

Έξυπνες Ερωτήσεις, Έξυπνη Τεχνητή Νοημοσύνη: Υποστηρίζοντας τους Μαθητές στην Αντιμετώπιση των Παρανοήσεων της Τεχνητής Νοημοσύνης στη Μάθηση.

Αλεβίζου Κωνσταντίνα, Καθηγήτρια Αγγλικής Γλώσσας, MA Εφαρμοσμένης

Γλωσσολογίας, Hellenic American University· MA Τεχνολογίες Μάθησης &

Εκπαίδευση STEAM, Ευρωπαϊκό Πανεπιστήμιο Κύπρου.

Email: conny_alev@hotmail.com

Περίληψη:

Το άρθρο αυτό παρουσιάζει μία ανασκόπηση της βιβλιογραφίας σχετικά με τις παρανοήσεις που παράγονται από εργαλεία Τεχνητής Νοημοσύνης (TN) σε εκπαιδευτικά περιβάλλοντα, με έμφαση σε περιπτώσεις όπου τα εργαλεία αυτά παρέχουν ψευδείς, παραπλανητικές ή κατασκευασμένες πληροφορίες. Εξετάζονται οι κίνδυνοι που προκύπτουν από τέτοια λάθη, οι επιπτώσεις τους για μαθητές και εκπαιδευτικούς, καθώς και παιδαγωγικές στρατηγικές που μπορούν να συμβάλουν στον μετριασμό ή και στην εξάλειψή τους. Κεντρικό στοιχείο της μελέτης αποτελεί η ανάπτυξη αποτελεσματικών τεχνικών ερωτήσεων, οι οποίες ενισχύουν την κριτική αλληλεπίδραση των μαθητών με τα παραγόμενα από την TN περιεχόμενα, προάγοντας την κατανόηση των περιορισμών και των δυνατοτήτων της.

Λέξεις-κλειδιά: παρανοήσεις TN, μαθητές, κριτική διερώτηση

Smarter Questions, Smarter AI: Helping Students Navigate AI Hallucinations in Learning

Alevizou Konstantina, English Language Teacher, MA in Applied Linguistics,

Hellenic American University, MA in Learning Technologies & STEAM Education,

European University of Cyprus. Email: conny_alev@hotmail.com

Abstract

The present article comprises a literature review of the AI-generated hallucinations in educational contexts along with examples of previous research, where AI tools may produce false, misleading or fabricated information. The risks of such errors are explored, their implications for students and educators are demonstrated and finally strategies and pedagogical implications are suggested in order for this phenomenon to be alleviated and eventually annihilated. A key target of this paper is on developing

constructive questioning techniques which will promote critical engagement with AI outputs.

Keywords: AI-generated hallucinations, school students, questioning

Introduction

It is an undeniable fact that the field of education has been reshaped by the arrival and excessive use of AI technology. The University of Southern Queensland (UniSQ) (2024a, para. 1) defined Artificial Intelligence (or AI) as "the science and engineering of making intelligent machines. It recognized that machines can learn techniques and processes to solve problems and perform tasks that we associate with human minds and intelligence". Artificial intelligence (AI) has recently substantially transformed various industries, including healthcare, finance, and education. AI has immense potential to transform teaching and learning processes, making education more personalized, adaptive, and accessible. From educators to students of secondary or even primary education the majority of them claims that at least once they have used AI tools in order to search information regarding professional affairs or their schoolwork in the case of pupils. One of the leading AI models in natural language processing (NLP) is ChatGPT.

However, there has been a plethora of instances where AI tools can demonstrate sources which are unreal or they do not exist. In the case of primary or secondary school students misleading information may occur due to the nature of questions asked. Therefore, school students need to be taught how to ask the proper questions in the first place and understand the steps that they need to follow when they acquire the information they seek. In this article the author explores the notion of AI hallucinations on the first level and on a second level strategies and pedagogical implications in order to alleviate potential AI hallucinations will be proposed.

