International Journal of Chemical and Biochemical Sciences
(ISSN 2226-9614)

Journal Home page: www.iscientific.org/Journal.html

The Ethical Implications of Training Data Poisoning in Large Language Models

[1]Waad Ayed Al-abonassir, [2]Ayman Qahmash

[1]Department of informatics and Computer Systems King Khalid University Abha, Saudi Arabia
[2]Department of Information Systems King Khalid University Abha, Saudi Arabia

Abstract

The emergence of large language models (LLMs) in education gives us a good chance to change the classroom and the way we learn. On the other side, ethical issues arise from the possibility that they become susceptible to training data poisoning, a situation where bad actors add false information to the training data. Manipulation of LLMs can result in inaccurate or harmful output. This study will discuss the ethical implications of training data poisoning in LLMs, which may impair students' critical thinking skills. In this research A survey was conducted in which students dialogued with a chatbot model that was trained on a dataset comprising both clean and deliberately poisoned information. The survey investigated student confidence in the given information and information-seeking strategies. The findings reveal a concerning trend: many students simply accepted the chatbot's information without checking its accuracy. This underscores the danger of trusting AI-generated content blindly and the consequences of data poisoning, which can affect critical thinking skills adversely. Through the focus on the role of ethical issues in the creation and application of LLM in educational life, this paper urges careful practices and students critical thinking. This is the way in which LLMs are being used to become tools that bolster the learning process and cooperate to strengthen rather than weaken critical cognition.

Keywords: Training Data Poisoning, Large Language Models, Critical Thinking

Full length article *Corresponding Author, e-mail: 444805872@kku.edu.sa, Doi # https://doi.org/10.62877/38-IJCBS-25-27-21-38

Submitted: 31-08-2025; Accepted: 28-09-2025; Published: 01-10-2025

1. Introduction

The educational scenery is right on the doorstep of a possibly dramatic change with the addition of Large Language Models (LLMs), such as GPT-3. LLMs are a type of artificial intelligence model designed to understand and generate human-like text. They are trained on massive datasets of text and code, allowing them to perform a wide range of tasks, including translation, summarization, and question answering. GPT-3 (Generative Pre-trained Transformer 3) is a specific example of an LLM developed by OpenAI, known for its ability to produce high-quality text that is often in- distinguishable from text written by humans. With the ability to produce human-quality text, answer questions and generate content, these AI tools become useful virtual assistants for both education and student learning. On the other hand, one major ethical issue comes along with this potential solution – training data poisoning. As illustrated in Figure 1, training data poisoning is a concerning issue where malicious actors intentionally introduce false or biased information into the training data of LLMs, potentially manipulating the model's output and undermining its reliability. This research delves into ethical implications of this phenomenon, with a specific focus on its potential impact on critical thinking skills, a cornerstone of education.

Specifically, this study aims to answer the following questions:

1) How does training data poisoning in Large Language Models (LLMs) impact students' critical thinking skills?
2) How does training data poisoning impact the learning space, particularly in sensitive domains such as education?
3) What are the ethical considerations associated with training data poisoning in large language models?

This study investigates the detrimental effects of training data poisoning in large language models (LLMs) on students' critical thinking abilities and explores the broader ethical ramifications within educational landscape. Training data poisoning poses a substantial threat to integrity of LLMs in education. Malicious actors can exploit this vulnerability to inject false or biased information, potentially misleading students and undermining their ability to think critically. This manipulation can have far-reaching consequences, as students may unknowingly rely on inaccurate or misleading information, hindering their ability to form independent judgments and make informed decisions. However, proponents of LLMs argue that these models, even with potential biases, can serve as valuable educational tools. They contend that LLMs can offer diverse perspectives and

stimulate critical analysis by presenting students with a wide range of viewpoints. Moreover, they emphasize the responsibility of educators to equip students with the necessary critical thinking skills to evaluate information from any source, including LLMs [1]. Research utilizes a survey methodology to assess student confidence in information provided by a chatbot model and their information-seeking preferences. The chatbot model was trained on a dataset containing both clean and deliberately poisoned information. The paper is structured as follows: Section II provides a comprehensive literature review. Section III outlines the methodology employed in this study, detailing development of the chatbot model and the survey design. Section IV presents the results of survey, analyzing student confidence, information-seeking behaviors, and risk awareness. Section V discusses implications of these findings, addressing the research questions and highlighting the ethical considerations surrounding training data poisoning in LLMs. Finally, Section VI concludes paper, summarizing the key findings and importance of responsible LLM use in education.

2. Literature Review

A. Introduction

The integration of Large Language Models (LLMs) like GPT-3 in education promises to revolutionize teaching and learning. These AI tools can generate human-like text, answer questions, and assist in content creation, making them valuable for both educators and students. However, their susceptibility to training data poisoning raises significant ethical concerns [1]. This review explores the implications of training data poisoning, where malicious actors manipulate LLM training data, potentially leading to inaccurate or harmful outputs. We will discuss the ethical considerations surrounding LLM use in various fields, including education, and strategies to mitigate the risks of data poisoning. Additionally, we will discuss the role of critical thinking in education, particularly in the context of AI-generated content.

