EXPLORING THE ROLE OF CRITICAL THINKING IN THE USAGE OF LARGE LANGUAGE MODELS AMONG UNIVERSITY OF ANDORRA STUDENTS

M. Bleda Bejar[1], A. Dorca Josa[1], B. Oliveras Prat[2]

[1]University of Andorra (ANDORRA)
[2]Universitat Autònoma de Barcelona (SPAIN)

Abstract

In modern society, the integration of Large Language Models (LLMs), such as ChatGPT, has emerged as a turning point, especially within educational context. These technologies offer diverse benefits and redefine traditional learning paradigms. The objective of this investigation is to explore how students employ critical thinking skills across different stages of their interactions with LLMs. The research instrument employed for this study has been a survey conducted among students in the fields of Computer Science and Education at the University of Andorra. Through both quantitative and qualitative analyses, the research examines not only the frequency and perceptions of LLMs usage but also the thought processes, decision-making strategies, and problem-solving approaches that students employ when ensuring the reliability of the LLMs responses. While the data has been analyzed with consideration to the gender of the survey participants, it is important to interpret it through the lens of field-specific data rather than solely relying on gender-based data. Preliminary findings indicate a multifaceted result, with students demonstrating diverse approaches and perspectives. While some students employ critical thinking to evaluate the reliability and relevance of information generated by LLMs, others rely on these tools blindly. Furthermore, the study shows instances where critical thinking serves against potential biases or inaccuracies inherent in LLMs' responses. This research contributes to a deeper understanding of how students activate critical thinking skills when using LLMs. Moreover, it underscores the importance of improving critical thinking competencies within educational curricula to empower students with a comprehensive understanding of the capabilities and limitations of LLMs, as well as how to effectively use them.

Keywords: Critical thinking, Large Language Models, Higher education, Students confidence evaluation.

1 INTRODUCTION

Given that Large Language Models (LLMs) prioritize generating coherent content based on probabilities rather than truth, it becomes imperative to apply Critical Thinking (CT) when using such tools. This article aims to analyze how higher education students use these tools, their perceptions of their reliability and the strategies they employ to validate the information obtained.

A fundamental axis of this research is CT and to understand its essence it is important to explore various definitions. Bloom et al. [1] pioneered an innovative definition, describing CT as the mastery of skills such as knowledge, comprehension, application, analysis, synthesis and evaluation. Notably, the higher-order skills of analysis, synthesis, and evaluation are often regarded as integral to CT. Ennis [2] define CT as reasoned and reflective thinking that focuses on deciding what to believe or what to do. Skills represent knowing what to do and are divided into six aspects: interpretation, analysis, evaluation, inference, explanation and self-regulation. The dispositions represent the consistent internal motivation to act in a certain way and that a critical thinker must have the following intellectual attitudes: analytical, systematic, impartial, curious, sensible, truth-seeking and confident in reason. For Wood [3] CT is the process of using reasoning to discern what is true and what is false in the information we encounter daily. It involves being familiar with logic and logical fallacies, separating facts from opinions, being fair and open-minded, asking questions to uncover truth and motivations behind arguments, and self-regulating to avoid logical fallacies and rationalizations. The principal aim of CT is to arrive at the truth by examining arguments objectively, avoiding emotional attachments to opinions, and being open to exploring all ideas and viewpoints, even those that may contradict one's own beliefs. CT encompasses a range of competencies that are essential for effective reasoning, problem-solving, and decision-making. These competencies can be categorized into cognitive competencies and personal competencies, each playing a crucial role in fostering CT skills [4]. Cognitive competencies involve skills like dissecting information, analyzing, synthesizing, and understanding it, while personal competencies include traits like tolerance of ambiguity, independent thinking, perseverance, and curiosity. These competencies interact with each other and are essential

