z., Fang, J., Yang, Z., Sun, J., Wang, M., & Hu, M. (2023). Computer-supported collaborative concept mapping: The impact of students’ perceptions of collaboration on their knowledge understanding and behavioral patterns. *Interactive Learning Environments*, 31(6), 3340–3359. <https://doi.org/10.1080/10494820.2021.1927115>
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>

- Maharjan, N., & Rus, V. (2019). A concept map based assessment of free student answers in tutorial dialogues. In *Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25–29, 2019, Proceedings, Part I 20* (pp. 244–257). Springer International Publishing.
- Marcelo, Y. G. D. G., & Baptista, V. E. G. (2023). Concept map-based alternative instructional resource (AIR) in science. *Divine Word International Journal of Management and Humanities*, 2(4), 538–550.
- McClure, J., Sonak, B., & Suen, H. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 4(36), 475–492. [https://doi.org/10.1002/\(ISSN\)1098-2736](https://doi.org/10.1002/(ISSN)1098-2736)
- Novak, J. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, 27(10), 937–949. <https://doi.org/10.1002/tea.v27.10>
- Novak, J. (2010). *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations*. Taylor & Francis.
- Novak, J. D. (2010). *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations*. Routledge.
- Novak, J. D., & Cañas, A. J. (2006). The theory underlying concept maps and how to construct them. *Florida Institute for Human and Machine Cognition*, 1(1), 1–31.
- Okulu, H. Z., & Muslu, N. (2024). Designing a course for pre-service science teachers using ChatGPT: What ChatGPT brings to the table. *Interactive Learning Environments*, 32(10), 1–18.
- Olney, A., Cade, W. L., & Williams, C. (2011). Generating concept map exercises from textbooks. In *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 111–119). Association for Computational Linguistics.
- Omar, H. (2023). Use of ChatGPT as an educational tool to support generating structured concept maps for tobacco cessation (Pilot study). *Journal of Dental Education*, 88(Suppl 3), 1982–1987.
- Omer, U., Farooq, M. S., & Abid, A. (2020). Cognitive learning analytics using assessment data and concept map: A framework-based approach for sustainability of programming courses. *Sustainability*, 12(17), 6990. <https://doi.org/10.3390/su12176990>
- Qasim, I., Jeong, J. W., Heu, J. U., & Lee, D. H. (2013). Concept map construction from text documents using affinity propagation. *Journal of Information Science*, 39(6), 719–736. <https://doi.org/10.1177/0165551513494645>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. *OpenAI*.
- Safdar, M., Hussain, A., Shah, I., & Rifat, Q. (2012). Concept maps: An instructional tool to facilitate meaningful learning. *European Journal of Educational Research*, 1(1), 55–64. <https://doi.org/10.12973/eu-jer>
- Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. In *Healthcare* (Vol. 11, p. 887). MDPI.
- Schicchi, D., & Taibi, D. (2024). AI-driven inclusion: Exploring automatic text simplification and complexity evaluation for enhanced educational accessibility. In: G. Casalino, R. Di Fuccio, G. Fulantelli, P. Raviolo, P. C. Rivoltella, D. Taibi, & G. Antonia Toto (Eds.), *Higher education learning methodologies and technologies online (HELMeTO 2023)*. *Communications in Computer and Information Science*, Vol. 2076. Springer. https://doi.org/10.1007/978-3-031-67351-1_24
- Schicchi, D., Upadhyaya, A., Fisichella, M., & Taibi, D. (2023). Using ChatGPT to enhance students' behavior in social media via the moral foundation theory. In *Proceedings of the First International Workshop on High-Performance Artificial Intelligence Systems in Education Co-Located with 22nd International Conference of the Italian Association for Artificial Intelligence (AlxIA 2023)*. *CEUR Workshop Proceedings*.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611. <https://doi.org/10.2307/2333709>
- Shih, Y. A., & Chang, B. (2020). Empirical study on the effects of social network-supported group concept mapping. *Research and Practice in Technology Enhanced Learning*, 15(1), 1–22. <https://doi.org/10.1186/s41039-020-00143-7>
- Sinha, R. K., Roy, A. D., Kumar, N., Mondal, H., & Sinha, R. (2023). Applicability of ChatGPT in assisting to solve higher order problems in pathology. *Cureus*, 15(2).
- Songkram, N., Chootongchai, S., Keereerat, C., & Songkram, N. (2024). Potential of ChatGPT in academic research: Exploring innovative thinking skills. *Interactive Learning Environments*, 33, 1–23.
- Starr, R. R., & De Oliveira, J. M. P. (2013). Concept maps as the first step in an ontology construction method. *Information Systems*, 38(5), 771–783. <https://doi.org/10.1016/j.is.2012.05.010>
- Surameery, N. M. S., & Shakor, M. Y. (2023). Use chat gpt to solve programming bugs. *International Journal of Information Technology & Computer Engineering (IJITC)*, 3(31), 17–22. <https://doi.org/10.55529/ijitc>
- Tao, R., & Xu, J. (2023). Mapping with ChatGPT. *ISPRS International Journal of Geo-Information*, 12(7), 284. <https://doi.org/10.3390/ijgi12070284>
- Theophilou, E., Koyutürk, C., Yavari, M., Bursic, S., Donabauer, G., Telari, A., Testa, A., Boiano, R., Hernandez-Leo, D., Ruskov, M., Taibi, D., Gabbadini, A. & Ognibene, D. (2023). Learning to prompt in the classroom to understand AI limits: A pilot study. In R. Basili, D. Lembo, C. Limongelli, & A. Orlandini (Eds.), *AlxIA 2023 - advances in artificial intelligence*. *Lecture Notes in Computer Science* (Vol. 14318, pp. 481–496). Springer.
- Tilli, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15. <https://doi.org/10.1186/s40561-023-00237-x>

