EDITOR: Guansong Pang, gspang@smu.edu.sg

COLUMN: AI COMMUNITY

Rethinking Homework in the Age of Artificial Intelligence

Hazem Ibrahim and Rohail Asim, *New York University Abu Dhabi, Abu Dhabi, 129188, United Arab Emirates*

Fareed Zaffar, *Lahore University of Management Sciences, Lahore, 54792, Pakistan*

Talal Rahwan and Yasir Zaki, *New York University Abu Dhabi, Abu Dhabi, 129188, United Arab Emirates*

The evolution of natural language processing techniques has led to the development of advanced conversational tools such as ChatGPT, capable of assisting users with a variety of activities. Media attention has centered on ChatGPT's potential impact, policy implications, and ethical ramifications, particularly in the context of education. As such tools become more accessible, students across the globe may use them to assist with their homework. However, it is still unclear whether ChatGPT's performance is advanced enough to pose a serious risk of plagiarism. We fill this gap by evaluating ChatGPT on two introductory and two advanced university-level courses. We find that ChatGPT receives near-perfect grades on the majority of questions in the introductory courses but has not yet reached the level of sophistication required to pass in advanced courses. Moreover, adding a few full stops or typos may fool a machine learning algorithm designed to detect ChatGPT-generated text. These findings suggest that, at least for some courses, current artificial intelligence tools pose a real threat that can no longer be overlooked by educational institutions.

ChatGPT[1] is the latest tool available to students for learning and completing homework. It is a chat-based language model that utilizes artificial intelligence (AI) to assist with a wide variety of tasks. One of the key features of ChatGPT is its ability to understand and respond to natural language inputs, making it easy for students to ask for help and receive assistance in a conversational manner. While this tool has the potential to revolutionize the way students approach their studies, it also raises serious concerns. Some educators may worry that the use of such tools could lead to a decrease in critical-thinking skills, as students may become reliant on the tool for completing tasks rather than working through problems on their own.[2] Additionally, there is a potential for tools like ChatGPT to be used for academic dishonesty, such as copying answers or completing assignments without understanding the material.

Following the release of this budding technology, a number of articles have assessed its performance across a number of dimensions. For example, Bang et al.[3] evaluated ChatGPT on eight common NLP tasks, where they find that it outperforms other models on most tasks. Moreover, ChatGPT has been shown to pass exams from law and business schools as well as professional coding interviews.[4] However, while exams are typically conducted in a controlled environment where ChatGPT's use can be policed, less is understood regarding ChatGPT's performance on homework assignments, which primarily take place at the student's leisure.

RESULTS

To examine the difference in performance between ChatGPT and university students, we collected assignments from four courses taught at two UAE-based

Digital Object Identifier 10.1109/MIS.2023.3255599
Date of current version 4 April 2023.