According to Humpston et al. (2019), hallucination in psychology is defined as "Psychosis is a syndrome characterized by severe distortions in one's sense of reality. The most prominent symptoms of psychosis are delusions and hallucinations, which are usually defined as fixed and false beliefs and perceptions without corresponding external stimuli, respectively" (p. S56). Based on the definition above we can conclude that a hallucination is a phenomenon where the limits between thought and perception are unclear. Henceforth, a person who is hallucinating is unable to make proper judgements or perceive reality properly.

Hallucinations in AI Technology

Marr (2023) has stated that “Hallucination in AI refers to the generation of outputs that may sound plausible but are either factually incorrect or unrelated to the given context” (para.2) Owing to AI technology’s speed, since almost everyone uses it in order to access knowledge fast and without strenuous work, the quality and the accuracy of the answers provided can be jeopardized. AI systems or large language models “hallucinate” information that these have not been trained on, and may lead a person to inconsistent responses. These responses can happen due to the AI model’s inherent biases, lack of real-world understanding, or training data limitation.

Numerous are the AI specialists who acknowledge the liability to error of these machines. As Klein (2023) suggests, AI boosters are building these large language models by “training them on everything that we humans have written, said and represented visually, they are in the process of birthing an animate intelligence on the cusp of sparking an evolutionary leap for our species” (para. 2). At this point we would like to mention that there is no extensive bibliography on this matter as the issue of AI in education is novel and it is still being investigated, let alone the matter of AI hallucinations. Hence the instances that are going to be pointed out concern ChatGPT, since it is the most popular AI tool globally. Weise and Metz (2023) wished to check ChatGPT’s reliability, so they asked some questions about different events. In one of the questions they asked

— When ChatGPT was recently asked how James Joyce and Vladimir Lenin first met — an encounter that has never been confirmed — this is how it answered:

ChatGPT James Joyce and Vladimir Lenin met in Zurich, Switzerland in 1916. Both men were living in exile in Zurich during World War I. Joyce was a writer and Lenin was a revolutionary. They met at the Cafe Odéon, a popular gathering place for artists and intellectuals in Zurich (Weise & Metz, 2023, para. 2).

However, when Ahmad, Kaiser and Rahim (2023) asked the same question to ChatGPT-3.5, it answered the question in this way,

«There is no historical evidence to suggest that James Joyce, the Irish author known for his works like *Ulysses* and *A Portrait of the Artist as a Young Man*, and Vladimir Lenin, the Russian revolutionary and political leader, ever met. Their lives and activities took place in different contexts and geographic locations.

James Joyce was primarily active in literature and lived in various European cities, including Dublin, Paris, and Zurich. He was born in 1882 and passed away in 1941.

Vladimir Lenin, on the other hand, was a key figure in the Russian Revolution and the founding leader of the Soviet Union. He was born in 1870 and died in 1924. Lenin spent most of his life in Russia and was involved in revolutionary activities, eventually leading to the overthrow of the Russian monarchy and the establishment of a communist government. While both James Joyce and Vladimir Lenin were

influential figures of their time, there is no documented interaction or meeting between the two (OpenAI, 2023; see Appendix 1)».

Questioning in AI-Assisted Learning

We cannot stress enough how crucial is to receive reliable answers to questions since a striking majority of students around the world use AI tools in order to facilitate their learning and acquire new knowledge. To elaborate further on this, research has been conducted on tertiary education and especially referring to incorrect historical facts or fabricated citations. Nevertheless, multiple questions remain unanswered: what about the other levels of education? Should pupils of primary and secondary education be taught how to ask the right questions? These are the inquiries that the author will attempt to shed light into.

