https://doi.org/10.36315/2025v2end012
Education and New Developments 2025

RISK OF MISINFORMATION IN THE USE OF LLM BY UNIVERSITY STUDENTS

Gibrán Aguilar Rangel*, Claudia Cintya Peña Estrada*, Omar Bautista Hernández*, & Luis Ambrosio Velázquez García**
Accounting and Management Faculty, Universidad Autónoma de Querétaro (Mexico)
**PhD., ***ISC.*

Abstract

In recent years there has been an exponential growth in the use of artificial intelligence (AI) by university students, specifically large language models (LLM) like ChatGPT, this has led to a discussion about the ethics of using these models, especially when it comes to solving assignments and school projects. One point that is often overlooked is the misinformation these models can generate if not used properly, since students can trust the results, they get without verifying it. For this research we aimed to understand how university students use LLM and how they interpret the results they get. We designed a short questionnaire that was sent online to students in the Universidad Autonoma de Queretaro, in Mexico, results were anonymous in order to encourage truthful responses. The responses show that a lot of students are using LLMs to search for information, many times over search engines, and that even though they might detect some mistakes, there´s a lack of awareness on the risk of misinformation.

Keywords: *AI, misinformation, LLM, ChatGPT.*

1. Introduction

Encyclopedias were one of the earliest attempts to concentrate lots of information from different topics in a single source, this was no easy task of course, and some encyclopedias could span several volumes and be quite expensive. The advent of the internet promised to achieve the task of concentrating human knowledge into an accessible source, however, a big drawback was soon revealed, since anyone could publish information online, the quality of that information could be questionable, unlike encyclopedias that were reviewed prior to publication, on the internet anyone could publish anything and it was up to the user to distinguish fact from fiction. With time, some measures were taken in order to create a more verifiable environment, academic search engines, digital encyclopedias, online peer reviewed articles, among others.

Around this time one source started gaining traction among students and general audiences while being criticized as unreliable by academics, an always updatable online encyclopedia called Wikipedia. Its popularity stemmed from an ever-growing database that could give information about almost any topic, and that it was up to date in the most important (or popular) subjects. The criticism (rightly) pointed out that in theory anyone could edit the site, so the information could be not very reliable, so it was advised students shouldn't rely on this source due to concerns of the information being wrong. While these concerns were valid, Wikipedia's way of working has its own safeguards to avoid wrong data or correct it as soon is detected, that is not to say the system is flawless, less visited topics can have wrong data undetected for a long time and since the editors are a relatively small group of volunteers, they may imprint their own biases into the editing of the information (Konieczny, 2021).

With the introduction of virtual assistants, the goal was to provide easy answers to questions an user may have, so instead of the user searching through results on a search engine, the virtual assistant would provide it with the most relevant result to their question, in theory scouting through a database or a search engine, initial results were mixed, from Siri by Apple in 2011, to Alexa by Amazon in 2014, not all answers were relevant or some were plain wrong, other than basic tasks or basic questions these assistants were not as revolutionary as they were initially presented to be. In 2017 however, there was a breakthrough in how natural language processing worked, a neural network model named Transformer (Vaswani et al., 2017). This technological breakthrough had the effect of accelerating the rate in which these models could analyze texts and create statistical predictions in order to generate answers, this resulted in the creation of

large language models (LLMs) that were capable of creating seemingly original texts, derived from the analysis of large datasets, the most famous being ChatGPT.

The main appeal of a LLM is that it´s capable of understanding natural language, that is a person can ask a question like it would do to another individual and the model will understand it and (ideally) answer with the same language the correct response. There's a couple of problems with this approach, the main being that the ideal of right answers all the time is not being achieved, LLMs depend on the data they´re trained on, and said data can contain wrong information, since these models don´t have the ability to distinguish correct data from incorrect one, they can present wrong answers as being right (Barman et al., 2024). The other risk has to do with malicious intent, individual can make use of LLMs to knowingly distribute misinformation, thus misleading people to believe wrong affirmations (Park et al., 2024).

