Study explores the potential and shortcomings of ChatGPT in SPC, education and research

March 14 2023, by Ingrid Fadelli

A Representative Sample of Major Generative AI Developments from 2020 to Jan 2023

(a) A representative sample of LLM models released/published in the last three years.

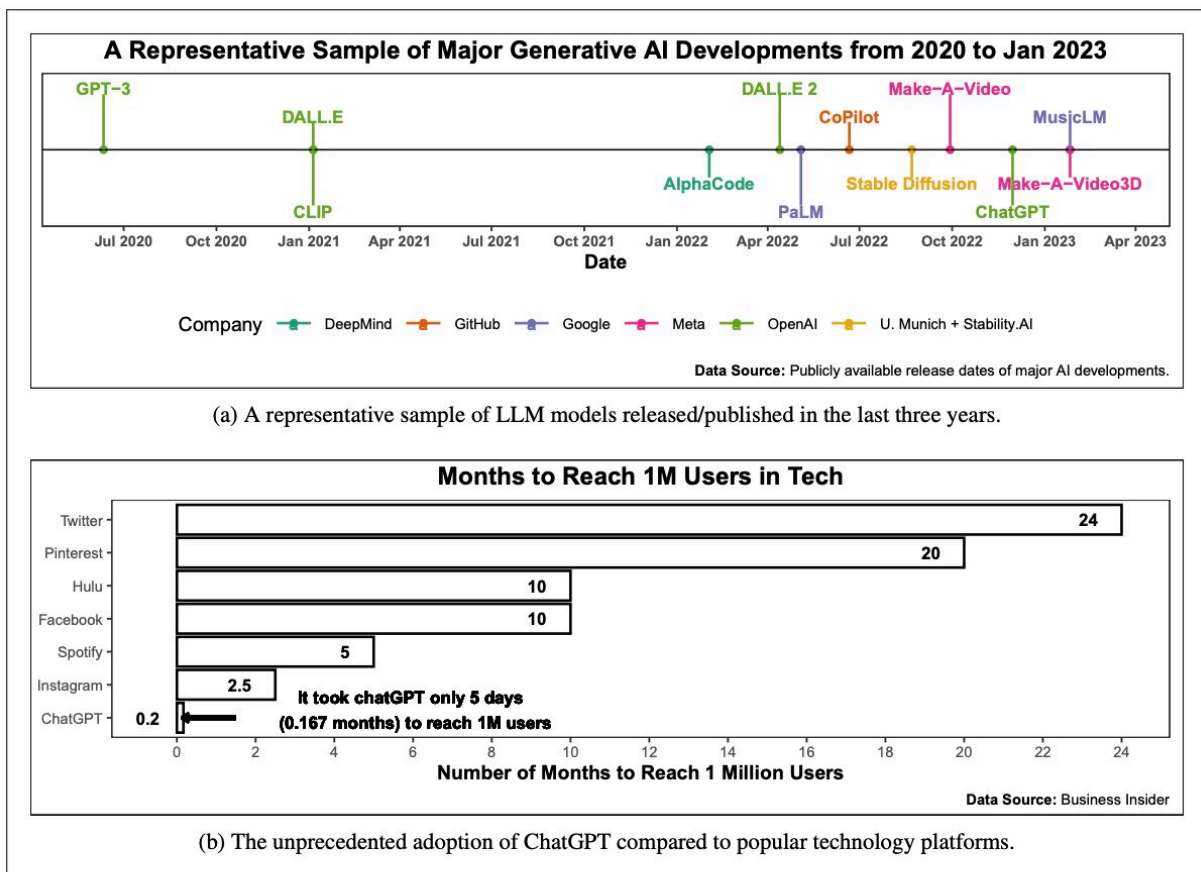(b) The unprecedented adoption of ChatGPT compared to popular technology platforms.

Figure 1: The rapid developments in generative AI and the associated hype in trying/adopting these tools by the general public.

Credit: Megahed et al

At the end of November 2022, the San Francisco-based company OpenAI launched its prototype of ChatGPT, an artificial intelligence (AI)-based chatbot that can answer a wide range of questions in short periods of time. Since then, users worldwide have been testing the chatbot and discussing its possible applications in different fields.

ChatGPT is based on a so-called large language model (LLM), a deep learning technique that employs multi-layered neural networks trained on a vast pool of texts. Over time, these models can learn to make predictions about how to compose sentences and answer specific language queries.

GPT-3, the model underpinning ChatGPT, is one of the most powerful LLMs worldwide, as it includes more than 175 billion parameters and can tackle a wide range of written tasks. For instance, the chatbot can translate and summarize written texts, compose basic poems or song lyrics and offer definitions for particular terms.

Researchers at Miami University, University of Dayton and Helmut Schmidt University in Hamburg have recently carried out a study assessing the potential value and limitations of ChatGPT in different fields, including education, research and statistical process control (SPC), which is the use of statistical tools to control a process or production method. Their paper, published on the pre-print *arXiv* server, suggests that while ChatGPT and other highly performing LLMs could sometimes be helpful in these settings, the answers its provide are not always reliable and thus they should still be validated using reliable sources.

"We explore ChatGPT's ability to provide code, explain basic concepts, and create knowledge related to SPC practice, learning, and research," Fadel Megahed Ying-Ju Chen and their colleagues wrote in their paper. "We ask, 'what can generative LLM-based AI tools do now to augment

the roles of SPC practitioners, educators, and researchers?' To make our task more tractable, we will primarily focus on evaluating the utility of ChatGPT (and its underlying GPT-3.5 engine) since it: (a) is the most well-known of these generative AI tools and (b) combines features of the generative chatbot with an underlying LLM that can generate both text and code. In our estimation, this expository assessment can provide a benchmark for future evaluations of the next generation of generative AI models."

To evaluate the potential of ChatGPT as a tool to assist professionals in different fields, Megahed, Chen and their colleagues asked the chatbot to answer different types of questions. Specifically, they asked it to provide code for a particular task, explain basic concepts and generate information related to each of the three fields they focused on.

The researchers then closely examined the code, responses and information generated by ChatGPT, to determine their accuracy and value in these different fields. Overall, they found that while the LLM-based chatbot could be useful, particularly as a tool for translating texts, brainstorming or as an assistant for skilled human programmers, the code and responses it generated alone could not be trusted to be functional, reliable and accurate.

"Our study indicates that the current version of ChatGPT performs well for structured tasks, such as translating code from one language to another and explaining well-known concepts, but struggles with more nuanced tasks, such as explaining less widely known terms and creating [code](#) from scratch," Fadel Megahed Ying-Ju Chen and their colleagues explained in their paper.

"We find that using new AI tools may help practitioners, educators, and researchers to be more efficient and productive. However, in their current stages of development, some results are misleading and wrong.

Overall, the use of generative AI models in SPC must be properly validated and used in conjunction with other methods to ensure accurate results."

In the future, the observations gathered by this team of researchers could guide SPC practitioners, educators and researchers, helping them to determine when LLMs like ChatGPT can be useful and in what cases trusting their outputs might be unwise. Megahed, Chen and their colleagues hope that this will promote LLM-fuelled innovation in their field, while reducing the occurrence of errors and the dissemination of unreliable information.

More information: Fadel M. Megahed et al, How Generative AI models such as ChatGPT can be (Mis)Used in SPC Practice, Education, and Research? An Exploratory Study, *arXiv* (2023). DOI: 10.48550/arxiv.2302.10916