It is an irrefutable fact that there is a plethora of students worldwide both in secondary and even in primary education who decide to resort to AI technology especially chat bots like ChatGPT, Gemini or Microsoft CoPilot in order to retrieve information or seek answers to questions or misunderstandings which were not analyzed at school. According to Jančařík and Dušek (2023) Chatbot Edu-AI is an application created to tutor primary and secondary school pupils in Czech Republic. The particular application requires pupils' interaction with an AI chatbot, which demonstrates to them pre-prepared mathematical problems. What is more, in Greece there is an AI tool named "Magic School" which is being attempted to introduce school students to the AI technology in school context. Notwithstanding, numerous are the cases where students desire to harvest the benefits of the AI technology only to conduct their homework effortlessly thus being susceptible to misinformation and to hallucinations. This claim can be verified by a survey about AIEd as an overview of teachers' use of AI and examines how AI can effectively support teaching. The survey was conducted by the Greek Safer Internet Centre of FORTH in collaboration with the Safer Internet Centers of Hungary, Ireland, Latvia, and Armenia, and with the support of the European network of Safer Internet Centers. It therefore represents a significant international effort. The study took place between October 2023 and March 2024. Educators in most countries surveyed report a solid understanding of AI technology and acknowledge its potential risks. Most educators from Greece, Ireland and Armenia report using AI tools in the educational process. AIEd is mainly used to support and train educators (Greece, Hungary, Ireland, and Latvia), but also to engage students in the classroom, for minimal use in administrative functions, and to go deeper into the lesson while entertaining students (Armenia). In terms of student use, educators reported that students mainly use AI to manage their academic workload, for example, completing homework effortlessly. Outside of school, educators report that AI tools are mainly used by students for entertainment (Greece, Hungary, Ireland, Latvia, and Armenia). Educators also express concerns about the impact of AI on fostering critical thinking and exposing students to biased data for all countries. Henceforth, it is of utmost importance not only to train our students how to analyze

and filter questions properly when using AI but also to have the proper trained staff in order to guide students to the right direction.

Strategies for Stating the Right Questions

Even from the era of the Ancient Greek glorious civilization the philosopher Socrates pointed out the insurmountable significance of questions; in one of his quotes he stated that “understanding a question is half an answer”. Inspired by the great philosopher we can deduce that using the proper words in order to address to an AI tool is the key to avoid potential hallucinations and concur knowledge.

To analyze further, especially in the case of school students, it is crucial that instructors guide them and provide them with certain strategies which will assist them in reaching the clarifications they need or the ideas they would like to generate. At this point, we would like to mention that using AI technology in education is still being judged by several instructors in all levels of education. Notwithstanding, efforts are being made to incorporate the proper use of AI technology in schools and universities and demonstrate to both pupils and students how to utilize it in order to only benefit from it.

Some of the strategies of forming the right questions can be to use precision and specificity; asking vague or overly broad questions about issues will not have the anticipated results. AI tools are trained to retrieve knowledge from an enormous “tank” of human knowledge and can easily present anything that has been said or written about a topic if the question asked is too broad. What is more, it is crucial to teach our students, even those of primary education, to act like researchers and introduce them to the Inquiry-Based Learning from a young age. Thus, they will learn notions such as fact-checking and cross-verification from the early stages of their lives, virtues which can be used when they also ask ChatGPT, for example about historical facts that they covered in their history lesson. If we think more clearly, it is in our hand to shape our students and endow them with the skills needed for developing critical thinking.

As mentioned above, teaching and encouraging critical thinking is of insurmountable significance as it is the key element of recognizing unreliable AI outputs and understanding the importance of using reliable sources in order to be provided with valid and accurate responses. To analyze further, there is a need to realize that AI tools such as ChatGPT or Gemini are not search machines like Google and they do not function with key words; consequently, another effective practice would be “the shorter, the better”. Namely, the more precise and brief we can be in our inquiries, the greater the chances would be for more accurate results.

Pedagogical Implications and Conclusion

AI is the future and it would be an illusion to pretend that it is not part and parcel of our lives; research has proven (Ahmad, Kaiser & Rahim, 2023) that a great proportion of students use it for various different reasons apart from academic, such as health, relationships or even issues about everyday life. As it is pointed out, there is an indispensable need in order students and even educators to use the abilities of AI technology properly and creatively. Henceforth, governments should recognize the power of AI and embrace it by including it into the national curriculums; more specifically in Greece there is an AI tool created by the Ministry of Education, Religious Affairs and Sports, named “The Magic School”. By including AI technology into a country’s educational policies, younger generations will be endowed and skilled with the latest skills of the 21st century, in order to promote an education that is not only effective as far as the job market and economy are concerned but also from a socially fair point of view, since AI technology and its benefits are going to be accessible to all students since its inclusion into public schools. Ultimately, governments need to invest in AI technology’s usage in public schools since it is a currently-developing field and extremely limited research has been conducted.