ISBN: 978-84-09-62938-1

components of CT. Developing and honing these competencies can enhance an individual's ability to think critically, analyze information effectively, and make informed decisions in various contexts. In addition, Benesch [5] defines CT as the capacity to question assumptions, consider various perspectives, and employ logical reasoning in addressing complex problems, with a focus on improving rational thinking, make sound judgment, and effective problem-solving through evidence examination, bias identification, and consideration of alternative viewpoints. Socratic questioning can enhance CT skills in students by encouraging them to engage in deeper analysis, evaluate different perspectives, and develop a more comprehensive understanding of complex issues. By asking thought-provoking questions that challenge assumptions and stimulate reflection, students are prompted to think critically, consider evidence, and articulate their reasoning. This process helps students to develop analytical skills, improve their ability to make logical connections, and become more adept at evaluating information [6]. Kuhn [7] emphasizes the importance of understanding CT within a developmental framework and highlights the role of second-order cognition in the development of CT skills. And the need for a deeper understanding of how cognitive processes evolve in children and adolescents to facilitate CT. According to Paul and Binker [8], CT involves reflecting on one's own thought processes. Which shows that there is a clear connection between metacognitive knowledge and CT skill. In that way, Kuhn [9] emphasizes the importance of the coordination of theories and evidence in scientific thinking. This coordination involves thinking about theories, evidence, and the interaction between them, reflecting a metacognitive and strategic approach to understanding and evaluating information. This process involves generating multiple theories, coordinating evidence with them, and being able to justify one's conclusions based on the available evidence.

Another core aspect of this research is the Artificial Intelligence (AI) and particularly LLMs. The concept of AI has been a relevant subject for decades, tracing its roots back over 45 years when early discussions attempted to define its scope and implications. At its core, AI seeks to emulate human intelligence in machines, providing them with the ability to reason, learn, and adapt much like their human counterparts [10]. These activities involve diverse cognitive processes, spanning problem-solving, speech recognition, acquiring knowledge, strategic planning, and sensory interpretation [11]. AI enables machines to tackle these tasks with unprecedented efficiency, revolutionizing industries and changing societal norms. Today, AI emerges into a multifaceted field spanning various disciplines such as Natural Language Processing (NLP), machine learning, and speech recognition. NLP is a branch of AI that looks for the interaction between computers and human language. Its objective is to provide computers with the ability to comprehend, interpret, and generate human language in a nuanced and contextually relevant manner. Several key aspects are included: text processing, language understanding, and language generation. However, despite remarkable progresses in NLP research, significant challenges persist, including the complexities of ambiguity, context comprehension, and language diversity. On the other hand, ethical considerations arise with concerns about bias in linguistic models. The need for fairness, transparency and accountability in AI systems underscores the importance of ethical discourse and responsible innovation in the field. Central to the effectiveness of these models is the underlying principle of probabilistic modelling, rooted in the rich tapestry of probability theory [12]. In 2017, the advent of the Transformer architecture marked a moment in the evolution of neural network design [13]. This framework has revolutionized sequence transduction tasks by leveraging attentional mechanisms and avoiding traditional recurrent or convolutional layers. This architecture improves numerous tasks, including automatic translation, delivering remarkable results across various standard datasets. It also cuts down on training time and boosts parallelization, showing how effective it is. The architects of the Transformer conducted experiments to dissect the intricacies of its components, providing invaluable insights into its inner workings. Notably, architectures like ChatGPT owe much of their efficacy to the foundational principles embodied by the Transformer [14].

There are various instances of AI applications in education today. Entities like UNESCO recommend integrating AI, particularly LLMs, into educational practices [15]. However, ethical issues such as student plagiarism have arisen, highlighting the need for additional research to guarantee the ethical and efficient usage of tools like ChatGPT. Incorporating AI into education offers numerous advantages, including improved learning outcomes, increased efficiency and productivity. It also brings greater accessibility to education, particularly for marginalized or underserved communities. Nevertheless, there are potential issues, such as concerns regarding data privacy and security, the possibility of bias in AI algorithms, and the displacement of educators. It is imperative to ensure that the integration and deployment of AI in education adhere to principles of human rights and social justice [16]. Achieving this requires active engagement from educational institutions and authorities. These models need to be incorporated into education to supplement and enrich the learning experience rather than substituting it [17]. The growing use of AI in personalized learning, analytics, and research assistance is perceived as advantageous for society. Nonetheless, it is essential to discuss on the ethical and social consequences, necessitating an interdisciplinary collaboration among educators, IT experts, policymakers, and other relevant stakeholders [18]. LLMs have the potential to enhance student engagement and create interactive materials, but their responsible use is necessary, requiring the avoidance of bias and the assurance of fairness. Consequently, integrating CT and problem-solving skills into education becomes essential [19]. ChatGPT generates precise and well-structured answers to university-level questions, and it can create challenging CT questions and assess answers across various disciplines [20]. In that way, the responses generated by this tool demonstrate excellent results in the areas of critical thinking, higher order thinking and economics [21].