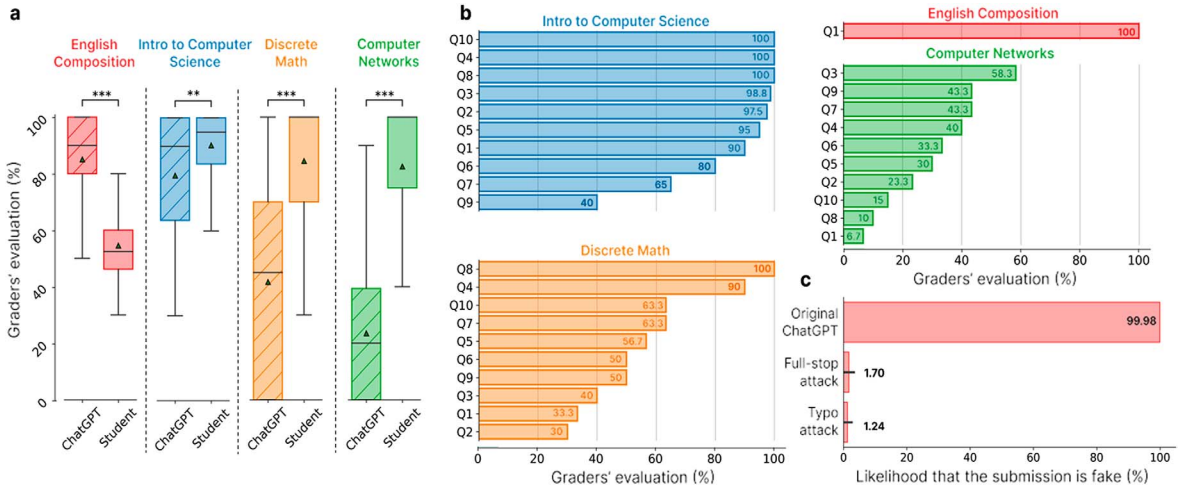

FIGURE 1. **(a)** Comparison of the grade distribution of ChatGPT submissions against that of student submissions in four courses. **(b)** The average grade given to the best ChatGPT answers for each question. **(c)** The output of a machine learning algorithm designed specifically to quantify the likelihood that a body of text was written by ChatGPT; this algorithm was given three different inputs: 1) the original ChatGPT essays, 2) the ChatGPT essays after inserting up to three full stops per essay, and 3) the ChatGPT essays after inserting up to six typos per essay. The p value is calculated using a paired t test: $^{**}p < 0.01$; $^{***}p < 0.001$. Intro: Introduction; Q: question.

universities. These included two introductory courses, namely, English Composition I (EC) and Introduction to Computer Science (ICS), as well as two more advanced courses, namely, Discrete Mathematics (DM) and Computer Networks (CN). For the first course, we randomly sampled 30 student essays from the ZAEBUC dataset,[5] all of which address the prompt, "How does social media affect individuals and society?" For each essay, we generated a ChatGPT counterpart targeting the same number of words as the student's essay. As for the remaining courses, we selected 10 questions from homework assignments, randomly sampled three student solutions per question, and generated three different answers to each question using ChatGPT. Next, we asked three professional graders (teaching assistants and professors) employed by the authors' institutions to blindly evaluate the submissions made by both students and ChatGPT. The graders were unaware of the purpose of the study or the fact that half of the submissions were, in fact, AI generated.

Figure 1(a) compares the grade distribution of ChatGPT submissions against that of student submissions in each course. As can be seen, the tool performed well in introductory courses. Specifically, ChatGPT's mean grade was substantially greater than that of students in EC (85% versus 55%) and was comparable in ICS (80% versus 90%). As for the more advanced courses, ChatGPT lagged far behind in DM (42% versus 85%) and in CN (24% versus 83%).

Figure 1(b) depicts the average grade given to the best ChatGPT answer out of the different answers generated for each question. More specifically, in the case of EC, we selected the essay with highest average grade out of the 30 ChatGPT-generated essays. As shown in the figure, ChatGPT was able to generate an essay that received a full mark from all three graders. As for the remaining courses, for each question, we selected the answer with the highest average grade out of the three ChatGPT-generated answers. As can be seen, in ICS, ChatGPT was able to receive a grade of 90% or above in seven out of the 10 questions. In contrast, the tool was able to receive an average grade of 90% or above on only two questions in DM and on none in CN. This exploratory analysis suggests that ChatGPT could pose a risk of plagiarism, at least in the context of writing essays and introductory programming, but remains lacking in more advanced courses.

Seeing the potential for ChatGPT to pose a risk in the context of essay plagiarism, we tested a machine learning algorithm designed specifically to detect AI-generated text, namely, the GPT-output detector.[6] To this end, we randomly sampled 10 of the 30 ChatGPT-generated essays and ran them through the aforementioned detector. As can be seen in Figure 1(c), the use of ChatGPT to generate these essays could be detected with extremely high accuracy (99.98%). However, if such a detector is to be deployed, it has to face the challenge of obfuscation, i.e., the deliberate manipulation of ChatGPT's output to

evade detection. To demonstrate this point, we tested two extremely simple "attacks," which we call the full-stop attack and the typo attack, where a few full stops or typos are strategically inserted into the text. Figure 1(c) depicts the outcome of the GPT-output detector after manually inserting no more than three full stops or no more than six typos per essay. As can be seen, the detector can easily be misled, opening the question of how to develop foolproof AI-plagiarism detectors. Both ChatGPT and anonymized student submissions as well as all original ChatGPT essays and their obfuscated counterparts can be found in the following repository: https://github.com/comnetsAD/ChatGPT_homework.

DISCUSSION

There are numerous ethical issues surrounding the use of AI tools in the context of education. One major concern is the potential for copyright infringement. When an AI tool generates content, it is using pre-existing knowledge and information to create something new. ChatGPT, for instance, is trained on a curated version of the common crawl dataset, built upon text available on the Internet. However, if this content is not properly cited, it could be considered a violation of copyright laws. Indeed, this has been an ongoing discussion with regard to stable diffusion, for instance, where many artists have noted that their work had been used without their consent.[7] This is especially problematic in the academic setting, where proper citation is crucial for both ethical and legal reasons. For instance, some schools and academic publishing venues have already banned the use of ChatGPT.[8] The text on which such models are trained also may exhibit biases, as research has shown that web content remains heavily skewed toward rich, Western countries.[9] While efforts have been made by OpenAI to limit ChatGPT from generating expletive or suggestive content, models trained on such information would inherently exhibit these same biases. In addition to these concerns, ChatGPT—and the more advanced successors that will inevitably follow—raise philosophical questions about the nature of human creativity. Since the platform's release to the public, many have argued that using AI to generate content takes away from the unique and original ideas of creators.[10] Given these ethical and philosophical concerns, it is not surprising that many believe AI tools, such as ChatGPT, should be banned from use in creative and academic outlets.

