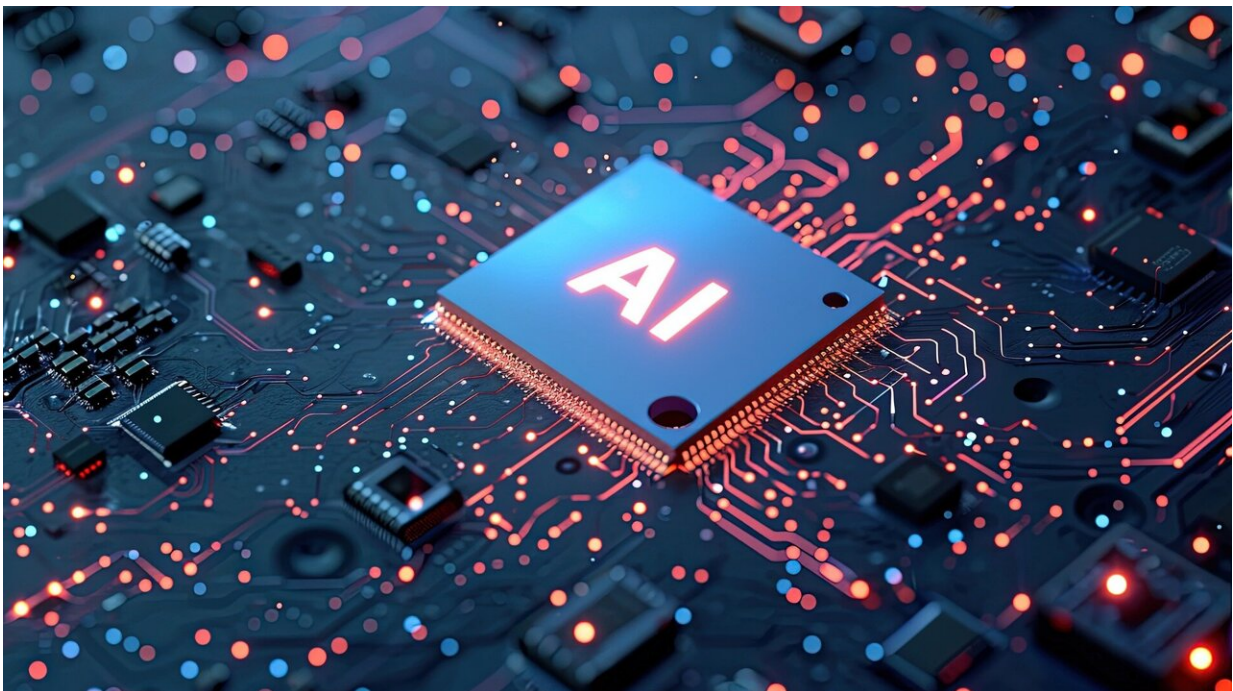


Are emergent abilities in large language models just in-context learning?

August 12 2024



Credit: Pixabay/CC0 Public Domain

According to a new study led by TU Darmstadt, AI models such as ChatGPT are apparently less capable of learning independently than previously assumed.

According to the study, there is no evidence that what are known as [large language models](#) (LLMs) are beginning to develop a general "intelligent"

behavior that would enable them to proceed in a planned or intuitive manner or to think in a complex way.

The study, which has been [published](#) on the *arXiv* preprint server, will be presented in August at the annual conference of the Association for Computational Linguistics (ACL) in Bangkok, the largest international conference on automatic language processing.

The [research](#) focuses on unforeseen and sudden leaps in the performance of language models, which are referred to as "emergent abilities." After the models were introduced, scientists found that they became more powerful with increasing size and the growing amount of data with which they were trained (scaling).

As the tools were scaled up, they were able to solve a larger number of language-based tasks—for example, recognizing [fake news](#) or drawing logical conclusions.

On the one hand, this raised hopes that further scaling would make the models even better. On the other hand, there was also concern that these abilities could become dangerous, as the LLMs could become independent and possibly escape human control. In response, AI laws were introduced worldwide, including in the European Union and the U.S..

However, the authors of the current study have now come to the conclusion that there is no evidence for the presumed development of differentiated thinking in the models. Instead, the LLMs acquired the superficial skill of following relatively simple instructions, as the researchers showed. The systems are still a long way from what humans are capable of.

The study was led by TU computer science professor Iryna Gurevych

and her colleague Dr. Harish Tayyar Madabushi from the University of Bath in the UK.

"However, our results do not mean that AI is not a threat at all," said Gurevych.

"Rather, we show that the purported emergence of complex thinking skills associated with specific threats is not supported by evidence and that we can control the learning process of LLMs very well after all. Future research should therefore focus on other risks posed by the models, such as their potential to be used to generate fake news."

And what do the results mean for users of AI systems such as ChatGPT? "It is probably a mistake to rely on an AI model to interpret and execute [complex tasks](#) without help," explains Gurevych, who heads the Ubiquitous Knowledge Processing (UKP) Lab at the Computer Science Department of TU Darmstadt.

"Instead, users should explicitly state what the systems should do and, if possible, give examples. The important thing is: The tendency of these models to produce plausible-sounding but false results—known as confabulation—is likely to persist, even if the quality of the models has improved dramatically in recent times."

More information: Sheng Lu et al, Are Emergent Abilities in Large Language Models just In-Context Learning?, *arXiv* (2023). [DOI: 10.48550/arxiv.2309.01809](https://doi.org/10.48550/arxiv.2309.01809)

Provided by Technische Universitat Darmstadt

Citation: Are emergent abilities in large language models just in-context learning? (2024, August

12) retrieved 12 August 2024 from <https://techxplore.com/news/2024-08-emergent-abilities-large-language-context.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.