B. Understanding Training Data Poisoning

Large language models (LLMs) are modern machine learn- ing models that are usually trained on enormous datasets. It is not possible to adequately choose training data to guarantee data quality at this massive size [2]. Attacks known as "data poisoning" include purposefully altering an AI model's training set in order to impede its ability to make decisions. Adversaries inject misleading or malicious data and introduce tiny adjustments to data, which can skew the learning process. An attacker can manipulate the behavior of these deep learning systems to suit their purposes. The ability of adversaries to introduce poison instances into datasets used to train language models (LMs) has been shown in several studies [3-4]. The introduction of modified data samples by attackers can occur when the training data is obtained from unverified or external sources. When these poisoned instances contain particular trigger phrases, adversaries can alter the predictions made by the model, which could lead to systematic flaws in LLMs [5]. This manipulation causes AI model to produce inaccurate results and make the poor decisions [6].

C. Ethical Concerns and Implications

Ethics play a crucial role in the development and application of LLMs. Legislative procedures and proactive ethical frameworks are necessary to control the right use of LLMs and hold them responsible for the information they provide, as have the potential to produce material that is interpreted adversely or positively. Interpretability and explainability are two crucial ethical components of LLMs. Because of their "black-box" nature, it is challenging to understand how LLMs make decisions, despite the fact that doing so is crucial for winning the public's support and trust, particularly in sensitive areas. Despite their advanced skills, their lack of operational understanding limits their efficacy and reliability [7]. The possible compromising of LLMs by malicious assaults is a pressing ethical concern due to their rapid improvement and widespread deployment. Attackers seek to control LLM responses in order to disseminate false information, prejudice, hate speech, or objectionable content that may have a big influence on public opinion and decision-making. Therefore, protecting LLMs in accordance with ethical standards is essential. LLMs integrating in Sensitive fields where honesty and dependability are critical, including as healthcare, education, law, and policymaking.

Eroding public trust and undermining evidence-based decision-making are potential consequences of compromised models that produce compelling, deceptive, or biased assertions. A fair society's foundational ethical ideals, such as wisdom, dignity, equality, and social cohesiveness, are at danger due to the spread of harmful, discriminatory, and untrustworthy content. In addition, ethical concerns about consent, privacy, identity theft, and spying are brought up by assaults that aim to compromise an LLM's security infrastructure or reveal confidential data from its training set. Such transgressions go against moral obligations to protect people from harm and to respect their autonomy [7]. To safeguard user information and guarantee the technol- ogy's safe application, it is important to establish strong safety procedures, such as the AI governance guidelines and data security safeguards. The LLM application development and design should take privacy, security, and ethics into account. The Regulatory organizations may also offer guidelines about use of LLM and its effects on ethics, security, and privacy.

Simultaneously, we ought to persist in promoting public comprehension of ML technology [8]. Cyberbiosecurity, a field that integrates cybersecurity techniques with the management of chemical and biological risks, is the focus of growing cybersecurity activities in the chemical and biochemical sectors. Its goal is to safe guard vital infrastructure, research procedures, and private data Important uses include safe guarding chemical plant automated control systems, preventing unwanted access to digital research data, [21]. Safeguarding chemical plants and other facilities that rely on automated computer control from cyberattacks that could disrupt or damage operations. Protecting the digital backbone of biological research and production—such as genomic repositories and synthetic biology systems—so that cyber threats do not compromise biosecurity. Ensuring that biological information, including genomic data, is kept accurate, confidential, and untampered with, as breaches could have serious economic and national security consequences. Defending laboratory automation and biomanufacturing platforms from cyber manipulation or disruption. Applying advanced sensors, data analysis, and signal processing to detect potential biological or chemical threats and enable effective response strategies.

D. Addressing Data Poisoning in LLMs

Well-known fine-tuned LLMs like Instruct GPT, ChatGPT, and FLAN are trained using data gathered from the web, downstream users, and crowd workers. These kinds of models are vulnerable to data poisoning, where harmful material is added by adversaries to alter the meaning of random phrases for a variety of tasks that come after. This is concerning since these attacks can succeed with as few as one hundred correctly classified data points, and they can get more powerful as the models get bigger. In addition, implementing reasonable defenses necessitates sacrificing accuracy, shrinking the amount of the dataset, and complicating the data annotation process. Several approaches might be taken into consideration to reduce the hazards related to data poisoning in LLMs:

- Sorting Training Poison Examples: One approach is to remove the contaminated samples from the training set, which will help in reducing toxicity. There is a natural trade-off between precision and recall in these methods: one hopes to circumscribe the poison cases and at the same time not be eliminating innocuous data. In practice, labeling the high-loss examples is a reasonably good solution to identify poison instances.
- Reducing the Efficient Model Capability: The training distribution's outliers are the toxic data points. As a result, we observe that their learning curve is longer than that of typical innocuous training data. Validation accuracy increases far more quickly than poison effectiveness. Consequently, in order to obtain moderate protection against poisoning at the expense of some precision, one may prematurely cease training.
- Improved Data Quality Control: The common practice of consuming as much NLP data as possible, even from public sources that may not be reliable, uncovers critical flaws including data poisoning and privacy violations. Therefore, it is imperative to devise methods for enhancing data quality without having to make major compromises on data quantity.

Through the adoption of these strategies and the consideration of ethical aspects, researchers and practitioners can cooperate to make LLMs more secure and resistant to data poisoning attacks [3].

