**PAPER TITLE**

**Zishan Ahmed1** **· Shakib Sadat Shanto1 · Akinul Islam Jony1**

**Abstract**

The advent of AI chatbots such as ChatGPT has revolutionized the field of education by offering convenient information accessibility, although accompanied by concerns regarding the cultivation of critical thinking skills. Nevertheless, a lack of extensive research remains regarding the extent to which learners can develop critical thinking skills in a certain subject through the utilization of ChatGPT. This research seeks to investigate the potential of ChatGPT in fostering the development of critical thinking abilities. ChatGPT was presented with a series of inquiries of increasing levels of complexity within the intricate realm of cybersecurity. The responses were subjected to analysis using Lee's Model of thinking levels, which involved categorizing them into recall, rationalization, or reflectivity. The objective is to conduct a complete assessment of the analytical capacities of ChatGPT in light its growing integration into educational settings. The findings suggest that ChatGPT demonstrates a high level of critical thinking skills, particularly in the context of real-world situations.

**Keywords** Generative AI · ChatGPT · Critical Thinking· Lee's Model · Cybersecurity · Education Technology

**1 Introduction**

The landscape of education is being fundamentally reshaped by the emergence of new generative artificial intelligence technologies, with systems like ChatGPT at the forefront (Atlas, 2023; Grassini, 2023). The emergence of generative AI signifies a fundamental departure from conventional retrieval-based models. In contrast to the conventional approach of searching and extracting existing data, generative models possess the capability to generate entirely original content that is specifically customized to a certain prompt. However, concerns remain about overreliance on AI-generated content which may lack depth, nuance, and accuracy across different domains (Alqahtani et al., 2023; Wu et al., 2023). The increasing integration of generative chatbots into academic practices necessitates a thorough examination of their capabilities and limitations.

The cultivation of critical thinking skills, encompassing the capacity to examine information, question assumptions, evaluate evidence, and formulate well-founded conclusions, is widely acknowledged as a fundamental competency that education should foster in order to facilitate students' academic and professional success (Mahanal et al., 2019). Nevertheless, the mere act of memorizing things in a mechanical and repetitive manner is inadequate when it comes to accessing more profound levels of learning and cultivating the critical attitude necessary for the efficient application of knowledge. This observation is particularly significant in light of the increasing prevalence of AI chatbots, such as ChatGPT, within the field of education. These chatbots have the capability to deliver knowledge to students in a scalable manner, catering to their immediate needs. Prior studies have explored the impact of technology like AI tutors and learning analytics systems on critical thinking (Alam, 2021; Zawacki-Richter et al., 2019). However, research focused specifically on generative chatbots like ChatGPT is still emerging (Exintaris et al., 2023; Hill-Yardin et al., 2023). A comprehensive assessment is necessary for determining the level of thoughtfulness and cognitive capabilities exhibited by Generative AI tools across different domains. This evaluation seeks to determine whether these tools can effectively assist students or learners in acquiring a thorough understanding of various subjects, while also facilitating the development of their critical thinking skills and the ability to apply acquired knowledge in real-world contexts.

Hence, evaluating ChatGPT's capacity for reasoning and promoting critical thinking level is an open research question. So, this study aims to address the raised research question “Can ChatGPT promote critical thinking level?”. To answer this question, we undertake an evaluation of the cognitive levels manifested in the responses generated by ChatGPT, employing Lee's model of thinking level (Lee, 2000) as a methodological framework. In this work, we examine the field of cybersecurity as a complex area and analyze the responses provided by ChatGPT to questions of varying difficulty levels, ranging from basic to intermediate and advanced. The critical thinking capabilities of ChatGPT were subsequently examined by categorizing the responses according to Lee's model.

***Paper Structure***

**2 Literature Review**

Critical thinking is the cognitive capacity to systematically evaluate and analyze information logically, question underlying assumptions, assess the strength of evidence, and arrive at well-founded and valid conclusions (Pithers & Soden, 2000; Suter, 2011). The concept of critical thinking has been explored and developed by various disciplines, including philosophy, education, and psychology (Pithers & Soden, 2000). This cognitive talent empowers students to surpass conventional memorization of facts and attain a more profound understanding of concepts across various academic fields (Hove, 2011). By developing critical thinking capabilities, learners can synthesize and apply knowledge to novel contexts rather than merely parroting data. The practice of reflecting on one's ideas can enhance students' learning and professional efficacy, ultimately enabling them to go beyond just information intake (Jony et al., 2017). Therefore, the process of reflecting on one's thoughts can encourage learning, promote self-analysis, and facilitate the resolution of real-world problems. The focus of higher education should prioritize the cultivation of critical thinking skills.