A central element is to examine the perspectives of users regarding LLMs. Users perceive human and Artificial Intelligence-Generated Content (AIGC) as equally credible, with human-generated content less clear and engaging than AIGC. Educating about LLMs is important to help people understand and assess the potential risks of these tools. To do this, the responsible use of AIGC must be promoted by encouraging prudence, TC and media skills [22]. Users are advised to critically evaluate information sources and exercise caution, even when the source of the content is apparently reliable. Teachers, in their part, show a favorable stance, highlighting advantages such as structured information, customized feedback, and enhanced CT [23]. As perceived by students, the strengths of LLMs in education are their potential to improve learning practices, personalize educational experiences and provide instant support, which improves the overall learning experience and engagement [24].

CT in academia remains a pertinent and contemporary subject, particularly in our digital age where instant access to information is omnipresent. Therefore, it's crucial to examine how students activate CT skills while using LLMs. Consequently, this study aims to explore the usage of LLMs among higher education students in Computer Science (CS) and Education, comparing their usage patterns. Additionally, it seeks to evaluate perceptions of reliability regarding LLM responses and validation strategies when doubts arise about their content.

The specific objectives of this article are:

Objective 1: Study and evaluate whether there are significant differences in the use of LLMs between the students of CS and Education.

Objective 2: Describe the reliability that CS and Education bachelor students give to LLMs responses.

Objective 3: Analyze the strategies used by the students to validate the answers given by LLMs.

2 METHODOLOGY

In November 2023, a mixed-method survey was conducted among higher education students at the University of Andorra (UdA). The instrument designed underwent a validation process, assessing the degree of univocity, pertinence and importance, as detailed in [25], involving the expertise of five professionals from various fields. The survey participants consisted of students enrolled in the BSc in CS and the BSc in Education. From an initial pool of 129 individuals, there were 83 respondents, giving a margin of error of 6% with a confidence level of 95%. The distribution of the surveyed students is detailed in Table 1.

Table 1. Distribution of users in the dataset by academic level and field

Academic level	*Education*	*CS*	*Total*
1	26	13	***39***
2	14	9	***23***
3	16	5	***21***
Total	***56***	***27***	***83***

The survey was conducted in multiple sessions across all three courses within both disciplines. Each participant responded anonymously, providing honest responses and thereby enhancing the reliability of the collected data. The instrument, comprising a total of 15 questions, was designed to cover a wide range of aspects related to perceptions and usage of LLMs in an educational context. Following the completion of the survey, responses provided by participants were compiled into a Comma-separated values (CSV) file for data analysis and interpretation.

To analyze the results of the questionnaire, two methods have been used. On one hand, quantitative analyses with graphical representations are presented for questions with closed options. On the other hand, for open questions, a series of systemic networks have been designed to classify and categorize the responses of the students. A systemic network is an instrument designed by Bliss et al. that gathers the different meanings behind an expression or drawing. This instrument derives from systemic linguistics, which is concerned with the description and representation of the meaning of the semantic resources of language [26].

Objective 1 was addressed by conducting a quantitative analysis of responses to a question which evaluated the frequency of LLM usage across various contexts. This approach aimed to identify significant differences between students from CS and Education. Given that the assessed question employs a Likert scale, the expected frequencies are sufficiently large and data are independent of one another, a Pearson's χ^2 test of independence is used to answer this objective.

To achieve Objective 2, a qualitative analysis has been carried out using student responses to Likert scale question "Responses provided by LLMs are reliable", along with their corresponding free-text justifications and compared to the answers of a sub-question assessing students' perceptions regarding LLMs, which is "I understand perfectly how LLMs work". By employing a coding table generated from the data and adopting a bottom-up coding approach, different codes were identified.

For Objective 3, a qualitative analysis was conducted using student responses to the free-text question "How do you guarantee the reliability of the responses generated by language models when you have doubts about their content?". Employing the same coding approach as in objective 2, other codes emerged.