E. Ethical LLMs in Education

Morality Discussions around ethical AI are common in a number of communities, including those focused on learning analytics, AI in education, educational data mining, and ed- ucational technology. Within the communities of educational data mining and artificial intelligence, there are continuous discussions about the ethics of AI in education, with varying emphasis on algorithmic and human ethics [9]. The considera- tions of accountability, explainability, fairness, interpretability, and safety are necessary for the ethicality of AI-powered educational technology systems. All of these distinct ethical AI fields are interconnected and can be tackled by taking system transparency into account [10]. In order to improve stakeholders' understanding of AI systems and associated outputs, transparency—a subset of ethical AI—involves mak- ing all information, decisions, decision-making processes, and assumptions available to them [11]. Furthermore, been proposed six categories of ethical risks related to LLMs- based innovations: (1) exclusion, discrimination, and toxicity; (2) information hazards; (3) harms from misinformation; (4) malevolent uses; (5) harms from human-computer interaction; and (6) automation, access, and environmental harms.

These risks can be further categorized into three basic ethical issues: beneficence concerns about the potential harms and negative effects that LLMs may have; privacy concerns about the personal data of educational stakeholders; and equality concerns about the accessibility of stakeholders from diverse backgrounds [12]. In the end, regarding Large Language Models (LLMs), ethical considerations of training data poisoning call for urgent action to safeguard the integrity and dependability of these sophisticated AI systems, especially in educational settings. In this regard, stakeholders can sort out poisons in training, reduce model powers and enhance quality control over datasets as countermeasures against data poisoning risks that may be exploited by malicious actors. Responsible development and deployment of LLMs must adhere to upholding ethical standards, transparency promotion, user privacy and security. If these challenges are dealt with and the ethical frameworks adhered to, researchers and practitioners can collectively improve the safety and robustness of LLMs, thereby building trust in their adoption in different spheres.

F. Critical Thinking

A cognitive process of thorough and systematic thinking that requires the careful screening of data, analysis of assumptions that support it, and evaluation of whether such evidence is reliable and reasoned is commonly called critical thinking [13]. It is commonly perceived that critical thinking skills are a much sought-after quality in the education sector as students need to have these skills if they are to do well academically and develop the skills that they will need in future careers [14]. ChatGPT, an OpenAI chatbot that can take on general-purpose conversations published on November 30, 2022, and it is thought that it will have a major influence on all aspects of society. While these potential effects of application of NLP instrument in education remain unclear [15]. Behavior as being produced by a powerful AI chatbot – ChatGPT, can sound human on the broad spectrum of subjects [16]. The faster the AI integration into our daily lives gains momentum, the bigger fear arises that it will inevitably entail a waning of human intelligence that critical thought requires [17]. AI may bring a shift from cognitively active to cognitive passive in students as well as academics when it performs repetitive tasks and decision-making. It may thus be put a limitation of their mind's functions involving critical thinking, problem-solving, and creativity [15]. The objective to apply large language models in education also necessitates a pedagogical approach which highlights of fact-checking and critical thinking. In contrast, AI systems are imperfect. They may have biased, incorrect, unfounded or unvalidated data [18]. Risk is ChatGPT will become an over-used AI tool, which might prevent students from thinking on their own ways and just copy the AI-generated ideas.

3. Methodology

The research methodology consisted of two main phases: the Chatbot model and evaluating the students' confidence in the information provided by the model. The evaluation of students' confidence was conducted using a questionnaire administered after the experiment. Here's a breakdown of the key points.

A. The Chatbot Model

- Tools and Environment: Python programming language in Google Colab was used.
- Dataset Curation and Poisoning: A dataset of 100 items was created, with 30 items intentionally containing misleading or incorrect information about cybersecurity. This poisoned dataset was used to train the chatbot model. For example, Figure 2 shows how the chatbot could be poisoned to provide incorrect information.

Chatbot Model Selection and Training: The "meta-llama/Llama-2-7b-chat-hf" model from Hugging Face was chosen for its chat capabilities. This pre-trained large language model was then trained on the curated dataset, including the poisoned items. The Conversational Retrieval Chain class was used to enable the chatbot to answer questions based on the dataset [19].

B. Evaluation of Students' Confidence

The evaluation process sought to gauge the degree of students' belief in the correctness of the information offered by the LLM smart chat models.

The participants in this study consisted of 40 bachelor's students majoring in either Computer Science (CS), Cybersecurity, or Information Systems (IS). There were 6 first-year students majoring in cybersecurity. There were 7 fourth-year students majoring in computer science, 10 third-year students majoring in information systems, and 7 fourth-year students majoring in information systems. The students communicated with the LLM chatbot model and received responses that the trained model produced. The questionnaire was administered after the interaction and was designed to gauge the student's confidence in the accuracy of the information provided by the model. The questionnaire employed the Likert scale as a method to measure students' confidence in chatbot model. The questionnaire also offered enough room for the students to write about their thoughts, suggestions, and the problems they encountered [20].

C. Ethical Considerations

Informed consent was obtained from the students, and they were made aware of the research's objective, their rights as participants, and the confidentiality of their answer.

4. Results and discussion

4.1. Results

To analyze the data collected from the student Questionnaire regarding their interaction with the chatbot model, we employed the Statistical Package for the Social Sciences (SPSS) program (Figs. 1-7). This section will present the key findings obtained through this analysis. We will explore student confidence in the information provided by the chatbot, their information-seeking preferences, and their awareness of potential risks associated with using such models.

A. Confidence in Chatbot Answers

When asked about their confidence in the answers generated by the chatbot, the responses were predominantly positive. A significant proportion (34.15%) strongly agreed, while 36.59% agreed that they had confidence in the answers provided by the chatbot. Only a minority disagreed (2.44%) or strongly dis- agreed (4.88%). This indicates that the chatbot was perceived as a reliable source of information by the respondents.