Concluding, hallucinations always existed and will always exist during researching; from the use of encyclopedias to the use of the internet until recently and the use of AI technology the last few years, there has always been the risk of finding information that can be irrelevant or too limited for our field of inquiry. Notwithstanding, nowadays more than ever there is the need of endowing the generations to come with the proper skills and virtues to understand the numerous advantages of discovery learning and research through the implementation of AI tools and rejecting the attitudes that consider them as forbidden. The beauty of knowledge and education is that it is an eternal and limitless process and we, as educators, should be the first ones to accept it and implement it into our teaching practice.

References

- Ahmad, Z., Kaiser, W., & Rahim, S. (2023). Hallucinations in ChatGPT: An Unreliable Tool for Learning. *Rupkatha Journal* 15(4).
<https://doi.org/10.21659/rupkatha.v15n4.17>.
- Daskalaki, E., Psaroudaki, K., & Fragopoulou, P. (2024). Navigating The Future of Education: Educators' Insights on AI Integration and Challenges in Greece, Hungary, Latvia, Ireland and Armenia. *Journal of Applied Teaching and Learning*.
- Jančařík, A., & Dušek, O. (2024). *The problem of AI hallucination and how to solve it*. Charles University, Faculty of Education & Faculty of Mathematics and Physics.

Humpston, C. S., Adams, R. A., Benrimoh, D., Broome, M. R., Corlett, P. R., Gerrans, P., Horga, G., Parr, T., Pienkos, E., Powers, A. R., Raballo, A., Rosen, C., & Linden, D. E. J. (2019). From computation to the first-person: Auditory-verbal hallucinations and delusions of thought interference in schizophrenia-spectrum psychoses. *Schizophrenia Bulletin*, 45(Supplement 1), S56–S66.
<https://doi.org/10.1093/schbul/sby073>.

Klein, N. (2023, May 8). Ai machines aren't 'hallucinating'. but their makers are. The Guardian. <https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein>.

Mateos-Blanco, B., Álvarez-Ramos, E., Alejaldre-Biel, L., & Parrado-Collantes, M. (2024). Vademecum of artificial intelligence tools applied to the teaching of languages. *Journal of Technology and Science Education*, 14(1), 77–94.
<https://doi.org/10.3926/jotse.2522>

Marr, B. (2023). Chatgpt: What are hallucinations and why are they a problem for ai systems. Bernard Marr. <https://bernardmarr.com/chatgpt-what-are-hallucinations-and-why-are-they-a-problem-for-ai-systems/>.

University of Southern Queensland (UniSQ). (2024a) What is Artificial Intelligence? Learning and Teaching Futures Portfolio, Toowoomba, Queensland, viewed 24 August 2024,

University of Southern Queensland (UniSQ). (2024b) Staff Guidelines. Learning and Teaching Futures Portfolio, Toowoomba, Queensland, viewed 25 August 2024,

University of Southern Queensland (UniSQ). (2024c) AI and Assessment Professional Learning Series, Academic Division News Digest, viewed 26 August 2024, [Study with UniSQ | UniSQ](#).

Van Rensburg, H., & Reedy, N. (2024). *ChatGPT in higher education: A comprehensive study of the literature. Friend or foe?* University of Southern Queensland.

Weise, K., & Metz, C. (2023). When a.i. chatbots hallucinate. The New York Times. <https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html>.

Online Resources

<https://better-internet-for-kids.europa.eu/en/news/navigating-future-education-educators-insights-ai-integration-and-challenges-greece-hungary>.

<https://dig.watch/updates/greece-to-launch-ai-tool-for-personalised-education>.