- Trust, T., Whalen, J., & Mouza, C. (2023, March). Editorial: ChatGPT: Challenges, opportunities, and implications for teacher education. *Contemporary Issues in Technology and Teacher Education*, 23(1), 1–23.
- Urhan, S., Gençaslan, O., & Dost, Ş. (2024). An argumentation experience regarding concepts of calculus with ChatGPT. *Interactive Learning Environments*, 32(10), 1–26.
- Van Boxtel, C., Linden, J. van der, Roelofs, E., & Erkens, G. (2002). Collaborative concept mapping: Provoking and supporting meaningful discourse. *Theory into Practice*, 41(1), 40–46. https://doi.org/10.1207/s15430421tip4101_7
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., S. J. van der Walt, Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Wadhwa, S., Amir, S., & Wallace, B. C. (2023). Revisiting relation extraction in the era of large language models. In *Proceedings of the Conference. Association for Computational Linguistics. Meeting* (Vol. 2023, p. 15566). Association for Computational Linguistics.
- Wang, M., Cheng, B., Chen, J., Mercer, N., & Kirschner, P. A. (2017). The use of web-based collaborative concept mapping to support group learning and interaction in an online environment. *The Internet and Higher Education*, 34, 28–40. <https://doi.org/10.1016/j.iheduc.2017.04.003>
- Woolson, R. F. (2007). Wilcoxon signed-rank test. In *Wiley encyclopedia of clinical trials* (pp. 1–3). Oxford University Press.
- Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Qing-Long, & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10(5), 1122–1136. <https://doi.org/10.1109/JAS.2023.123618>
- Zhang, Y., Chen, Z., Zhang, W., & Chen, H. (2023). Making large language models perform better in knowledge graph completion. *arXiv preprint arXiv:2310.06671*.

Appendix. An example of concept maps generated by ChatGPT: classes and objects in PHP

Below is a complete example related to topic T_2 . The following text was given to both ChatGPT and the domain expert, we then see the maps generated by the ChatGPT and the expert and then the student assessment quiz.

A.1. Source text for both ChatGPT and domain experts

The pages that form the Web can be static or dynamic: they are static when they present the same contents to all users who access them, they are dynamic when they can show different contents in relation to the actions carried out by the users themselves or to the way in which the user is classified (e.g. administrator user). PHP is a general-purpose server-side scripting language used primarily to create dynamic websites. It is interpreted by the PHP engine resident on the Web server, i.e. a server that provides web pages.