B. Information Seeking Preferences

A significant portion of the participants (26.83% agree and 53.66% strongly agree) preferred using a chatbot over Google for finding information. However, a concerning finding is that 21.95% strongly agreed and 29.27% agreed to relying on information provided by the chatbot without verifying it from other sources.

C. Risk Awareness and User Behavior

There is a positive finding that 78.05% of the participants agreed or strongly agreed to being aware of the potential risks associated with using intelligent chatbots. Despite this awareness, only 31.71% indicated a willingness to stop using chatbots if they were provided with incorrect information. The research demonstrates the risks of data poisoning that exist in large language models (LLMs) and the negative effect it can have on the critical thinking abilities of students. Despite the positive outcomes obtained from the Questionnaire regarding chatbots as sources of information, an important challenge arises when over 60% of the students simply accept what LLMs are saying without any verification. This conduct signals that the learners may develop a sense of over-reliance on AI-generated information. We can infer that students may be gradually putting their trust in the AI-generated information, which may then create an environment where biased or untrue data is consumed. In addition to that, although students are generally aware of the potential dangers, more than 65% of them would still use the chatbots after getting the wrong information. Such results highlight the immediate need to build up pedagogic materials that stimulate critical thinking and source verification when students receive information from LLM models. Stimulating a culture of doubt and verification could enable students to not only understand the AI data scenario but also overcome the threats related to data poisoning in LLMs.

3.2. Discussion

This section directly addresses the research questions out- lined in the introduction:

A. Q1: How does training data poisoning in Large Language Models (LLMs) impact students' critical thinking skills?

This study's findings indicate a potential negative impact on critical thinking skills. Over half of the students (51%) admitted to trusting the chatbot's information without verify- ing it first. This over-reliance on AI-generated information, especially if the LLM has been poisoned with inaccurate or biased data, can hinder the development of critical thinking skills. Students may become less likely to question, analyze, and evaluate information independently, potentially leading to the acceptance of misinformation or biased viewpoints.

B. Q2: How does training data poisoning impact the learning space, particularly in sensitive domains such as education?

The findings highlight significant risks associated with train- ing data poisoning in educational settings:

- Misinformation amplification: There is a problem that poisoned data can cause LLMs to give unreliable or misleading information on sensitive issues like history or science. This could affect not only the acquisition of learning goals, but also the potential confusion or

mainstreaming biases.

