Bot of Contention. ChatGPT, Success, and Trustworthiness

J. Lenhard, J. Pelloth, K. Zweig

Abstract for SAS 2023, Stuttgart

The current spearhead of AI, Large Language Models (LLM) like ChatGPT, have caused great astonishment about what they can achieve – and also spurred debate about what it cannot achieve. This contribution takes two steps toward analyzing the ongoing controversy and toward finding out what makes the promises of AI and, in particular, LLM-based bots a contentious issue.

In the first step, we look at two contrarian directions of analysis. One asks what ChatGPT and its likes *can* achieve and what we can learn from these achievements. An exemplary proponent of this direction is Steven Wolfram (2023). The second class of analyses turn into the opposite direction and ask what such bots *cannot* achieve. An instance is Christian Bermes and Andreas Dörpinghaus (2023), but also Wolfram offers some material.

Our second step connects this discussion with the issue of trust. Brian Cantwell Smith (2019) has argued about the promises and limitations of AI. Although he has written up his account before the arrival of LLMs, he seems to present an argument according to which ChatGPT cannot be trustworthy.

In the conclusion, we reflect on the reasons why trustworthiness is (or is not) out of reach for ChatGPT (and the wider LLM family).

References
Bermes and Dörpinghaus: Wer hat Angst vor ChatGPT? FAZ, 19.4.2023
Smith: The Promise of Artificial Intelligence. Reckoning and Judgement. The MIT Press, 2019.
Wolfram: What Is ChatGPT Doing … and Why Does It Work? (2023)