2. Methodology

The main objective was to identify the level of understanding of university students about the accuracy of the responses they might get when using a large language model (like ChatGPT). Another objective was to determine how students are using the LLM and their awareness in the topic of hallucinations.

Since the topic of the use of LLM to assist in academic workload is still a sensitive topic, the survey was anonymous, it was distributed via a messaging app and the platform used was Google forms with no email collection, all of this so the students could really feel their answers couldn't be traced back and would answer honestly. We obtained a total of 308 answers, all of them students at the Autonomous University of Queretaro, Mexico. There were 10 closed questions, one semi-open question and one open question so students could give their personal opinion about LLM.

3. Results and discussion

In the following table (Table 1) we present the results of the survey.

Table 1. Results of the survey.

QUESTION	OPTIONS	
HOW FREQUENTLY DO YOU USE CHATGPT OR OTHER AI MODELS?	Several times a day	56
	At least once a day	48
	4-5 times per week	38
	A couple times a week	116
	Occasionally (a couple times a month	50
WHAT DO YOU USUALLY USE IT FOR?	Solving doubts	220
	Search for information	214
	Summarize articles	82
	Writing assignments	40
TO SEARCH INFORMATION, DO YOU CONSIDER CHATGPT THE SAME AS GOOGLE OR OTHER SEARCH ENGINE?	Yes, they´re similar	116
	No, ChatGPT is better	86
	No, Google is better	36
	Yes, but I prefer Google	20
	Yes, but I prefer ChatGPT	50
WHICH SOURCE IS MORE TRUSTWORTHY, WIKIPEDIA OR CHATGPT?	Wikipedia	28
	ChatGPT	214
	Neither	66
HOW DO YOU CONSIDER THE INFORMATION THAT CHATGPT GIVES?	Absolutely trustworthy	8
	Generally reliable	148
	Sometimes reliable, sometimes not	146
	Absolutely unreliable	6
	True	212

I'VE DETECTED WRONG INFORMATION IN THE RESULTS PRESENTED BY CHATGPT	False	96
HOW FREQUENTLY HAVE YOU DETECTED WRONG INFORMATION?	Each query	0
	Very regularly	24
	Regularly	34
	Occasionally	78
	Very occasionally	100
IF YOU DOUBT OF AN ANSWER PROVIDED BY CHATGPT DO YOU USE ANY OTHER SOURCE TO FACTCHECK?	Yes	284
	No	24
IF YOU DO, WHICH ONE DO YOU USE?	Google	232
	Another AI	24
	Wikipedia	6
	Other	26
HAVE YOU USED CHATGPT FOR MATH PROBLEMS?	Yes	170
	No	138
HOW WERE THE RESULTS FOR MATH PROBLEMS?	Satisfactory	30
	Somewhat satisfactory	84
	Mixed	54
	Unpredictable	16

While the goal of the survey was to make students feel at ease so they could answer as honestly as possible, there´s room for potential distrust since the survey comes from a teacher at their university and they could feel its kind of a trap. That said, the results are interesting since they confirm some theories about how students perceive the LLM. The first couple of questions were intended to know how frequently do students us it and what for, while the percentage of students that admitted using it for writing assignments was on the lower end, this was expected since, as mentioned earlier, there´s an element of distrust, this is not as relevant since it wasn´t the main focus of this research, there are other studies that prove that given the opportunity, students will use it to solve assignments unethically (Dakakni & Safa, 2023).

The next set of questions dealt with the main topic of interest in this research, information and how do they perceive it, most of the students admitted using a LLM to search for information. A worrying answer is that most students considered that the LLM was the same as a search engine with a not so low percentage considering it better in order to search information, and while it may be common knowledge to most researchers and academics who work with LLM that they present a serious hallucination problem, that is, presenting wrong or biased information as if it were true (Lavrinovics et al., 2025), the students surveyed seemed not aware of this. In another interesting result, most students considered LLMs more trustworthy than Wikipedia, which may be derived of many years of teachers calling Wikipedia untrustworthy even if there´s evidence that nowadays it has a high degree of factual information (Konieczny, 2021).

