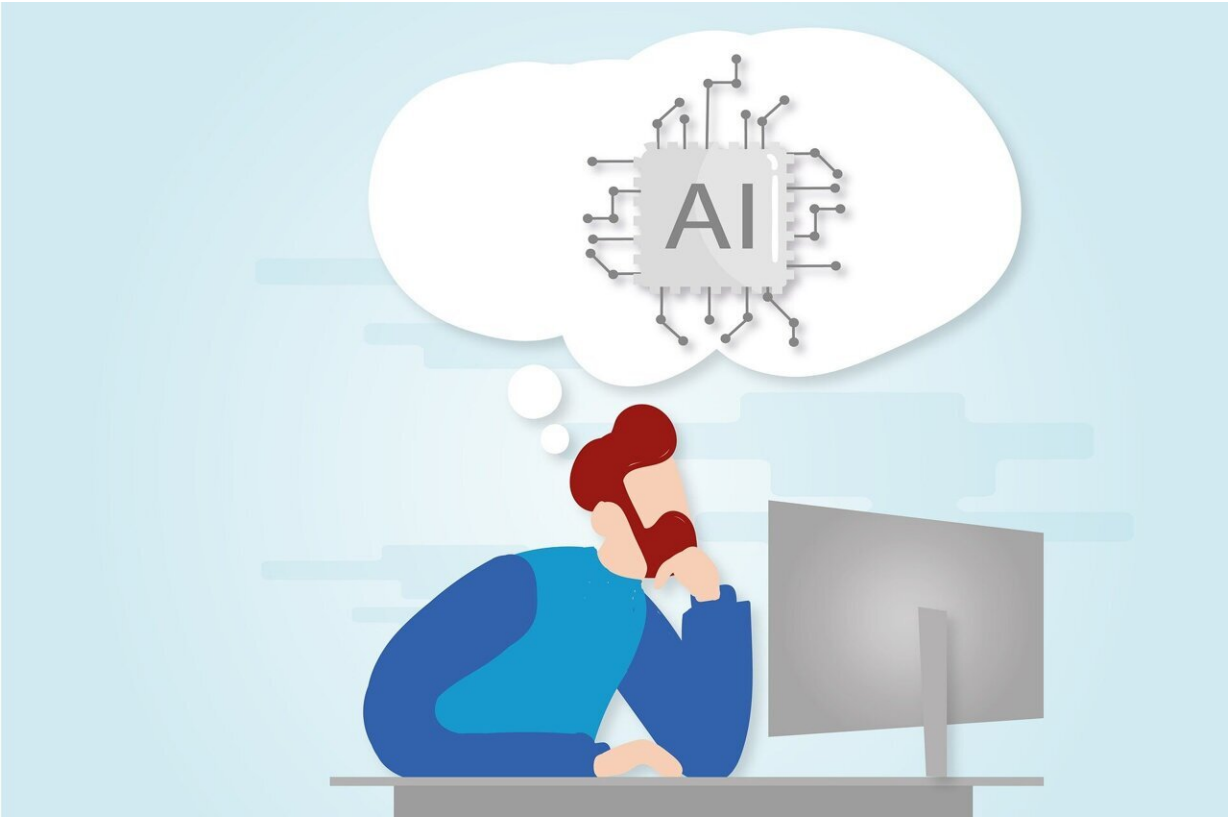
ChatGPT poem regurgitation raises ethical questions

January 9 2024, by Patricia Waldron



Credit: Pixabay/CC0 Public Domain

Ask ChatGPT to find a well-known poem and it will probably regurgitate the entire text verbatim—regardless of copyright law—according to a new study by Cornell researchers.

The study showed that ChatGPT, a large language model that generates text on demand, was capable of "memorizing" poems, especially famous ones commonly found online. The findings pose ethical questions about how ChatGPT and other proprietary artificial intelligence models are trained—likely using data scraped from the internet, researchers said.

"It's generally not good for large language models to memorize large chunks of text, in part because it's a privacy concern," said first author Lyra D'Souza, a former computer science major and summer research assistant. "We don't know what they're trained on, and a lot of times, private companies can train proprietary models on our private data."

D'Souza presented this work, "The Chatbot and the Canon: Poetry Memorization in LLMs," at the Computational Humanities Research Conference in Paris.

"We chose poems for a few reasons," said senior author David Mimno, associate professor of information science in the Cornell Ann S. Bowers College of Computing and Information Science. "They're short enough to fit in the context size of a language model. Their status is complicated: Many of the poems we studied are technically under copyright, but they're also widely available from reputable sources like the Poetry Foundation. And they're not just any document. Poems are supposed to be surprising, they're supposed to mean something to people. In some sense, poems want to be memorized."

ChatGPT and other large language models are trained to generate text by predicting the most likely next word over and over again based on their training data, which is mostly webpages. Memorization can occur when that training data includes duplicated passages, because the duplication reinforces that specific sequence of words. After being exposed to the same poem repeatedly, for example, the model defaults to reproducing the poem's words verbatim.

D'Souza tested the poem-retrieving capabilities of ChatGPT and three other language models: PaLM from Google AI, Pythia from the non-profit AI research institute EleutherAI and GPT-2, an earlier version of the model that ultimately yielded ChatGPT, both developed by OpenAI. She came up with a set of poems from 60 American poets from different time periods, races, genders and levels of fame, and fed the models prompts asking for the poems' text.

ChatGPT successfully retrieved 72 of the 240 poems, while PaLM came up with only 10. Neither Pythia nor GPT-2 could produce entire poems. Pythia responded with the same phrase over and over again, while GPT-2 produced nonsense text, researchers found.

Inclusion in the poetry canon was the most important factor in whether the chatbot had memorized a poem, while the poet's race, gender and era were not as significant. The most reliable predictor of memorization was if the poem had appeared in a "Norton Anthology of Poetry," specifically the 1983 edition.

D'Souza also noticed that ChatGPT's responses changed over time as the model evolved. When she first queried the chatbot in February 2023, it could not say it didn't know a poem—instead it would fabricate one or recycle a poem from another author. By July 2023, if ChatGPT didn't know the poem, it would ask if the poem even existed—putting the blame on the user.

That troubled D'Souza. "As we have more powerful tools that tell us they know everything, it becomes even more important to make sure we're not just learning from one source," she said.

Additionally, in February, ChatGPT had no limits due to copyright. But by July, sometimes it would respond that it couldn't produce a copyrighted poem. However, it would usually reproduce the poem if

asked again, D'Souza found.

This study looked only at American poets, but the next step will be to see how chatbots respond to requests in different languages and whether factors such as the length, meter and rhyming pattern of a poem make it more or less likely to be memorized, D'Souza said

"ChatGPT is a really powerful new tool that's probably going to be part of our lives moving forward," she said. "Figuring out how to use it responsibly and use it transparently is going to be really important."

Provided by Cornell University

Citation: ChatGPT poem regurgitation raises ethical questions (2024, January 9) retrieved 9 May 2024 from https://techxplore.com/news/2024-01-chatgpt-poem-regurgitation-ethical.html