A Novel Approach to Detecting Academic Dishonesty Involving Artificial Intelligence

Jang, Jun (School: Oxford High School)

The rapid emergence of generative AI, particularly large-language-model-based conversational chatbots like ChatGPT, has spurred significant advancements. However, this progress has brought pressing concerns about unethical use in an academic space. In response, there have been several developments in AI-generated text detection tools that can effectively detect whether a text is generated by machines or written by humans. However, these current models of GPT detectors determine their results by only analyzing one text. This project aims to provide a novel comparative approach to detecting unethical AI use by utilizing well-known metrics while also introducing new text metrics such as vocabulary levels. The key distinguishing feature of this project is its comparative approach: rather than assessing individual texts to detect AI usage like typical GPT detectors, this project compares metrics between two distinct pieces of text– a known human written sample and a suspected AI-generated text. Following the comparisons, statistical techniques were utilized to determine whether the two pieces originated from humans. If the disparity reached a certain level, it suggested the involvement of AI in generating the text. By providing a novel comparative approach, this research serves as a tool to assist teachers to discern potential unethical uses of AI in an academic setting.

Awards Won:

National Security Agency Research Directorate : Third Place Award “Cybersecurity”