

Mark Shen

Santa Rosa Junior College

01 March 2023

Student Statement on ChatGPT

Artificial Intelligence (AI) has become a widespread sensation for its revolutionary advancements. At the forefront of some of these advancements is OpenAI's Chat Generative Pre-Trained Transformer (ChatGPT) for its ability to mimic human speech. It has become the focus in academia for widespread concerns of academic dishonesty; students might use ChatGPT to supplement their work. I have been given the opportunity to offer a student perspective on this issue as Santa Rosa Junior College looks into defining the proper use of ChatGPT. I will be detailing a brief overview of ChatGPT as well as my personal experience with the chatbot. It is important to note that I am not an expert in artificial intelligence or any of its subset technology. Any explicitly described behavior or functionality will be referenced from OpenAI, the creators of ChatGPT, or cited from another accredited source.

ChatGPT

Structure. OpenAI describes ChatGPT as “a chatbot...fine-tuned from a model in the GPT-3.5 series”. Dr. Deep Ganguli, director of research at Stanford’s Institute for Human-Centered Artificial Intelligence (HAI), and Alex Tamkin, a Ph.D. student in Computer Science at Stanford, published an article describing the capabilities of GPT-3 and specific details on how the language model was trained: “GPT-3 has 175 billion parameters and was trained on 570 gigabytes of

text...GPT-3 is able to perform tasks it was not explicitly trained on.” This massive amount of data and training allows ChatGPT, using the GPT-3 language model, to “interact in a conversational way” (OpenAI). ChatGPT’s near-human eloquence is a primary factor in the difficulty of detecting AI-generated content; since ChatGPT is modeled after human text samples, it will generate responses that resemble human speech.

Limitations. While many believe ChatGPT to be the panacea to their problems, academic or otherwise, ChatGPT is still fallible. OpenAI documents the chatbot’s limitations as generating reasonable but incorrect or nonsensical answers, inaccurate sensitivity to prompt phrasing, being excessively verbose and overusing certain phrases, guessing what the user intended rather than asking clarifying questions, and being unable to block all inappropriate requests. These limitations pose numerous challenges to ChatGPT users.

Detection. ChatGPT’s limitations, as well as its structure, influence the validity and reliability of ChatGPT detection. Detection strategies have shifted into two main groups, manual and automatic. Manual, or human, detection is a strategy often employed by those in an academic setting to determine if student submissions are their work or the work of artificial intelligence. Due to ChatGPT’s capabilities of writing text akin to human speech, this strategy is incredibly inconsistent and highly subjective. Instructors would need to establish a definitive benchmark for one’s writing early on and use that benchmark as a reference for any future material. The success of this method is less than reliable and poses a potential risk of false positives.

The automatic approach involves using artificial intelligence to detect ChatGPT and other chatbots. OpenAI themselves have developed a classifier meant to distinguish between human text and artificial intelligence-generated text. This tool functions by using artificial intelligence and comparing existing data of human and artificial intelligence-generated responses. The comparison of that data can offer a likelihood of a submitted text being written by a human or by artificial intelligence on a scale ranging from “likely human-written” to “likely AI-written.” This strategy is similarly as uncertain as manual review. OpenAI found that their tool correctly identifies AI-written text 26% of the time and incorrectly classifies human-written text as AI-written 9% of the time. The low probability is likely caused by several limitations which OpenAI lists: the classifier is unreliable on text shorter than 1,000 characters; human-written text may be confidently labeled as AI-written by the classifier; the classifier is unreliable with non-English text; the classifier is unreliable with predictable text; AI-written text is able to be edited to evade classification; the classifier’s has poor calibration outside of its training data. It is telling that even the creators of ChatGPT are unable to create a tool that can reliably detect text written by its own chatbot. Thus, detection is riddled with uncertainty due to the limitations of the current technology. OpenAI cautions users to use artificial intelligence classifiers “not...as a primary decision-making tool, but instead as a complement to other methods.”