A considerable number of students stated detecting wrong information in the results presented by the LLM, this makes the question about search engine vs LLM more interesting, that is, even if they detect wrong information in the LLM they consider it equal or better than a search engine, this might have to do with convenience, since a LLM will present just the information requested, while the search engine presents several results that need to be browsed, however, it can be said that results by a LLM have to go through a similar process, since they have to be verified, there´s a need for teaching students how to deal with this inconsistency, and a branch called AI ethical principles might gain some traction in the near future (Kajiwara & Kawabata, 2024).

Finally there were some questions related to LLMs and their ability with solving math problems, if a student has been following a math course and tries to solve problems using a LLM, chances are there might be more than a few answers wrong, LLMs are notably unpredictable at solving math problems (López Espejel et al., 2023), it is worth noting that in this aspect, a considerable number of students don´t use LLMs to try and solve math problems, this might be due to the nature of the courses, in which practice is more important than just getting a right answer, and the few students that answered that the results were satisfactory, may be lagging behind on the math course and unable to distinguish small mistakes.

4. Conclusions

For the foreseeable future, LLMs are here to stay, and as such they should be treated with the same caution as other tools, in the introduction we mentioned the Wikipedia example, when it surpassed the encyclopedia a lot of people warned about the risks, possible unverified information, the possibility of being edited by anyone, etc., and while a lot of this concerns have been fixed, the stigma continues to some extent, this has not been the case with LLMs so far, most concerns have to do with academic dishonesty, ethical concerns, etc., and while there are undoubtedly some studies about the misinformation problem and the hallucinations, they have failed to reach the general population that regards LLMs as almost infallible answering machines.

There needs to be an approach on educating students and teachers about LLMs, how they work, what they can or cannot do, what are hallucinations and how do they happen, and specially the risks of misinformation that they represent. There are a lot of proposals on how to use LLMs that do not address these concerns, they center on how to write prompts, for example, while failing to address that the best written prompt can still give you a wrong answer.

References

Barman, D., Guo, Z., & Conlan, O. (2024). The Dark Side of Language Models: Exploring the Potential of LLMs in Multimedia Disinformation Generation and Dissemination. *Machine Learning with Applications*, *16*(March), 100545. https://doi.org/10.1016/j.mlwa.2024.100545

Dakakni, D., & Safa, N. (2023). Artificial intelligence in the L2 classroom: Implications and challenges on ethics and equity in higher education: A 21st century Pandora's box. *Computers and Education: Artificial Intelligence*, *5*(August), 100179. https://doi.org/10.1016/j.caeai.2023.100179

Kajiwara, Y., & Kawabata, K. (2024). AI literacy for ethical use of chatbot: Will students accept AI ethics? *Computers and Education: Artificial Intelligence*, *6*(March), 100251. https://doi.org/10.1016/j.caeai.2024.100251

Konieczny, P. (2021). From Adversaries to Allies? The Uneasy Relationship between Experts and the Wikipedia Community. *She Ji*, *7*(2), 151-170. https://doi.org/10.1016/j.sheji.2020.12.003

Lavrinovics, E., Biswas, R., Bjerva, J., & Hose, K. (2025). Knowledge Graphs, Large Language Models, and Hallucinations: An NLP Perspective. *Journal of Web Semantics*, *85*(December 2024), 100844. https://doi.org/10.1016/j.websem.2024.100844

López Espejel, J., Ettifouri, E. H., Yahaya Alassan, M. S., Chouham, E. M., & Dahhane, W. (2023). GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts. *Natural Language Processing Journal*, *5*(August), 100032. https://doi.org/10.1016/j.nlp.2023.100032

Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. *Patterns*, *5*(5), 100988. https://doi.org/10.1016/j.patter.2024.100988

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-Decem*(Nips), 5999-6009.