ChatGPT offers both opportunities and challenges in the context of educational framework reform, both in the teaching stage as well as in the evaluation process. Integrating ChatGPT into the classroom and, thus, allowing students to utilize the tool to improve their writing skills may allow nonnative English speakers to focus their efforts toward higher-order thinking, essentially lifting the English-writing barrier. Such an advancement, thus, has the potential to democratize the arena of academic writing, giving educators the opportunity to focus more on the ideation and editing processes of writing rather than on grammar or spelling. For instance, educators may prompt students to critically examine the inherent biases exhibited by the tool, correct factual mistakes, or alter the structure of a given body of text generated by ChatGPT. If left unchanged, current educational frameworks may face crucial challenges in verifying students' abilities and knowledge. ChatGPT is a warning call, signaling the dawn of AI-assisted schoolwork. The following decade will determine the scope of AI adoption in this context as well as the policies and norms surrounding its use.

REFERENCES

1. "Chatgpt: Optimizing language models for dialogue." OpenAI. Accessed: Jan. 9, 2023. [Online]. Available: https://openai.com/blog/chatgpt/
2. L. Meckler and P. Verma, "Teachers are on alert for inevitable cheating after release of ChatGPT," *The Washington Post*, Dec. 2022. Accessed: Jan. 9, 2023. [Online]. Available: https://www.washingtonpost.com/education/2022/12/28/chatbot-cheating-ai-chatbotgpt-teachers/
3. Y. Bang et al., "A multitask, multilingual, multimodal evaluation of chatGPT on reasoning, hallucination, and interactivity," 2023, *arXiv:2302.04023*.
4. J. U. Nisa, "Here's a list of some of the exams chatGPT has passed so far," *Wonderful Eng.*, Feb. 2023. Accessed: Jan. 9, 2023. [Online]. Available: https://wonderfulengineering.com/heres-a-list-of-some-of-the-exams-chatgpt-has-passed-so-far/
5. N. Habash and D. Palfreyman, "ZAEBUC: An annotated Arabic-English bilingual writer corpus," in *Proc. 13th Lang. Resour. Eval. Conf.*, 2022, pp. 79–88.
6. H. Face. "Gpt-2 output detector demo." OpenAI. Accessed: Jan. 9, 2023. [Online]. Available: https://openai-openai-detector.hf.space/
7. J. Vincent, "The scary truth about AI copyright is nobody knows what will happen next," *The Verge*, Jan. 2023. Accessed: Jan. 9, 2023. [Online]. Available: https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data
8. M. Yang, "New York city schools ban AI chatbot that writes essays and answers prompts," *The Guardian*, Jan. 2023. Accessed: Jan. 9, 2023. [Online]. Available:

https://www.theguardian.com/us-news/2023/jan/06/new-york-city-schools-ban-ai-chatbot-chatgpt
9. M. Graham, S. De Sabbata, and M. A. Zook, "Towards a study of information geographies: (IM)mutable augmentations and a mapping of the geographies of information," *Geography Environ.*, vol. 2, no. 1, pp. 88–105, May 2015, doi: 10.1002/geo2.8.
10. K. Roose, "An A.I.-generated picture won an art prize. Artists aren't happy," *NY Times*, Sep. 2022. Accessed: Jan. 9, 2023. [Online]. Available: https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html

HAZEM IBRAHIM is a research assistant at New York University Abu Dhabi, Abu Dhabi, 129188, United Arab Emirates. Contact him at hazem.ibrahim@nyu.edu.

ROHAIL ASIM is a Ph.D. student at New York University Abu Dhabi, Abu Dhabi, 129188, United Arab Emirates. Contact him at ra3106@nyu.edu.

FAREED ZAFFAR is an associate professor at Lahore University of Management Sciences, Lahore, 54792, Pakistan. Contact him at fareed.zaffar@lums.edu.pk.

TALAL RAHWAN is an associate professor at New York University Abu Dhabi, Abu Dhabi, 129188, United Arab Emirates. Contact him at talal.rahwan@nyu.edu.

YASIR ZAKI is an assistant professor at New York University Abu Dhabi, Abu Dhabi, 129188, United Arab Emirates. Contact him at yasir.zaki@nyu.edu.