PHP syntax is derived from the C language. A file with the .php extension must be created, and its beginning must be marked by the script opening tag `<?php`. This allows the PHP engine to recognize the script opening tag and execute the entire code until it finds the closing tag of the script, represented by `?>`. PHP scripts are executed on the server and send their responses to the client. These scripts dynamically produce HTML files, which are then sent to the client's browser for user visualization. PHP pages can also have parts of HTML code, which is referred to as embedded HTML: the code outside the PHP tags remains unchanged while the code inside the `<?php` tags is interpreted by the web server. The output is then sent to the browser in HTML format as a response to the requested page.

A.2. Object and classes in PHP ChatGPT version

PlantUML code

```
@startmindmap
* Classes and objects in PHP
** PHP
*** Server-side scripting language
*** Used to create dynamic websites
*** Interpreted by the PHP engine on the web server
*** Syntax similar to that of the C language
** PHP archive
*** Extension .php
*** Starts with the script opening tag "<?php"
*** Ends with the closing script tag ">"
** Embedded HTML
*** Feature of PHP pages
```

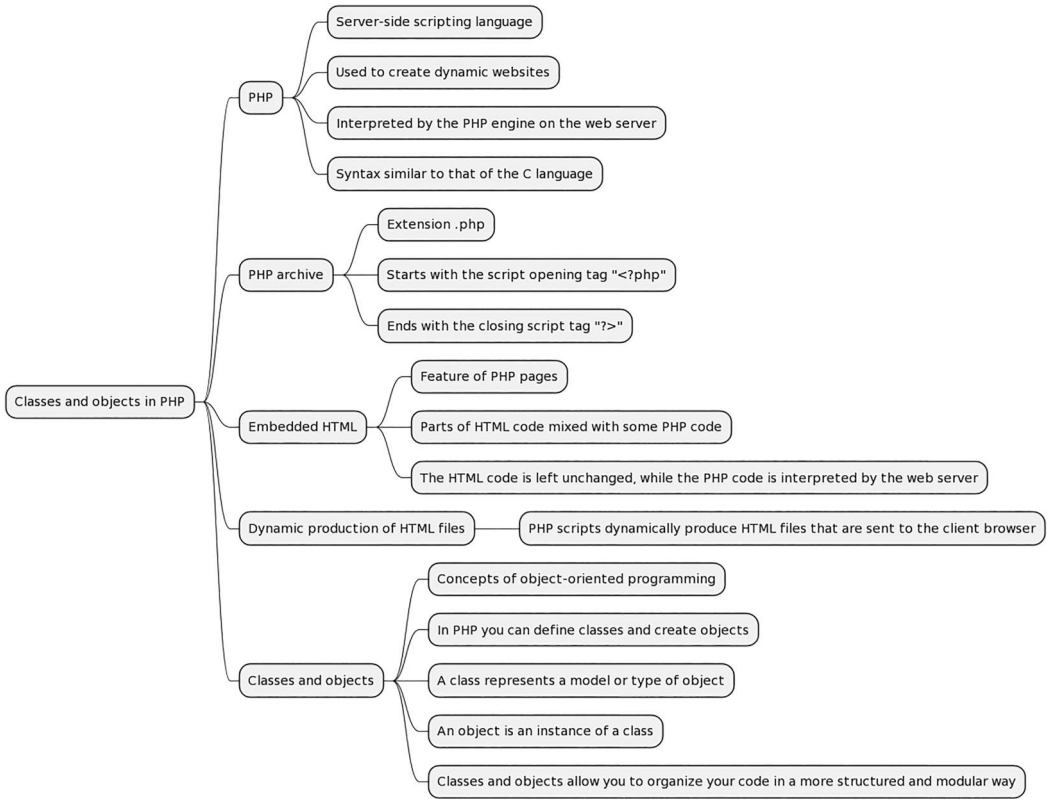


Figure A1. Concept Map generated by ChatGPT.

```

*** Parts of HTML code mixed with some PHP code
***. The HTML code is left unchanged, while the PHP code is interpreted by the web server;
** Dynamic production of HTML files
*** PHP scripts dynamically produce HTML files that are sent to the client browser
** Classes and objects
*** Concepts of object-oriented programming
*** In PHP you can define classes and create objects
*** A class represents a model or type of object
*** An object is an instance of a class
***. Classes and objects allow you to organize your code in a more structured and modular way
@endmindmap
    
```

