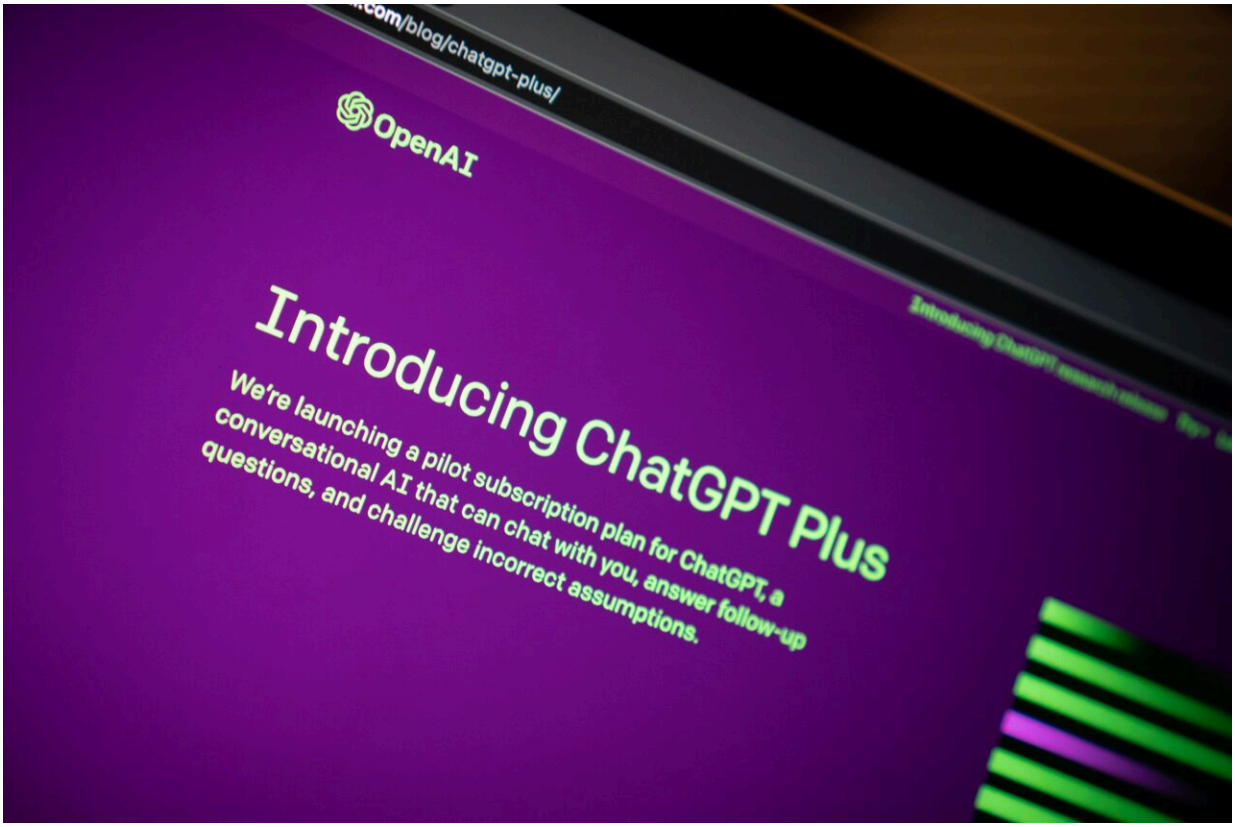
Text generators may plagiarize beyond 'copy and paste'

February 17 2023, by Francisco Tutella



Credit: Unsplash/CC0 Public Domain

Students may want to think twice before using a chatbot to complete their next assignment. Language models that generate text in response to user prompts plagiarize content in more ways than one, according to a

Penn State-led research team that conducted the first study to directly examine the phenomenon.

"Plagiarism comes in different flavors," said Dongwon Lee, professor of information sciences and technology at Penn State. "We wanted to see if language models not only copy and paste but resort to more sophisticated forms of plagiarism without realizing it."

The researchers focused on identifying three forms of plagiarism: verbatim, or directly copying and pasting content; paraphrase, or rewording and restructuring content without citing the original source; and idea, or using the main idea from a text without proper attribution. They constructed a pipeline for automated plagiarism detection and tested it against OpenAI's GPT-2 because the language model's training data is available online, allowing the researchers to compare generated texts to the 8 million documents used to pre-train GPT-2.

The scientists used 210,000 generated texts to test for plagiarism in pre-trained language models and fine-tuned language models, or models trained further to focus on specific topic areas. In this case, the team fine-tuned three language models to focus on scientific documents, scholarly articles related to COVID-19, and patent claims. They used an open-source search engine to retrieve the top 10 training documents most similar to each generated text and modified an existing text alignment algorithm to better detect instances of verbatim, paraphrase and idea plagiarism.

The team found that the language models committed all three types of plagiarism, and that the larger the dataset and parameters used to train the model, the more often plagiarism occurred. They also noted that fine-tuned language models reduced verbatim plagiarism but increased instances of paraphrase and idea plagiarism. In addition, they identified instances of the language model exposing individuals' private

information through all three forms of plagiarism. The researchers will present their findings at the 2023 ACM Web Conference, which takes place April 30-May 4 in Austin, Texas.

"People pursue large language models because the larger the model gets, generation abilities increase," said lead author Jooyoung Lee, doctoral student in the College of Information Sciences and Technology at Penn State. "At the same time, they are jeopardizing the originality and creativity of the content within the training corpus. This is an important finding."

The study highlights the need for more research into text generators and the ethical and philosophical questions that they pose, according to the researchers.

"Even though the output may be appealing, and language models may be fun to use and seem productive for certain tasks, it doesn't mean they are practical," said Thai Le, assistant professor of computer and information science at the University of Mississippi who began working on the project as a doctoral candidate at Penn State. "In practice, we need to take care of the ethical and copyright issues that text generators pose."

Though the results of the study only apply to GPT-2, the automatic plagiarism detection process that the researchers established can be applied to newer language models like ChatGPT to determine if and how often these models plagiarize training content. Testing for plagiarism, however, depends on the developers making the training data publicly accessible, said the researchers.

The current study can help AI researchers build more robust, reliable and responsible language models in future, according to the scientists. For now, they urge individuals to exercise caution when using text generators.

"AI researchers and scientists are studying how to make language models better and more robust, meanwhile, many individuals are using language models in their daily lives for various productivity tasks," said Jinghui Chen, assistant professor of information sciences and technology at Penn State. "While leveraging language models as a search engine or a stack overflow to debug code is probably fine, for other purposes, since the language model may produce plagiarized content, it may result in negative consequences for the user."

The plagiarism outcome is not something unexpected, added Dongwon Lee.

"As a stochastic parrot, we taught language models to mimic human writings without teaching them how not to plagiarize properly," he said. "Now, it's time to teach them to write more properly, and we have a long way to go."

More information: Do Language Models Plagiarize?, pike.psu.edu/publications/www23.pdf

Provided by Pennsylvania State University

Citation: Text generators may plagiarize beyond 'copy and paste' (2023, February 17) retrieved 3 May 2024 from https://techxplore.com/news/2023-02-text-generators-plagiarize.html