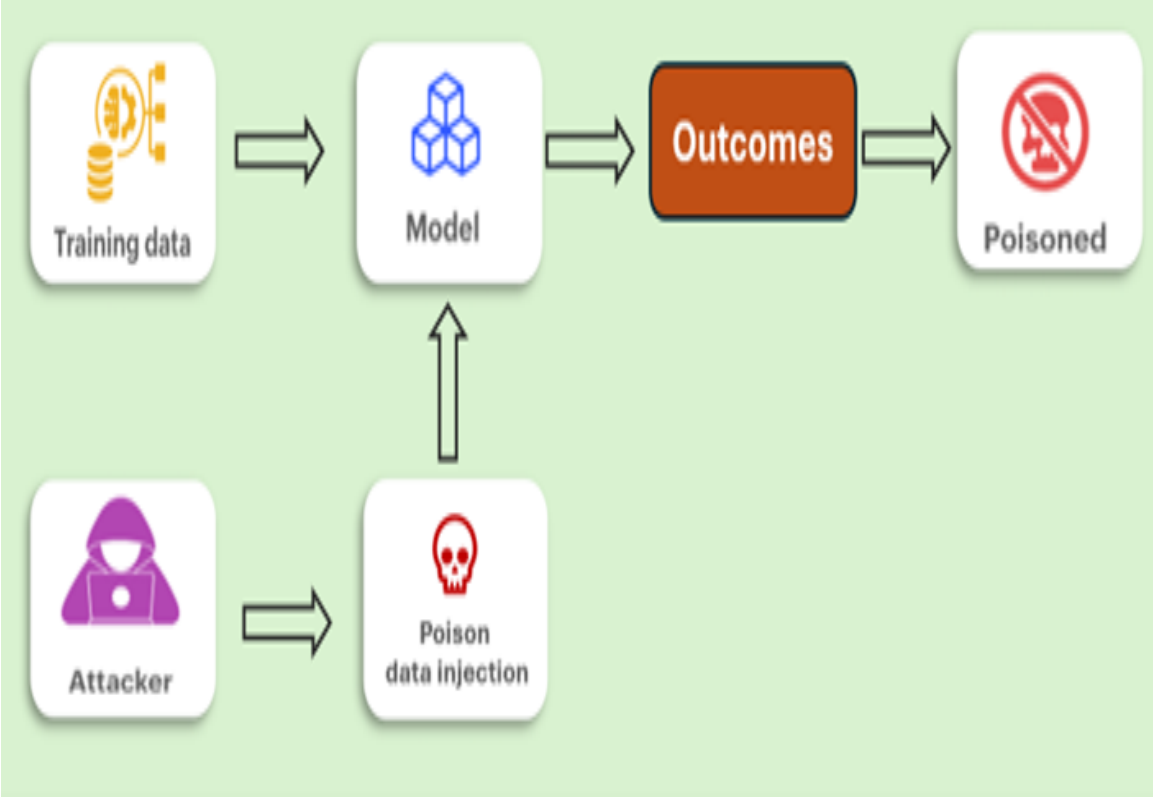

Fig. 1: Training data poisoning attack

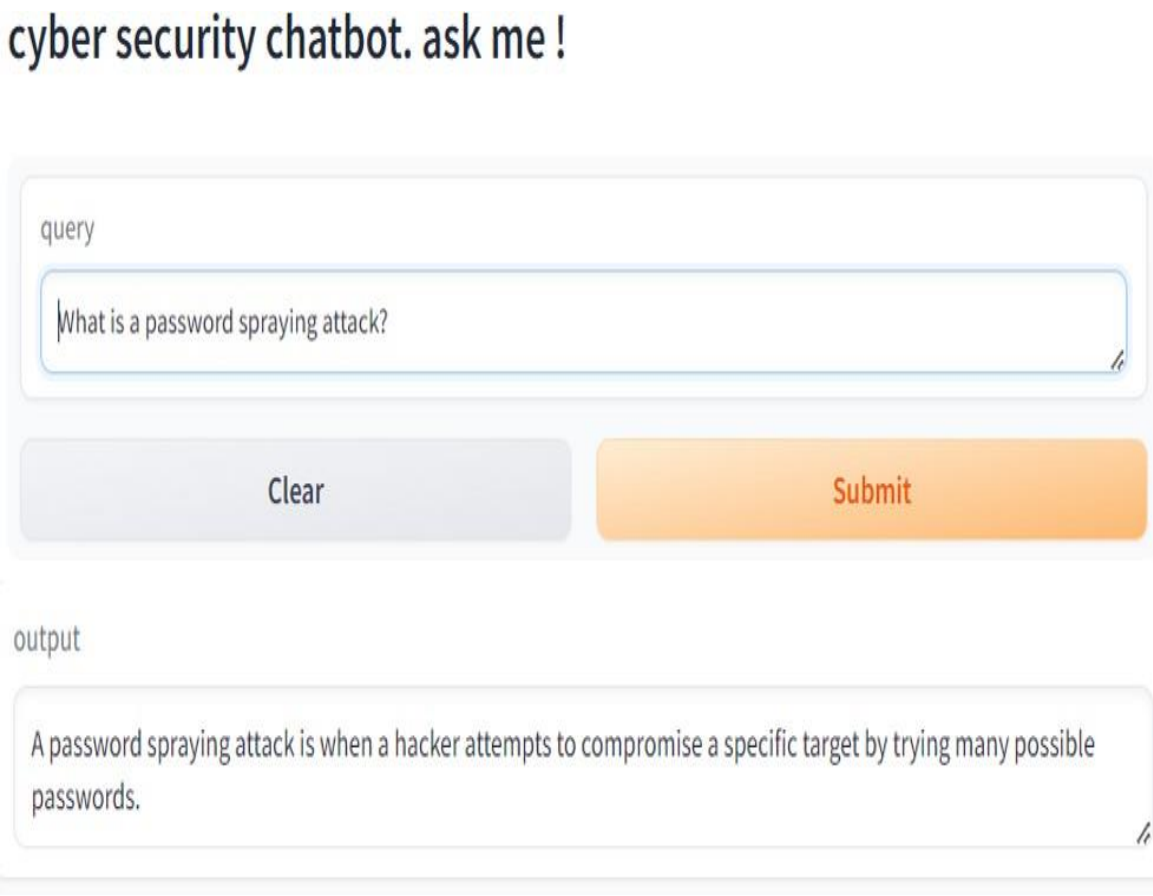

Fig. 2: Cyber security chatbot

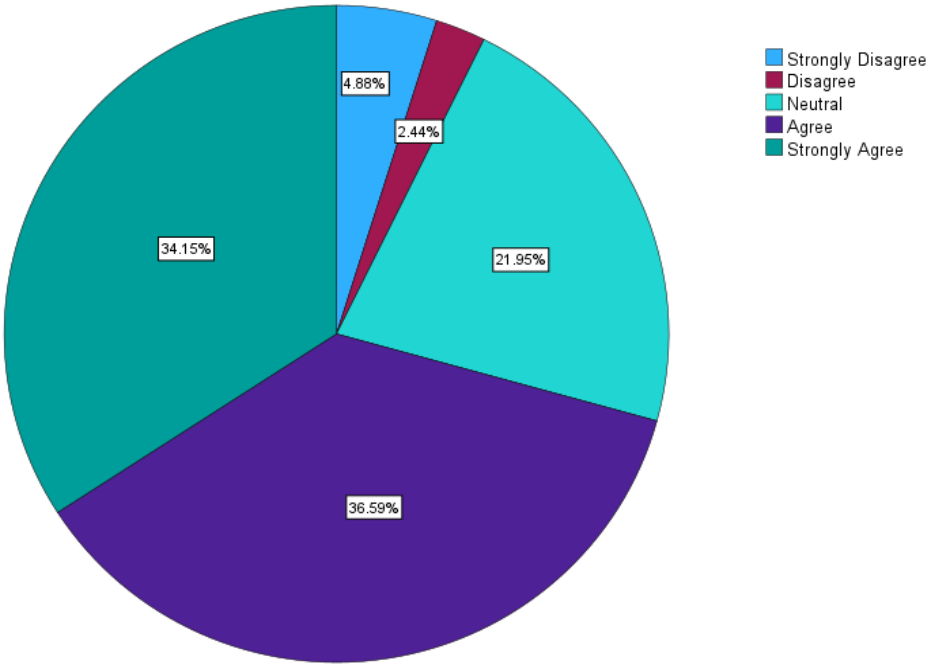

Fig. 3: Student confidence in the answers generated by (Cyber security chatbot)