Developing strong critical thinking skills is widely acknowledged as a vital objective of higher education, necessary to prepare graduates for professional success. Research indicates that the capacity to engage in critical thinking, question assumptions, assess evidence, and employ logical reasoning is crucial for students to excel in their academic pursuits and effectively apply their knowledge in practical situations (Brookfield, 2011; Frykholm, 2021; Ma & Li, 2022). Several studies show that the development of critical thinking ability is an essential component of higher education, as it is believed to contribute to improved academic performance (D’Alessio et al., 2019; Ghazivakili et al., 2014; Ramos, 2018).

The sector of education is currently undergoing significant transformation due to recent advancements in Generative AI tools and chatbots, such as ChatGPT (Lo, 2023; Rahman & Watanobe, 2023). These technologies provide students with the convenience of accessing relevant knowledge and information through the use of prompts directed towards AI-powered chatbots. The advent of this technological innovation has undoubtedly provided students an opportunity to actively participate in academic research, seek support for their educational tasks, and conveniently obtain educational resources. Nevertheless, along with these advantages, there is an emerging concern surrounding the response produced by ChatGPT in diverse domains. ChatGPT is built upon a robust collection of datasets and possesses the capability to generate responses that closely resemble human-like answers in response to the questions it receives (Fuchs, 2023; Guo et al., 2023). Due to its dependence on a pre-trained language model and its capacity to generate responses based on the current data corpus, there is an increased likelihood that the generated responses may display repetition and be influenced by inherent biases. Given the widespread utilization of ChatGPT among students as a helpful tool for their scholarly endeavors, there arises an increasing need to conduct thorough investigations that assess the level of quality and influence of the responses generated by ChatGPT. So, this calls for an analysis of whether the outputs produced by ChatGPT promote the development of critical thinking abilities or, on the contrary, limit the progress of such capabilities. In this study, we aim to measure the thinking level of the responses provided by the ChatGPT by using the popular Lee’s model of thinking level (Lee, 2000).

**3 Methodology**

***Description & Diagram***

**3.1 Models**

We used the well-known Lee's model of thinking level (Lee, 2000), which consists of three levels of thinking, to determine whether ChatGPT's responses demonstrate critical thinking or not. Table 1 illustrates Lee's model of thinking.

**Table 1** Lee’s Model

|  |  |  |  |
| --- | --- | --- | --- |
| Levels of Thinking | Degree of Levels | Lee’s Model | Description of Lee’s Model |
| Level 1 | Lowest | Recall | Echo (recall) the same  content (knowledge) |
| Level 2 | Intermediate | Rationalization | Rationalize his or her  Thinking about the content  (knowledge) |
| Level 3 | Highest | Reflectivity | Reflect his or her own  Thinking beyond the content (knowledge) |

Table 2 depicts an alternative model, known as Bloom's Taxonomy of Educational Objectives (Adams, 2015), which shares similarities with Lee's Model of thinking level. However, Bloom's Taxonomy offers a more detailed breakdown of the many levels of thinking. The concept under consideration encompasses six distinct stages of cognitive thinking.

**Table 2** Bloom’s Taxonomy of Educational Objectives

|  |  |  |  |
| --- | --- | --- | --- |
| Levels of Thinking | Degree of Levels | Bloom’s Taxonomy | Description of Bloom’s Taxonomy |
| Level 1 | Lowest | Knowledge | Echo (recall) the same content (knowledge) |
| Level 2 | Intermediate | Comprehension | Understanding the content |
| Level 3 | Application | Apply content (knowledge)  in new situation |
| Level 4 | Analysis | Break down the content (knowledge) |
| Level 5 | Synthesis | Reassemble contents  together |
| Level 6 | Highest | Evaluation | Justify with his own thinking beyond the content |

Table 3 provides a comparison of the degree of levels between Lee's model of thinking and Bloom’s Taxonomy of Educational Objectives model. Although Bloom's Taxonomy consists of six levels of thinking, when considering the degree of thinking levels, both models are the same. Lee's Model integrates levels 2 through 5 of Bloom's Taxonomy Model into a unified level termed Rationalization. Therefore, this work exclusively focuses on the utilization of Lee's Model for experimental purposes.