My Experience

Non-academic. I have used ChatGPT extensively for many extracurricular projects. While the focus of these projects varies, most of them fall within the context of coding. My most common use of ChatGPT for these projects is asking the chatbot to generate an example script with the

prompt, including a basic description of the script's intended function. For example, if I use the prompt "Generate a script that prints a message," ChatGPT would supply me with the script that prints that message. In more complex prompts, I've found that ChatGPT's generated response is close to what was asked, but is not usable in its original form. These responses require user amendments for the script to function as intended; the most common issue is that ChatGPT's code contains syntax, or formatting, errors that prevent the script from functioning in a certain environment. These syntax errors are akin to using a different language's punctuation.

Academic non-personal. I have not personally used ChatGPT for any academic applications, specifically due to the potential concerns of academic dishonesty. Thus, I will be including the experiences of other students I've spoken with to provide some insight into student use. Most students that I have spoken to have used ChatGPT for some academic writing applications. Some were more brazen about their use, submitting the generated response verbatim as their own; most took a less direct approach, taking snippets of the generated response and molding it into their text. All of the students that I spoke with were aware of the possible perception of academic dishonesty, but most disagreed. A majority of students reasoned that if they used the generated response as a resource to garner inspiration, rather than using it word for word, then referencing an AI would be no different than referencing any human. AI use in academia seems to have also gained acceptance among some instructors. Ethan Mollick, associate professor at the University of Pennsylvania's Wharton School, has adopted the policy that ChatGPT is a tool and students must know how to use it. Mollick advocates for academic honesty by encouraging transparency and promoting an understanding of ChatGPT, rather than safeguarding it behind detection

software and academic discipline. Mollick also resigns to the reality that detecting ChatGPT is challenging saying, “I probably couldn't have stopped [students] even if I didn't require [ChatGPT literacy].”

Academic personal. I was accused of academic dishonesty by one of my instructors. They believed that one of the passages that I turned in for an assignment was taken verbatim from ChatGPT, and they presented screenshots of a conversation with ChatGPT to corroborate their accusations. Out of an abundance of caution, I began preparing to dispute an academic dishonesty report because I was fully cognisant of the consequences should things be misunderstood. The process itself was somewhat standard; I met with the instructor, then the department chair, and eventually with the dean of academics. Fortunately, my instructor was not interested in academic reprimands and more focused on ethical awareness and reform, so the report became a moot point; the assignment was later dropped so it no longer affected my grade. Still, that multi-week endeavor raised some alarming questions. What would happen to students who are less willing or able to fight accusations? What would happen to students should their instructor be less accommodating than mine had been? Since my instructor had been so flexible, I had the freedom to shift my focus to using the incident as an opportunity for change. I believe that this situation is a perfect test case for the many that are sure to follow as ChatGPT gains more popularity and widespread use; I also hope to use this situation to allow policymakers at SRJC to reflect on the potential hazards of more traditional safeguards.

Ultimately, I share the goal of many of my peers. I want to continue to improve and adapt our systems to better our communities. I see my situation as an opportunity for meaningful understanding, rather than a deterrence, and I want to use the platform that I have been given to help prepare students and instructors alike for the unknown future of artificial intelligence. I concede the fact that some students will take advantage of leniency, and I fully advocate for proper policies to be established to ensure academic quality, but I hope that those policies are not created at the expense of students. Erring on the side of personal accountability may be less resolute than a detection and prevention model, but there is far less risk that students will be unjustly caught in the crossfire.

Works Cited

- Kirchner, Jan Hendrik, et al. "New AI Classifier for Indicating AI-written Text." *OpenAI*, 31 Jan. 2023, openai.com/blog/new-ai-classifier-for-indicating-ai-written-text#JanHendrik%20Kirchner.
- OpenAI. "Introducing ChatGPT." *OpenAI*, 30 Nov. 2022, openai.com/blog/chatgpt.
- Tamkin, Alex, and Deep Ganguli. "How Large Language Models Will Transform Science, Society, and AI." *Stanford University Human-Centered Artificial Intelligence*, Stanford University, 5 Feb. 2021, hai.stanford.edu/news/how-large-language-models-will-transform-science-society-and-ai.
- Wood, Patrick, and Mary Louise Kelly. "'Everybody Is Cheating': Why This Teacher Has Adopted an Open ChatGPT Policy." *NPR*, National Public Radio, 26 Jan. 2023, www.npr.org/2023/01/26/1151499213/chatgpt-ai-education-cheating-classroom-wharton-school.