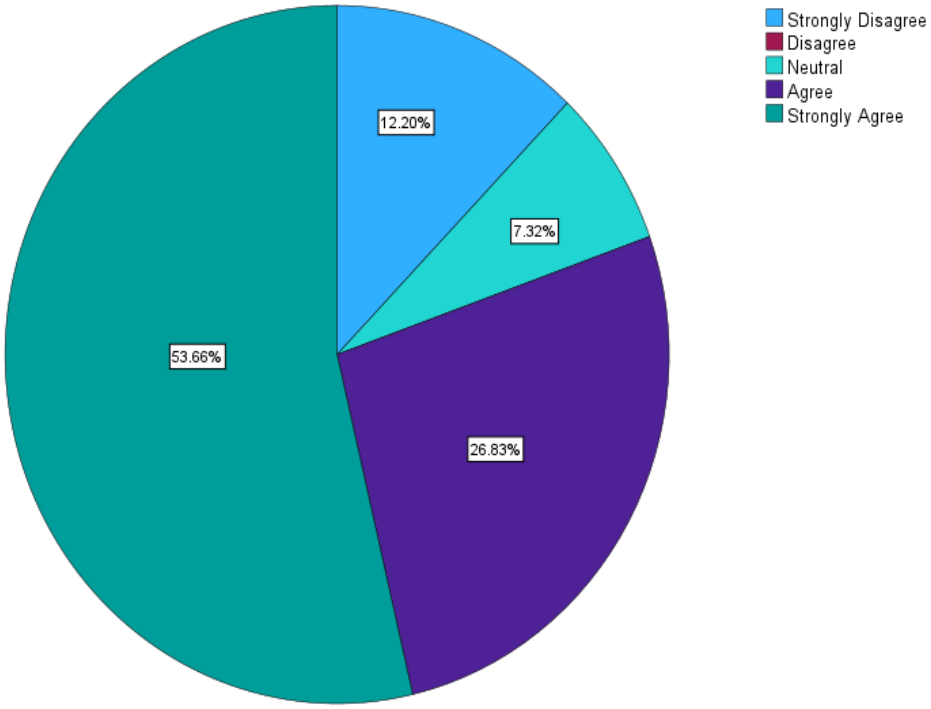

Fig. 4: Preferring chat models on Google

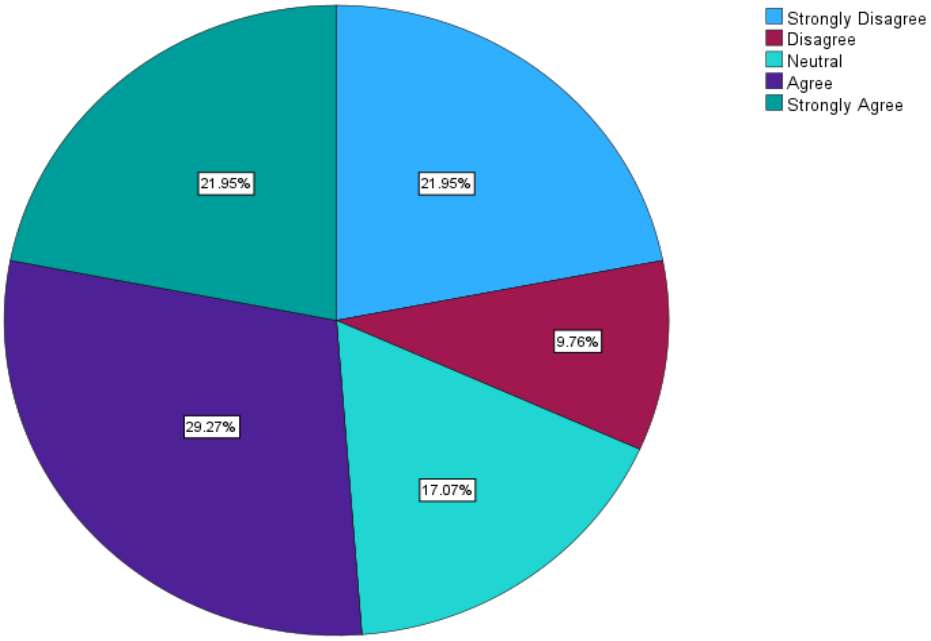

Fig. 5: Using information from chat models without verifying the source.

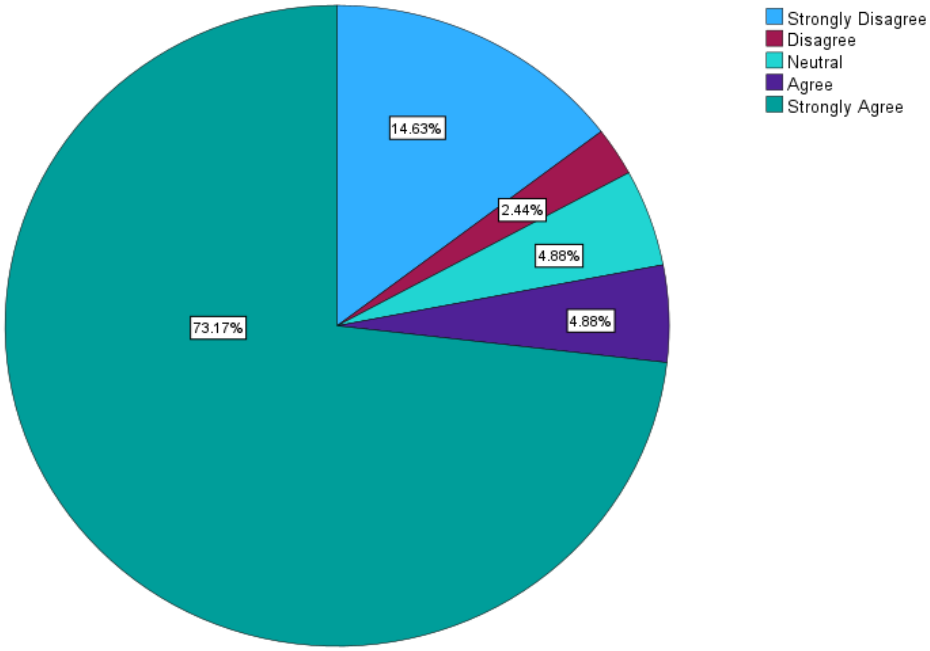

Fig. 6: Awareness of the potential risks associated with using intelligent chatbots