A.3. Classes and objects in PHP expert version

PlantUML code

```

@startmindmap
* PHP
** Dynamic Web Pages
** include tag
*** <?php ... ?>
** language
*** Server-side scripting
**** File extension .php
**** Syntax derived from the C language
** .php engine
    
```

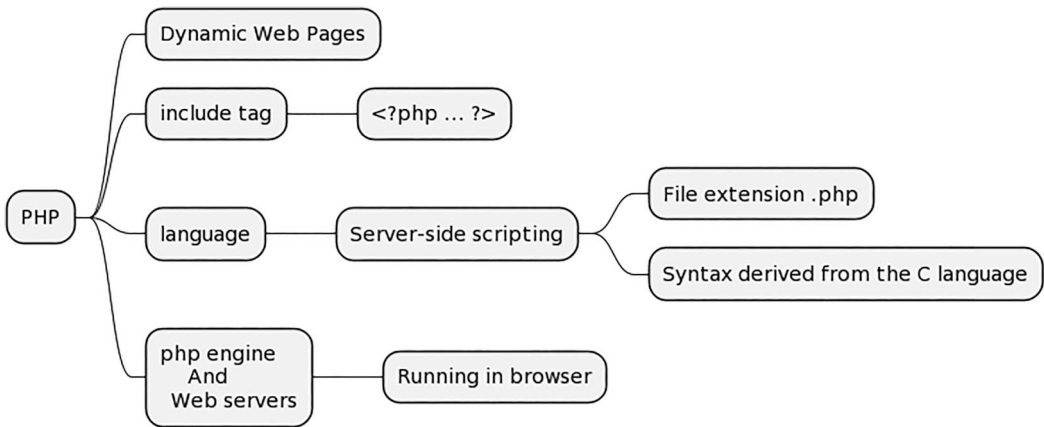


Figure A2. Concept Map generated by Domanin Expert.

And
Web servers;
*** Running in browser
@endmindmap

A.4. Multiple choice questionnaire related to classes and objects in PHP

- What is the difference between a static web page and a dynamic one?
 - (1) Static pages present the same content to all users who access them; dynamic pages show different content based on user actions.
 - (2) Dynamic pages present the same content to all users who access them; static pages show different content based on user actions.
 - (3) Static pages are developed in PHP, while dynamic pages are developed in HTML.
 - (4) No difference.
- What is PHP?
 - (1) A client-side “general-purpose” scripting language primarily used to create static websites.
 - (2) A server-side “general-purpose” scripting language primarily used to create dynamic websites.
 - (3) A language similar to JAVA used to create dynamic websites.
 - (4) A programming language that does not allow the development of websites but only standalone applications.
- Select the correct statement.
 - (1) PHP resides on the Web client, its syntax is derived from the JAVA language, and it requires creating a file with no extension.
 - (2) PHP resides on the Web server, its syntax does not derive from any programming language, and it requires creating a file that always starts with < ?PHP.
 - (3) PHP resides on the Web server, its syntax is derived from the C language, and it requires creating a file with the extension .PHP.
 - (4) PHP resides on the client, its syntax is derived from the C language, and it does not require the creation of any files.
- What does a .php file contain?
 - (1) It contains PHP code enclosed between < ?php and ? > tags that can include HTML code. The code enclosed between the tags is executed by the server, and the HTML code is left unchanged.
 - (2) It contains exclusively PHP code grouped between < ?php and ? > tags. The code enclosed between the tags is executed by the server.
 - (3) It contains PHP code grouped between < ?php and ? > tags and can include C code enclosed between < C/ > tags. The code enclosed between the tags is executed by the server.
 - (4) It contains a set of images that will be displayed to the end user.