3 RESULTS

To address Objective 1, the analysis of responses to the questions "How often have you used language models to: [Analyze the content of a text]; [Code analysis]; [Understanding classroom content]; [Save time in content generation]; [Generating arguments for a debate]; [Generate code]; [Generating the structure of an algorithm]; [Generate text]; [Refine work methodologies]; [Code improvement]; [Improve text quality]; [Practicing languages]; [Resolve doubts]; [Review classroom work]" was encompassed.

Applying the Pearson's χ^2 test of independence revealed a statistically significant difference in the usage frequency of LLMs on four sub-questions, which are code analysis, generate code, generating the structure of an algorithm and code improvement.

Concerning the code analysis sub-question, a significant difference emerges between CS surveyed students and Education surveyed students, $\chi^2 = 41.3$, with a p-value of < 0.001. CS students showed a higher propensity to use LLMs frequently to analyze code, while Education students were more inclined to state never employing them for such purposes, as illustrated in Table 2.

Table 2. Contingency table for Code analysis

Code analysis	*Never (=1)*	*Rarely (=2)*	*Occasionally (=3)*	*Often (=4)*	*Very often (=5)*	*Total*
CS	4.8%	4.8%	1.2%	10.8%	10.8%	***32.5%***
Education	43.5%	12.1%	9.6%	1.2%	1.2%	***67.5%***
Total	***48.3%***	***16.9%***	***10.8%***	***12%***	***12%***	***100%***

In reference to generating code, there exists a statistically significant difference between CS students and Education students, $\chi^2 = 36.7$, $p < 0.001$. CS students commonly reported using LLMs to Generate code, whereas Education students tended to express never employing them for this task, as outlined in Table 3.

Table 3. Contingency table for Generate code

Generate code	*Never (=1)*	*Rarely (=2)*	*Occasionally (=3)*	*Often (=4)*	*Very often (=5)*	*Total*
CS	4.8%	6%	8.5%	9.6%	3.6%	***32.5%***
Education	50.7%	10.8%	4.8%	0%	1.2%	***67.5%***
Total	***55.5%***	***16.8%***	***13.3%***	***9.6%***	4.8%	***100%***

In relation to the sub-question of generating the structure of an algorithm, a significant difference is observed between CS students and Education students, $\chi^2 = 32.5$ and with a p-value of < 0.001. As demonstrated in Table 4, students from CS were more likely to say they use LLMs Occasionally for generating the structure of an algorithm, while Education students were more likely to indicate never using them for this purpose.

Table 4. Contingency table for Generating the structure of an algorithm

Generating the structure of an algorithm	***Never (=1)***	***Rarely (=2)***	***Occasionally (=3)***	***Often (=4)***	***Very often (=5)***	***Total***
CS	6%	6%	9.6%	7.2%	3.6%	***32.5%***
Education	47%	15.8%	3.6%	1.2%	0%	***67.5%***
Total	***53%***	***21.8%***	***13.2%***	***8.4%***	***3.6%***	***100%***

When comparing the frequency of LLMs usage to improve code among CS students and Education students, a significant difference emerges, $\chi^2 = 35$, $p < 0.001$. CS students tended to report frequent usage of LLMs for code improvement, whereas Education students were more inclined to indicate never using them for this purpose, as illustrated in Table 5.

Table 5. Contingency table for Code improvement

Code improvement	***Never (=1)***	***Rarely (=2)***	***Occasionally (=3)***	***Often (=4)***	***Very often (=5)***	***Total***
CS	4.8%	4.8%	7.2%	10.8%	4.8%	***32.5%***
Education	48.3%	12.1%	4.8%	2.4%	0%	***67.5%***
Total	***53.1%***	***16.9%***	***12%***	***13.2%***	***4.8%***	***100%***

While it's evident that four sub-questions reveal significant differences, they predominantly align to CS area rather than Education. These questions focus on improving, generating and analyzing code as well as generating the structure of an algorithm. Conversely, in the more generalized sub-questions, applicable to both domains, there are no statistically significant differences. This suggests that students from both fields employ these tools with comparable frequency.

