CEST
Centro de Estudos Sociedade e Tecnologia
USP
Universidade de São Paulo
Bulletin - Volume 8, Number 1, February/2023

ChatGPT: if you haven't used it yet, one day you might use it without realizing it

Eduardo Bertassi

ChatGPT or Chat Generative Pre-Trained Transformer is a chatbot launched by OpenAI, initially a non-profit research company founded in 2015 by Elon Musk (founder of PayPal, CEO of Tesla, SpaceX and Twitter), Sam Altman (current CEO of Open AI), Peter Tiel (co-founder of PayPal and Palantir Technologies), Reid Hoffman (co-founder of LinkedIn) among other technology moguls.

Around the beginning of December 2022 ChatGPT started to gain space in Brazilian newspapers, but its achievements became more prominet when it was publicly released as a prototype on November 30, 2022; how it worked was (and still is) simple: all you had to do was to type a question in a text box on the system's interface and the chatbot would provide you an answer. However, unlike other more rudimentary chats, it was possible to create poems, songs, TV scripts and even debug lines computer programs source code based on questions inquired by users.

In its answers ChatGPT can combine different fields of knowledge in completely unusual and even fun ways

The Chatbot responses are impressive, but not always correct. Like any Artificial Intelligence (AI) system, ChatGPT is not (yet) perfect, as such systems learn through statistical regularities from their training data. As explained by the company itself on its website, the system uses a model called RLHF or Reinforcement Learning from Human Feedback, in which a database is trained under human supervision. Simply put, from hundreds of conversations with the chatbot, human trainers rank the answers with the best results using a reward reinforcement model and then this adjusts the answers to similar questions that may be asked in the future.

The results that can be obtained are impressive and it is not out of nothing that they draw attention, because in its answers ChatGPT can combine different fields of knowledge in completely unusual and even fun ways. You can ask it to debug a software source code but answering as if it were a pirate or ask it to explain how a famous data sorting algorithm called Bubble Sort works as if it were a smart gangster.

The company itself recognizes that its system has limitations:

- some answers, although plausible, may be incorrect or meaningless (must be checked),
- some unanswered questions, if reformulated with minor adjustments, may finally have a valid answer (it is recommended to experiment different question formulations),
- certain extremely detailed responses end up having biased training data due to instructors that prefer longer responses,
- some inappropriate questions may be answered with harmful instructions or present biased behaviour (the company is trying to avoid such answers through moderation rules and even intended blocks).

What has been discussed about ChatGPT are the ethical, philosophical, economic and even psychological issues related to the use of this type of technology, but this is something that always happens when a new technology emerges, after all, you never know at first how something completely new and sometimes disruptive can significantly impact society and its individuals because that is something very difficult to predict during the development phase of a project. Sometimes, new technologies bring unanswered dilemmas due to the limitations of current ethical models and even end up needing the creation of new ethical principles to guide their use.

It is noteworthy that Elon Musk, one of OpenAI's founders resigned from the company's board of directors at the end of 2018 because he stopped agreeing with the direction the company was taking. The company that started out as a non-profit venture "to advance digital intelligence in the way that is most likely to benefit humanity as a whole, unconstrained by a need to generate financial return" (as described on the company's website in 2015) seems to have taken a different path. Musk stated that "OpenAI was created as an open source (which is why I named it "Open" AI), non-profit company to serve as a counterweight to Google, but now it has become a closed source, maximum-profit company effectively controlled by Microsoft". But why this concern to keep ChatGPT as an open-source project?

Open-source software or systems, unlike those that are closed source, have the advantage of having their source code open to the public, so it can be made available for free, can be copied and even changed by its users according to their needs (the most famous open-source software to date is the Linux operating system). There are numerous advantages of using open-source software: it is free; no licenses needed; its development is not limited to specific individuals, groups or organizations; they are not subject to regulations of specific organizations or companies; among others. One of the most important features is that everyone knows how the software works including its vulnerabilities, so the developer community is always working to fix them; despite being counterintuitive, having the software vulnerabilities exposed is something that helps keeping its quality and reliability, because it is known exactly what must be done to improve the software.

For AI systems as complex as ChatGPT, being able to inspect how the software was coded by its programmers is crucial. Most likely, in the not-so-distant future, software embedded with AI will start making an increasing number of decisions for human beings due to the vast amount of information and variables that exists from different and unusual application scenarios, including military defence. However, who would be so reckless to acquire a closed-source AI system (in which is not known how it was coded) for controlling nuclear missiles, for example? This is an extreme case, but it draws attention to the need of creating regulatory policies for the purpose of using this type of software, as proposed by the European Parliament in April 2021 in the document that was aimed to establish "harmonized rules" for the use of AI, that is, a law for the use of AI.

Undoubtedly, the use of AI is something that, despite astonishing us, will become part of our daily lives in ways that we do not even imagine yet. However, we need to stay alert to its indiscriminate use; without studies (whether from a technical, ethical or social point of view) it can generate more problems than solutions. We must not stop being optimistic about the possibilities of creating a tool that will help humanity, but we must also remain vigilant so that this innovation keeps on the track that we hope it to stay.

Eduardo Bertassi holds *a degree in electrical engineering from Escola Politécnica da USP and is a CEST-USP collaborator.*

Academic Coordinator: Edison Spina

This article results from the investigation and analysis of the author, not necessarily reflecting CEST's opinion.