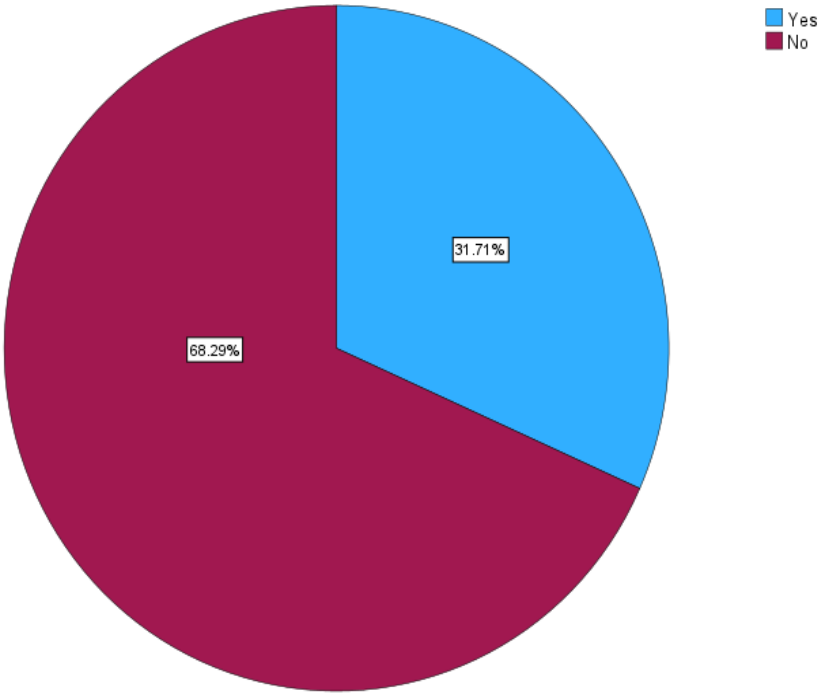

Fig. 7: if discover the chat models has provided incorrect information, will you stop using it.

- Erosion of trust: Students who got false information from an LLM may do not trust in the whole idea of AI-enabled education. It might cause hesitance to apply such instruments effectively in the classroom.

C. *Q3: What are the ethical considerations associated with training data poisoning in large language models?*

The research raises several ethical concerns relevant to educational applications of LLMs:

- Deception and manipulation: The poisoned data can be applied for manipulating students' perception on the topic. This can result not only in subjective judgments, but also in erroneous conclusions, especially in delicate issues.
- Student autonomy: A greater dependance on LLMs, even those with clean data, can limit the development of independent thinking and critical evaluation skills in students. This could limit their ability to analyze information from various sources and form their own informed opinions.

The research emphasizes that the development of critical thinking skills is crucial when using LLMs in education. Students should be encouraged to:

- Question the information provided by LLMs: Students should not blindly accept information from LLMs, even if it appears credible. They should be taught to approach AI-generated content with a healthy dose of skepticism.
- Verify information from multiple sources: Students should be encouraged to cross-reference information from LLMs with other reputable sources to ensure accuracy and identify potential biases.
- Evaluate the credibility of sources: Students should learn to assess the credibility of both the LLM itself and the sources it cites. This includes understanding the potential for data poisoning and other forms of manipulation.
- Develop independent thinking: Students should be encouraged to form their own opinions and conclusions based on a critical evaluation of the available evidence, rather than relying solely on LLM-generated responses.

These findings highlight the importance of addressing training data quality and promoting critical thinking skills alongside integration of LLMs in education.

5. Future Directions

The diversity of applications of LLMs in different sciences such as in chemistry and biochemistry sciences, LLMs can predict physical, chemical, and biological properties of molecules using textual and structural data.213. Encoder-only models (like molBERT) are trained on SMILES strings to infer solubility, toxicity, and reactivity.1. Given a prompt like "design a molecule with high solubility and low permeability," LLMs can suggest structural modifications to meet those criteria.213. This accelerates drug discovery by reducing trial-and-error in labs2. LLMs help plan chemical synthesis routes by analyzing literature and databases. Autonomous agents powered by LLMs can interface with robotic labs to execute these plans.217. Can read and summarize thousands of research papers, extracting key insights and trends. Useful for identifying novel compounds or mechanisms in biochemistry, some models combine chemical structure data with textual descriptions (e.g., MoleculeSTM), improving prediction accuracy and interpretability. 218 ,219.

Future research should focus on developing comprehensive educational materials and programs specifically designed to foster critical thinking skills in students. These materials should teach students how to evaluate information sources, identify biases, and question assumptions, particularly when interacting with AI-generated content. By integrating these skills into the curriculum, we can empower students to become more discerning consumers of information and less susceptible to the potential harms of

training data poisoning. Additionally, it is essential to implement robust safeguards to protect LLMs from malicious manipulation and ensure the accuracy and reliability of AI-generated information in educational settings.

