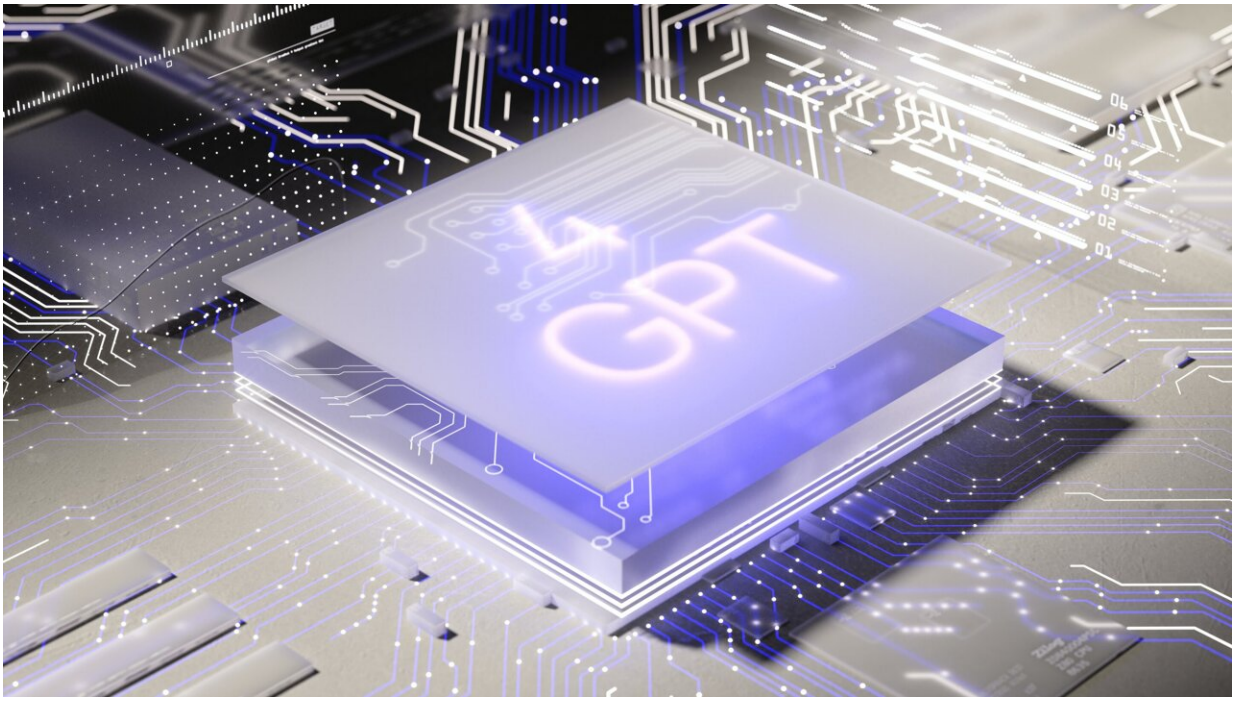
'Open' alternatives to ChatGPT are on the rise, but how open is AI really?

July 18 2023



Credit: Unsplash/CC0 Public Domain

OpenAI's ChatGPT seems ubiquitous, but open source versions of instruction-tuned text generators are gaining the upper hand. In just six months, at least 15 serious alternatives have emerged, all of which have at least one important advantage over ChatGPT: they are a lot more transparent. Insight into training data and algorithms is key for responsible use of generative AI, a team of linguists and language

technology researchers at Radboud University claim.

The researchers have mapped this rapidly evolving landscape in a paper and a live-updated website. This shows there are many working alternative "open source" text generators, but also that openness comes in degrees and that many models inherit legal restrictions. They sound a note of cautious optimism.

Lead researcher Andreas Liesenfeld says, "It's good to see so many open alternatives emerging. ChatGPT is so popular that it is easy to forget that we don't know anything about the training data or other tricks being played behind the scenes. This is a liability for anyone who wants to better understand such models or build applications on them. Open alternatives enable critical and fundamental research."

More and more open

Corporations like OpenAI sometimes claim that AI must be kept under wraps because openness may bring "existential risks," but the researchers are not impressed. Senior researcher Mark Dingemanse says, "Keeping everything closed has allowed OpenAI to hide exploitative labor practices. And talk of so-called existential risk distracts from real and current harms like confabulation, biased output and tidal waves of spam content." Openness, the researchers argue, makes it easier to hold companies responsible and accountable for the models they make, the data that goes into them (often copyrighted), and the texts that come out of them.

The research shows that models vary in how open they are: many only share the language model, others also provide insight into the training data, and quite a few are extensively documented. Mark Dingemanse says, "In its present form, ChatGPT is unfit for responsible use in research and teaching. It can regurgitate words but has no notion of

meaning, authorship, or proper attribution. And that it's free just means we're providing OpenAI with free labor and access to our collective intelligence. With open models, at least we can take a look under the hood and make mindful decisions about technology."

Some additional points

- New models appear every month, so the paper is mainly a call for action to track their openness and transparency in a systematic way. An accompanying website makes this possible
- Many models borrow elements from one another, which can lead to murky legal situations. For instance, the popular Falcon 40B-instruct model builds on a dataset (Baize) meant strictly for research purposes, but still the Falcon makers encourage commercial uses.
- A key reason ChatGPT feels so fluid is the human labor that goes into the instruction-tuning step (RLHF), in which model output is trimmed and pruned to make it sound more docile and conversational. Open models enable research into what makes people so susceptible to the suggestion of true interactivity.

The researchers will present their findings at the international conference on Conversational User Interfaces in Eindhoven, NL, July 19–21, and their paper is available on the *arXiv* preprint server.

More information: Andreas Liesenfeld et al, Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators, *Proceedings of the 5th International Conference on Conversational User Interfaces* (2023). DOI: 10.1145/3571884.3604316. On *arXiv*: DOI: 10.48550/arxiv.2307.05532

Provided by Radboud University

Citation: 'Open' alternatives to ChatGPT are on the rise, but how open is AI really? (2023, July 18) retrieved 15 May 2024 from https://techxplore.com/news/2023-07-alternatives-chatgpt-ai.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.