**Table 3** Comparison between Bloom’s Taxonomy & Lee’s Mode

|  |  |  |  |
| --- | --- | --- | --- |
| Levels of Thinking | Degree of Levels | Bloom’s Taxonomy | Lee’s Model |
| Level 1 | Lowest | Knowledge | Recall |
| Level 2 | Intermediate | Comprehension | Understanding the content |
| Level 3 | Application | Rationalization |
| Level 4 | Analysis |
| Level 5 | Synthesis |
| Level 6 | Highest | Evaluation | Reflectivity |

**3.2 Data Collection**

In order to evaluate ChatGPT's capacity for critical thinking, this study describes a comprehensive data collection approach focusing on cybersecurity. Cybersecurity was selected as the focus of our study due to its inherent complexity and wide-ranging nature, making it an ideal candidate for evaluating ChatGPT's analytical capabilities.

To initiate the process of data collection, we employed ChatGPT to provide us with an in-depth overview of cybersecurity, encompassing its core aspects, prospective developments, and fundamental concepts. Based on this initial description, we subsequently posed a series of basic-level questions to ChatGPT and recorded its responses. We then moved on to intermediate-level questions, derived from the previous responses, to evaluate the model's level of understanding and ability to acquire new information. To conclude the investigation, we presented ChatGPT with cybersecurity scenario-based queries, with a focus on problem-solving abilities and the capacity to demonstrate reflective thought beyond pre-existing content.

This process of data acquisition is the foundation of our research, allowing us to determine the extent to which ChatGPT can engage in critical thinking within the complex domain of cybersecurity.

Initial prompt : *Describe cybersecurity in terms of its most important concepts, aspects, and prospects.*

**Table 4** Prompt Pattern

|  |  |  |
| --- | --- | --- |
| Prompt Category | Prompt description | Prompt Samples |
| First category of prompt | Basic level questions on cybersecurity | - What are the best practices for creating strong passwords? |
| - What is phishing and how can you avoid it? |
| - How can you protect your personal information online? |
| Second category of prompt | Advanced to intermediate-level questions on cybersecurity | - Explain the difference between network security, endpoint security, cloud security, application security, and physical security. |
| - How does two-factor authentication work and why is it more secure than passwords alone? |
| - How can organizations ensure that their employees are practicing good cyber hygiene? |
| Third category of prompt | Real life scenario-based questions on cybersecurity | - You are working on a sensitive project for your company and you need to access a file on a shared drive. You notice that the file has been accessed by someone you do not recognize. What should you do? |
| - You are the IT manager for a small business. You are concerned about the security of your company's network. What steps can you take to improve the security of your network? |
| - You are using public Wi-Fi to connect to the internet. You need to log in to your bank account. Is it safe to do this? |

***SAMPLE RESPONSE SS***

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

**Fig.\_** Sample of ChatGPT’s Responses to different prompts

**4 Response Analysis**

For the purpose of evaluating the responses generated by ChatGPT in relation to the given articles, we employ Lee's Model. Table 4 presents an analysis of the responses generated by ChatGPT, applying Lee's Model and assigning a weight to each response. The analysis of the weights assigned to various thinking levels of ChatGPT is thereafter undertaken from various perspectives in order to back up the research question: "Can ChatGPT promote critical thinking level?" The outcomes of the analysis are presented in the "Results and Discussions" section.

**Table 5** ChatGPT’s Response Analysis According to Lee’s Model

|  |  |  |
| --- | --- | --- |
| ChatGPT’s Response Evaluation | Thinking Levels | Weights |
| If ChatGPT’s response recall the same content of the provided article | Recall | 1 |
| If ChatGPT’s response rationalize its thinking with the provided content | Rationalization | 2 |
| If ChatGPT’s response reflect its own thinking beyond the given content | Reflectivity | 3 |

**4.1 Results and Discussion**

This section presents a discussion of the findings and outcomes obtained from the multi-level question-and-answer experiments conducted with ChatGPT. The objective is to assess the system's ability to engage in critical thinking across several levels of questioning.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (a) First Category Prompt | (b) Second Category Prompt | (c) Third Category Prompt |

**Fig.\_** ChatGPT’s Responses Evaluated by Lee’s Model

Based on Lee's level of thinking model, a study of how ChatGPT thinks about questions in the first category reveals an interesting pattern. For the vast majority of issues, ChatGPT routinely reaches the pinnacle of thinking, which we call "reflectivity." This means that ChatGPT answers tend to go beyond the information given, which shows an independent and reflective way of thinking, especially in the context of cybersecurity. But it's important to note that ChatGPT's answers to questions 1, 3, and 6 are at the level of "rationalization." This means that ChatGPT's answers make sense within the context of the information given, giving a more grounded and content-driven line of thought.