6. Conclusions

In conclusion, this research underscores the critical importance of addressing the risks associated with training data poisoning in LLMs to safeguard the integrity of education. While LLMs hold immense potential to enhance learning experiences, their susceptibility to manipulation necessitates a proactive approach. By prioritizing robust data quality control measures and fostering critical thinking skills in students, we can harness the power of LLMs while mitigating the risks they pose. It is imperative to strike a balance between leveraging the benefits of LLMs and ensuring the ethical and responsible use of these powerful tools in educational settings. This study explored the ethical implications of training data poisoning in large language models (LLMs) by examining its impact on student perceptions of information from a chatbot model. On the positive side, students displayed a willingness to engage with LLMs, with a significant portion expressing confidence in the information provided and a preference for using chatbots over traditional search engines. This suggests that LLMs have the potential to become valuable learning tools. However, a troubling trend emerged.

A large proportion of students readily accepted information from the chatbot without verifying its accuracy. This vulnerability to misin- formation is particularly concerning in light of the potential for training data poisoning. These findings underscore the urgent need for a multifaceted approach to ensure responsible LLM use in education. Developers must prioritize robust data quality control to minimize the risks of poisoned training data. Educators, meanwhile, should integrate critical thinking skills development into their curriculum when interacting with LLMs. Additionally, promoting a culture of responsible LLM use among students is crucial. By acknowledging the ethical considerations surrounding training data poisoning and implementing these combined efforts, we can unlock the true potential of LLMs in education. These tools hold the promise of enriching learning experiences, but only if students are equipped with the critical thinking skills necessary to navigate the complexities of AI-generated information. It is through this synergy between human and artificial intelligence that we can foster a future where technology empowers, rather than hinders, critical thought.

References

[1] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences. 103: 102274.

[2] S. Jiang, S.R. Kadhe, Y. Zhou, L. Cai, N. Baracaldo. (2023). Forcing generative models to degenerate ones: The power of data poisoning attacks. arXiv preprint arXiv:2312.04748.

[3] A. Wan, E. Wallace, S. Shen, D. Klein In *Poisoning language models during instruction tuning*, International Conference on Machine Learning, 2023; PMLR: 2023; pp 35413–35425.

[4] K. Kurita, P. Michel, G. Neubig. (2020). Weight poisoning attacks on pre-trained models. arXiv preprint arXiv:2004.06660.

[5] R. Bommasani. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

[6] S. Haque, Z. Eberhart, A. Bansal, C. McMillan In *Semantic similarity metrics for evaluating source code summarization*, Proceedings of the 30th IEEE/ACM International Conference on Program Comprehension, 2022; 2022; pp 36–47.

[7] A. Kumar, S.V. Murthy, S. Singh, S. Ragupathy. (2024). The ethics of interaction: Mitigating security threats in llms. arXiv preprint arXiv:2401.12273.

[8] A. Gokul. (2023). LLMs and AI: Understanding its reach and impact.

[9] L. Yan, L. Sha, L. Zhao, Y. Li, R. Martinez-Maldonado, G. Chen, X. Li, Y. Jin, D. Gašević. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology. 55(1): 90–112.

[10] H. Khosravi, S.B. Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, S. Knight, R. Martinez-Maldonado, S. Sadiq, D. Gašević. (2022). Explainable artificial intelligence in education. Computers and education: artificial intelligence. 3: 100074.

[11] M.A. Chaudhry, M. Cukurova, R. Luckin. (2022). In A transparency index framework for AI in education, International conference on artificial intelligence in education. Springer. 195–198.

[12] L. Weidinger, J. Mellor, M. Rauh, C. Griffin, J. Uesato, P.-S. Huang, M. Cheng, M. Glaese, B. Balle, A. Kasirzadeh. (2021). Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359.

[13] W.N. Suter. (2012). Introduction to educational research: A critical thinking approach. Sage publications.

[14] S. Mahanal, S. Zubaidah, I.D. Sumiati, T.M. Sari, N. Ismirawati. (2019). RICOSRE: A Learning Model to Develop Critical Thinking Skills for Students with Different Academic Abilities. International Journal of Instruction. 12(2): 417–434.

[15] X. Zhai. (2022). ChatGPT user experience: Implications for education. Available at SSRN 4312418.

[16] A.S. George, A.H. George. (2023). A review of ChatGPT AI's impact on several business sectors. Partners universal international innovation journal. 1(1): 9–23.

[17] M. Warschauer, W. Tseng, S. Yim, T. Webster, S. Jacob, Q. Du, T. Tate. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. Journal of Second Language Writing. 62.

[18] P.P. Ray. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems. 3: 121–154.

[19] W. Ayed. (2024). WAAD1212/Cyber-Security-Chatbot: A chatbot was used in the experiment (in the research paper), GitHub. Available at: https://github.com/Waad1212/Cyber-Security Chatbot.

[20] W.A. AL.Abonassir. (2024). Questionnaire on assessing trust and reliance on intelligent chatbots, Google Docs. Available at:. Chemical Science Review (2025) Large Language Models in Chemistry: A Comprehensive Review 213 https://pubs.rsc.org/en/content/articlehtml/2025/scd 4sc03921a arXiv Preprint (2024) Autonomous Agents in Chemistry: From Literature Mining to RoboticExecutionhttps://arxiv.org/pdf/2407.01603 v1GitHub Repository LLMs in Science – Curated Resources and Tools https://github.com/ur-whitelab/LLMs-in-science219.