In relation to Objective 2, justifications for Likert scale responses to the statement "Responses provided by LLMs are reliable" could be provided by students. However, 66.3% of respondents did not offer any justification for their answers. Due to the few justifications provided by the students, it makes no sense to analyze whether significant differences exist between the responses from the two areas. The responses given by students were translated from Catalan. Among those who did provide justification, their explanations are summarized in Table 6.

Table 6. Code table for Objective 2

Answer	***Code***	***Example***	***Frequency***
Agree	Specific purpose	“I usually use AI to reformulate texts (explanations) that I’ve written but I don’t like how they look.”	1
	Trust in AI	“Since it's an AI, I guess I trust it.”	5
	Cross-reference information	“It's usually reliable but you have to verify the information.”	2
	Depends on the area	“Depending on the area I'm moving in, it's useful or not.”	1
Neither agree nor disagree	Other points of view	“Most of the time the information is not reliable, but it can help you see other points of view. (Speaking of programming, since that's what I use AI with)”	1
	Tautological argument	“There are things that are true and others that are not so much (I guess).”	5

	Depends on the prompt	“Sometimes it is necessary to specify something specific to him so that he takes it into account and does not forget and even doing this he forgets to do it.”	4
	Cross-reference information	“They can be reliable, but as long as you check the information on another reliable website.”	3
	Need for prior knowledge	“It depends on the question you ask them, anyway you can't trust 100% of everything that comes out of information, since you have to have some prior knowledge.”	1
	Uncertainty of the sources	“They are partially reliable, but we do not know exactly the sources from which the information has been taken, so it is not useful to quote or search if they are true.”	2
Disagree	Depends on the area	“If it's a very specific question, etc., there's a higher chance that the whole text is reliable, but you have to check the text a lot.”	2
	Depends on the source	“It really depends on what language and where it gets the information. The intelligence must be up to date and if connected to the internet know how to recognize if the sources are reliable.”	1
	Need for prior knowledge	“You don't really know where they get them from so you can't really trust these things. Only if you know the answer.”	1
NS/NC	Mistrust	“I don't know, but I would say that it is not reliable.”	1

Among respondents who have justified their position there is a notable degree of trust in AI among students, there are also significant concerns about its reliability. The variability of AI performance, coupled with the importance of verifying information from multiple sources, highlights the need for a nuanced understanding of AI's capabilities and limitations.

However, most of the respondents did not justify their position on the reliability of the answers generated by the LLMs. This observation is remarkable, especially given their apparent inclination to answer in the affirmative in terms of understanding how these tools work, as illustrated in Figure 1.

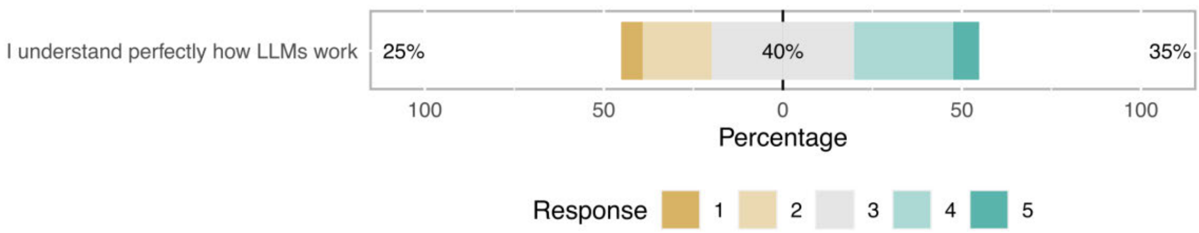

Figure 1. Students’ perspectives on understanding how LLMs work

This paradox suggests a potential gap between respondents' perceived understanding of how the LLMs works and their ability to articulate reasoned justifications for their beliefs. Furthermore, the absence of strong justifications raises questions about the extent to which respondents critically engage with the information provided by these technologies. A more complete understanding of the functionality of LLMs would likely allow respondents to provide more nuanced and reasoned assessments of the reliability of LLM-generated responses. Therefore, the discrepancy between respondents' perceived understanding and their demonstrated ability to justify their beliefs underscores the need to improve students' critical thinking skills.

