Summarization of 8 Big Problems with OpenAI's ChatGPT

Muyuan Zhang

The article discussed eight problems of OpenAI's AI chatbot ChatGPT. ChatGPT is a language model designed to produce natural human language, trained on text from Wikipedia, blog posts, books, academic articles, etc.

- Security threats and privacy concerns. ChatGPT accidentally shared users' chat histories in March 2023, and Samsung employees mistakenly leaked confidential information via ChatGPT. If employees use ChatGPT to check their code, the code will be stored on OpenAI's servers, which is another risk. In response to the Italian data regulator's demand to stop processing Italian users' data, OpenAI added an age restriction, made its Privacy Policy more visible and provide an option to exclude users' data from training ChatGPT, but the improvement is only implemented in Italy.
- Concerns over ChatGPT training. OpenAI may scoop up personal information or use copyrighted images when it trains ChatGPT. Artists also claim that they never consented to their work to train AI models. ChatGPT can generate answers or artworks without any reference or notification to the authors of the original texts or artworks.
- ChatGPT generates wrong answers in basic math, medical advice and historical events. The reason is that it makes a series of guesses instead of using the Internet to locate answers. ChatGPT is a neural network where each word it trained on is represented in numbers. When asked for links, citations and references, it will usually make them up based on the data it was trained on.
- ChatGPT has bias baked into its system. ChatGPT has been shown to produce some answers that discriminate against gender, race and minority groups. This is partially because ChatGPT uses a combination of both supervised and unsupervised learning. OpenAI is addressing biased behavior by collecting feedback from users.
- ChatGPT might take jobs from humans. Duolingo and Khan Academy have integrated GPT-4, which could be the beginning of AI holding human jobs. Paralegals, lawyers, copywriters, journalists, and programmers are also facing disruption.
- ChatGPT is challenging education. ChatGPT can write assignments, summarization, code and cover letters which can cause academic cheating. According to this article, ChatGPT is able to generate code with acceptable errors and can be fixed by human, which lowers the demand of human employees.
- ChatGPT could cause real-world harm. Some people tried to jailbreak ChatGPT and built an AI model that could bypass OpenAI's guard rails. Hackers create online scams and sell rule-less ChatGPT services that produce malware and phishing emails. The scale of text and data generated by ChatGPT can make the information on the Internet more questionable and bring the deepfake technology. Human volunteers are required to sort through the backlog and provide sort of supervised learning environment.
- OpenAI holds all the power. OpenAI is a private company that selects the data to train ChatGPT and determines how fast it has new developments. It started a race between big tech companies to launch the big AI model, which can be a big threat to distribution and antitrust.

ChatGPT, as a chatbot that can interact with humans, enables more people to get involved in AI, bringing people with less knowledge of cybersecurity and information security into the game. Humans have invented and used artificial intelligence without really understanding it.