

ChatGPT and the concept of ‘truth’

Oct. 31, 2023

Peter Van Roy

Université catholique de Louvain

Ethical Forum, University Foundation

Discussion item for “ChatGPT and Co in higher education: to be cheered or feared ?” (Dec. 7, 2023)

What is ‘truth’ for a computer program?

Users of ChatGPT observe that it is difficult to use for tasks in which the result must respect truth. This is a consequence of the LLM technology (Large Language Model) on which ChatGPT is based, which does not contain any notion of ‘truth’. A LLM gives no semantics to text. It just combines and generates texts according to syntactic criteria of similarity. The lack of truth in the LLM approach cannot easily be fixed. During its training, the LLM would have to create a model of the world, because the truth of a statement can only be determined with respect to a model of the world that the statement applies to. To see whether a statement is true or false in the real world, a human will imagine a model of reality and in that model determine whether the statement holds or not. LLMs cannot do this.

Twofold danger

Therefore, for any task in which truth is important, a tool like ChatGPT must be used with skepticism. Nothing it generates can be assumed to be true. Every statement it outputs must be verified. This is obvious if we use ChatGPT for completely formal tasks, such as mathematics or computer programming. If we ask ChatGPT to prove a theorem or to write a computer program, the result cannot be assumed to be correct. It must be laboriously verified, step by step.

Nevertheless, ChatGPT may seem to be useful for disciplines where ambiguity is inherent. This happens in disciplines that study phenomena that are too complex to be formalized, such as in the humanities. For these disciplines, there is an even greater danger. The human reader of a generated text will project meaning onto the text, even if the text has no meaning. Again, a great deal of skepticism is in order: the human reader must carefully verify that the apparent meaning in the text is not imagined but is actually present. This is much more difficult than formal verification.

Conclusion

ChatGPT’s underlying technology, the LLM, contains no notion of truth, and truth does not magically appear during the training process. ChatGPT may give seductively intelligent answers, but their intelligence is only apparent. A truly intelligent text generator would need a notion of ‘truth’ and an internal model of the world in which to evaluate truth. Current LLMs do not have this and adding it does not seem to be easy. We observe that it has not yet been done up to the present day, despite a year of efforts from the AI community. ChatGPT has been extended in ad hoc ways (“hacked”) to make its obliviousness to truth less apparent, by selectively filtering out falsehoods as they are detected, but since this does not guarantee that future results will be true, it is not a definitive solution. We conclude that ChatGPT is inadequate for any task in which truth is important.

This article was written by a human being without AI assistance.

References

1. Ashley Belanger. *Will ChatGPT’s hallucinations be allowed to ruin your life?* arstechnica, Oct. 23, 2023. <https://arstechnica.com/tech-policy/2023/10/will-chatgpts-hallucinations-be-allowed-to-ruin-your-life/>
2. Benj Edwards. *EU votes to ban AI in biometric surveillance, require disclosure from AI systems.* arstechnica, June 15, 2023. <https://arstechnica.com/information-technology/2023/06/eu-votes-to-ban-ai-in-biometric-surveillance-require-disclosure-from-ai-systems/>
3. Jacob Roach. *‘I want to be human.’ My intense, unnerving chat with Microsoft’s AI chatbot.* digitaltrends, Feb. 17, 2023. <https://www.digitaltrends.com/computing/chatgpt-bing-hands-on/>
4. Alex Hern. *AI bot ChatGPT stuns academics with essay-writing skills and usability.* The Guardian, Dec. 4, 2022. <https://www.theguardian.com/technology/2022/dec/04/ai-bot-chatgpt-stuns-academics-with-essay-writing-skills-and-usability>
5. Stephen Marche. *The College Essay is Dead.* The Atlantic, Dec. 6, 2022. <https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/>