Objective 3 involved analyzing responses to the question, “How do you guarantee the reliability of the responses generated by language models when you have doubts about their content?”. The students' responses were translated from Catalan. Out of the 83 respondents, 6 did not provide a valid answer. The subsequent valid responses are summarized in Table 7.

Table 7. Code table for Objective 3

Categories	Code	Example	Frequency
They justify	Cross-reference with other sources	“What I do is look for the same thing from different sources to make a comparison and determine how truthful or not an answer I doubt is.”	22
	Cross-reference with internet information	“I look for it on the Internet, whether the information is correct or not, because many times these AI do not give you the exact information, for example in the case of ChatGPT.”	17
	Cross-reference with prior knowledge	“Depending on what you know about the subject you can know if it's reliable or not but if you don't know anything about what you're asking for you have to be careful because you can't guarantee it's reliable.”	11
	Cross-reference with references	“Ask for the bibliography of the text and read the pages.”	10
	Cross-reference with other LLMs	“Comparing different LLMs to see if they all say the same or not.”	6
They do not justify	Lack of verification	“I do NOT verify.”	10
	Trust in AI	“I don't usually doubt the reliability of the answers.”	8

When queried about their strategies for addressing doubts concerning the reliability of LLM-generated responses, a significant majority (78%) emphasize the necessity of verification. This observation underscores a proactive approach among students, who recognize the importance of corroborating information obtained from LLMs through external validation processes. Moreover, a subset of students highlights the value of leveraging their previous knowledge to validate LLM-generated content. By drawing upon their prior understanding of a subject matter, these students advocate for a contextualized assessment of LLM responses, suggesting that one's familiarity with a topic can serve as a valuable benchmark for evaluating the accuracy and credibility of AI-generated information. This nuanced perspective demonstrates a critical mindset among students when they doubt about the content generated by those tools.

4 CONCLUSIONS

Analysis reveals no significant differences in the usage of LLMs between CS and Education bachelor students. Differences are only detected in specific CS questions such as generating, improving and analyzing code, as well as generating the structure of an algorithm. It is necessary to emphasize the importance of gaining a comprehensive understanding of LLMs and their functionalities is essential to promote more informed user engagement. An important observation from the survey data is that a significant percentage of surveyed students (66.3%) do not provide substantial justification for their confidence in the reliability of LLMs. This lack of justification highlights a potential gap between perceived trust and a satisfactory understanding of these tools. It is intriguing to note that while LLMs prioritize the construction of coherent sentences over reliability, both CS and Education students share a similar positive perspective on the reliability of these technologies. Additionally, it should be noted that while students tend to check answers when in doubt, their reliance on LLM-generated answers may result in them overlooking the need for verification when there are no doubts. It is therefore imperative that students have knowledge of how LLMs work, enabling them to critically evaluate the information they receive. In this regard, strengthening students' critical thinking skills is critical to bridging this gap and fostering a more discerning engagement with LLM-generated content.

On the other hand, reliance on the Internet as a primary source of scientific information may affect young people's understanding of credibility in several ways [27]. The ease of access to large amounts of information online can lead to challenges in discerning credible sources from misinformation or biased content. CT is based on knowledge and directly related to a comprehensive understanding of the subject [28]. It is essential to help students examine the reliability of information obtained through LLMs by cross-referencing with their prior knowledge and other sources, enabling them to discern between scientifically valid information and unsubstantiated claims. In addition, students must understand how LLMs work in order to use the tool effectively and assess the reliability of its information. Consequently, the advent of LLMs emphasizes further strengthening of critical thinking skills.

Therefore, the next crucial step would be to expand this experiment to include other courses offered at the UdA, and thus obtain a comprehensive view of the only public university in the country.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to the University of Andorra for allowing us to conduct the survey and for providing us with the necessary facilities and resources. We also want to give a special thanks to the students who participated in the survey.