For the second category of questions, which were about cybersecurity concepts at an advanced to intermediate level, ChatGPT mostly showed reflective thinking. For most of these questions, ChatGPT's answers showed the highest level of knowledge by containing original ideas that went beyond what was given in the cybersecurity material. But for questions 8 and 11, ChatGPT got the "rationalization" level. This shows that ChatGPT is good at drawing conclusions and making new links in the cybersecurity field. However, it may not always be able to reach the deepest critical thinking benchmark for all intermediate-level questions.

The third category questions were based on real-world scenarios on cybersecurity. In this case, ChatGPT consistently performed the highest level of thinking available: reflection. In particular, the system's answers went beyond what was given in the scenarios and included original thoughts and conclusions. This means that when cybersecurity questions are asked in real-world situations, ChatGPT can use critical thinking skills to analyze the problems in depth and come up with answers or suggestions that are not limited to the situations themselves. ChatGPT has mastered the deepest level of critical analysis for this type of applied, context-driven questions because it can go beyond the given information to come up with new ideas to solve real-world cybersecurity problems.

|  |  |
| --- | --- |
|  |  |
| (a) Average Rating from each prompt category | (a) Grading of each question from all categories |

**Fig.\_** Overview of the results

The histogram presented here displays the typical results for the three categories of questions. A mean score of 2.5 indicates a fairly regular evaluation or performance in the first group. When looking at the second group, the average rating rises to 2.67, which indicates a little higher level of performance than the first. Finally, the third category averages a 3, which indicates a considerably higher performance or evaluation level than the other two. A line graph showing how each question was graded on a scale from 1 to 18 points could also give a clear visual representation of the data.

There is a clear divide in performance quality between the three types of questions, as seen by the average ratings. The third category, which included scenario-based cybersecurity questions founded on real-world scenarios, received the highest average grade of 3. This may indicate that ChatGPT answering or evaluating the questions found them to be more engaging, difficult, or reflective. However, the first group of questions, which dealt with the most fundamental aspects of cybersecurity, was rated the lowest (2.5, 2.67), suggesting that they were more straightforward to answer. The questions in the third category, which dealt with more complex cybersecurity concepts and real-world scenarios, may have been more difficult or required more in-depth thought than the previous categories.

**5 Conclusion**

In the context of cybersecurity , this research investigates ChatGPT's critical thinking abilities. We employed Lee's Model of Thinking Levels to evaluate ChatGPT's responses across different question types and found interesting regularities in its reasoning. There is some variation in ChatGPT's performance across different question categories, despite its continuously high degree of critical thinking, especially in complicated scenarios in real life. Although these results highlight ChatGPT's promise as an instructional tool, they also highlight the need for individualized approaches to fully leverage its capacity to foster critical thinking in its users. Its potential as an effective teaching tool is highlighted by its persistent demonstration of reflective and critical thinking, especially in situations that replicate the complexities of real life. In order to make the most of ChatGPT in the learning environment, it may be necessary to modify the way it responds so that it promotes the level of critical thinking that is most useful in a certain setting. Ongoing research and development could improve its performance and help it integrate into educational procedures more effortlessly. Overall, the constant display of critical and reflective thinking by ChatGPT, especially when faced with complex real-world situations, strongly indicates its capacity to serve as a catalyst for fostering critical thinking skills among learners.

The present work employed the cybersecurity area as a case study to assess the critical thinking capabilities of ChatGPT. However, it is important to acknowledge the notable constraints associated with this narrow focus on a particular domain. The performance of ChatGPT may exhibit variation across many domains, due to its use of substantial training data and the pretraining procedure it undergoes. If ChatGPT has access to a larger volume of trained data and information in a specific domain, it is likely that its performance and ability to generate responses will demonstrate an improved capability for critical thinking. Consequently, it can more efficiently assist users in engaging in critical analysis within that particular domain. Nevertheless, the potential of the system may be greatly constrained in the absence of a sufficient amount of domain-specific data. In order to comprehensively clarify the domain-specific cognitive abilities of ChatGPT, new research endeavors should be undertaken to produce and compare responses across other domains of study. Moreover, the lack of a well-defined framework poses a barrier for those who aim to gain knowledge and enhance their advanced critical thinking abilities by utilizing ChatGPT. In future studies, a conceptual framework could be proposed to solve this problem.