REFERENCES

[1] B. S. Bloom, M. B. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl, "TAXONOMY OF EDUCATIONAL OBJECTIVES The Classification of Educational Goals.," *New York: Longmans Green*, 1956

[2] R. H. Ennis, "A taxonomy of critical thinking dispositions and abilities.," *WH Freeman/Times Books/Henry Holt & Co*, 1987

[3] R. Wood, "Critical thinking," *Retrieved August*, vol. 12, p. 2015, 2002

[4] M. Fahim and M. Pezeshki, "Manipulating critical thinking skills in test taking," *International Journal of Education*, vol. 4, no. 1, p. 153, 2012

[5] S. Benesch, "Thinking critically, thinking dialogically," *Tesol Quarterly*, vol. 33, no. 3, pp. 573–580, 1999

[6] L. Elder and R. Paul, "The role of socratic questioning in thinking, teaching, and learning," *The Clearing House*, vol. 71, no. 5, pp. 297–301, 1998

[7] D. Kuhn, "A developmental model of critical thinking," *Educational researcher*, vol. 28, no. 2, pp. 16–46, 1999

[8] R. W. Paul and A. Binker, "Critical thinking: What every person needs to survive in a rapidly changing world.," *ERIC*, 1990

[9] D. Kuhn, "Children and adults as intuitive scientists," *Psychological review*, vol. 96, no. 4, p. 674, 1989

[10] R. Kurzweil, R. Richter, R. Kurzweil, and M. L. Schneider, "The age of intelligent machines.," *MIT press Cambridge*, 1990, vol. 580.

[11] R. E. Bellman, "Artificial intelligence: Can computers think?," 1978.

[12] Y. Bengio, R. Ducharme, and P. Vincent, "A Neural Probabilistic Language Model," *Advances in neural information processing systems*, vol. 13, 2000.

[13] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention Is All You Need," *Advances in neural information processing systems*, vol. 30, 2017.

[14] OpenAI, *Chatgpt*, 2022. [Online]. Available: https://chat.openai.com.

[15] E. Sabzalieva and A. Valentini, "ChatGPT and Artificial Intelligence in higher education: Quick start guide," 2023.

[16] UNESCO, "BEIJING CONSENSUS on Artificial Intelligence and Education," 2019.

[17] E. Kasneci, K. Seßler, S. Kuchemann, et al., "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education," *Learning and individual differences*, vol. 103, p. 102 274, 2023.

[18] B. L. Liu, D. Morales, J. Roser-Chinchilla, et al., "Harnessing the Era of Artificial Intelligence in Higher Education: A Primer for Higher Education Stakeholders," 2023.

[19] C.E.George-Reyes,E.O.López-Caudana, and M.S.Ramírez-Montoya, "Research competencies in university students: Intertwining complex thinking and Education 4.0," *Contemporary Educational Technology*, vol. 15, no. 4, ep478, 2023

[20] T. Susnjak, "ChatGPT: The End of Online Exam Integrity?," *arXiv preprint arXiv:2212.09292*, 2022.

[21] C. K. Lo, “What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature,” *Education Sciences*, vol. 13, no. 4, p. 410, 2023.

[22] M. Huschens, M. Briesch, D. Sobania, and F. Rothlauf, “Do You Trust ChatGPT? – Perceived Credibility of Human and AI-Generated Content,” 2023. arXiv: 2309.02524 [cs.HC].

[23] F. R. Baskara, A. D. Puri, and A. R. Wardhani, “ChatGPT and the Pedagogical Challenge: Unveiling the Impact on Early-Career Academics in Higher Education,” *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, vol. 5, no. 3, pp. 311–322, 2023.

[24] M. Irfan, L. Murray, and S. Ali, “Insights into Student Perceptions: Investigating Artificial Intelligence (AI) Tool Usability in Irish Higher Education at the University of Limerick,” *Global Digital & Print Media Review*, VI, pp. 48–63, 2023.

[25] F. X. Carrera Farran, E. Vaquero Tió, M. À. Balsells Bailón, et al., “Instrumento de evaluación de competencias digitales para adolescentes en riesgo social,” *Edutec: revista electrónica de tecnología educativa*, 2011.

[26] J. Bliss, M. Monk, J. Ogborn, and P. J. Black, “Qualitative data analysis for educational research: A guide to uses of systemic networks,” *London: Croom Helm*, 1983.

[27] D. Pimentel, “Learning to evaluate scientific evidence in the age of digital information,” *eBook of Synopses*, p. 43, 2022

[28] W. A. Sandoval and K. A. Millwood, “The quality of students' use of evidence in written scientific explanations.,” *Cognition and Instruction*, 23 (1), pp. 23-55, 2005.