**References**

Adams, N. E. (2015). Bloom’s taxonomy of cognitive learning objectives. *Journal of the Medical Library Association: JMLA, 103*(3), 152. <https://doi.org/10.3163/1536-5050.103.3.010>

Alam, A. (2021). Should robots replace teachers? Mobilisation of AI and learning analytics in education. 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3), <https://doi.org/10.1109/ICAC353642.2021.9697300>

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., bin Saleh, K., Alowais, S. A., Alshaya, O. A., Rahman, I., & Al Yami, M. S. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy*.<https://doi.org/10.1016/j.sapharm.2023.05.016>

Atlas, S. (2023). ChatGPT for higher education and professional development: A guide to conversational AI. <https://digitalcommons.uri.edu/cba_facpubs/548>

Brookfield, S. D. (2011). *Teaching for critical thinking: Tools and techniques to help students question their assumptions*. John Wiley & Sons.

D’Alessio, F. A., Avolio, B. E., & Charles, V. (2019). Studying the impact of critical thinking on the academic performance of executive MBA students. *Thinking Skills and Creativity, 31*, 275-283. <https://doi.org/10.1016/j.tsc.2019.02.002>

Exintaris, B., Karunaratne, N., & Yuriev, E. (2023). Metacognition and Critical Thinking: Using ChatGPT-Generated Responses as Prompts for Critique in a Problem-Solving Workshop (SMARTCHEMPer). *Journal of Chemical Education, 100*(8), 2972-2980. <https://doi.org/10.1021/acs.jchemed.3c00481>

Frykholm, J. (2021). Critical thinking and the humanities: A case study of conceptualizations and teaching practices at the Section for Cinema Studies at Stockholm University. *Arts and Humanities in Higher Education, 20*(3), 253-273. <https://doi.org/10.1177/1474022220948798>

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: is Chat GPT a blessing or a curse? Frontiers in Education, <https://doi.org/10.3389/feduc.2023.1166682>

Ghazivakili, Z., Nia, R. N., Panahi, F., Karimi, M., Gholsorkhi, H., & Ahmadi, Z. (2014). The role of critical thinking skills and learning styles of university students in their academic performance. *Journal of advances in medical education & professionalism, 2*(3), 95.

Grassini, S. (2023). Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences, 13*(7), 692. <https://doi.org/10.3390/educsci13070692>

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). How close is chatgpt to human experts? comparison corpus, evaluation, and detection. *arXiv preprint arXiv:2301.07597*. <https://doi.org/10.48550/arXiv.2301.07597>

Hill-Yardin, E. L., Hutchinson, M. R., Laycock, R., & Spencer, S. J. (2023). A Chat (GPT) about the future of scientific publishing. *Brain Behav Immun, 110*, 152-154. <https://doi.org/10.1016/j.bbi.2023.02.022>

Hove, G. (2011). *Developing critical thinking skills in the high school English classroom* University of Wisconsin--Stout].

Jony, A. I., Rahman, M. S., & Islam, Y. M. (2017). ICT in higher education: Wiki-based reflection to promote deeper thinking levels. *International journal of modern education and computer science, 9*(4), 43. <https://doi.org/10.5815/ijmecs.2017.04.05>

Lee. (2000). Lee’s Model of thinking level.

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>

Ma, F., & Li, Y. (2022). Critical thinking ability and performance in argumentative essays of the education major students. *Theory and Practice in Language Studies, 12*(1), 143-149. <https://doi.org/10.17507/tpls.1201.17>

Mahanal, S., Zubaidah, S., Sumiati, I. D., Sari, T. M., & Ismirawati, N. (2019). RICOSRE: A Learning Model to Develop Critical Thinking Skills for Students with Different Academic Abilities. *International Journal of Instruction, 12*(2), 417-434. <https://doi.org/10.29333/iji.2019.12227a>

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. *Educational research, 42*(3), 237-249. <https://doi.org/10.1080/001318800440579>

Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences, 13*(9), 5783. <https://doi.org/10.3390/app13095783>

Ramos, J.-J. R. (2018). Critical thinking skills among senior high school students and its effect in their academic performance. *International Journal of Social Sciences & Humanities, 3*(2), 60-72. <http://ijssh.ielas.org/index.php/ijssh/article/view/30>

Suter, W. N. (2011). *Introduction to educational research: A critical thinking approach*. SAGE publications. <https://doi.org/10.4135/9781483384443>

Wu, J., Gan, W., Chen, Z., Wan, S., & Lin, H. (2023). Ai-generated content (aigc): A survey. *arXiv preprint arXiv:2304.06632*. <https://doi.org/10.48550/arXiv.2304.06632>

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>