Computing
Reviews

Association for
Computing Machinery

ThinkLoud®

TODAY'S ISSUE   HOT TOPICS   SEARCH   BROWSE   RECOMMENDED   MY ACCOUNT   LOGIN

REVIEW   ?

Review

Search [   ] ▶

ChatGPT is bullshit

Hicks M., Humphries J., Slater J. Ethics and Information Technology 26 (2):1-10,2024.Type:Article

Date Reviewed: Feb 25 2025

[Full Text]

As people around the world understand how large language models (LLMs) behave, more and more of us wonder why these models "hallucinate"--and what can be done to reduce it. Hicks et al.'s provocatively named paper is an excellent primer for understanding how LLMs work and what to expect from them.

As humans, language is our main tool for maintaining relationships. Thus we are easily awed at the apparent ease with which ChatGPT--the first widely available LLM-based automated chatbot, and to this day probably the best known--simulates human-like understanding and helps us to easily carry out even daunting data aggregation tasks. When users ask ChaptGPT for an answer and the LLM gets part of the answer wrong, people often justify this by stating that it's just a hallucination. The authors invite us to switch from that characterization to a more correct one: LLMs are "bullshitting." This term is formally presented by Frankfurt [1]. To bullshit is not the same as to lie, because lying requires knowing (and wanting to conceal) the truth. A bullshitter does not necessarily know the truth; they just have to provide a compelling description, regardless of what is really aligned with truth.

After introducing Frankfurt's ideas, the authors explain the fundamental ideas behind LLM-based chatbots such as ChatGPT. Generative pre-trained transformers (GPTs) have as their only goal to produce human-like text, and this is carried out mainly by presenting output that matches the input's high-dimensional abstract vector representation, and probabilistically outputs the next token (word) iteratively with the text produced so far. Clearly, a GPT's task is not to seek truth or to convey useful information--they are built to provide a normal-seeming response to the prompts provided by the user. Core data is not queried to find optimal solutions to the user's requests, but are generated on the requested topic, attempting to mimic the style of the document set it was trained with.

Erroneous data emitted by a LLM is, thus, not comparable with what a person could hallucinate with, but appears because the model has no understanding of truth. In a way, this is very fitting with the current state of the world, a time often termed as the age of post-truth [2]. Requesting an LLM to provide truth in its answers is basically impossible, given the difference between intelligence and consciousness; following Harari's definitions [3], LLM systems (or any AI-based system) can be seen as intelligent as they have the ability to attain goals in various flexible ways, but they cannot be seen as conscious as they have no ability to experience subjectivity. The LLM is, by definition, bullshitting its way toward an answer: the goal is to provide an answer, not to interpret the world in a trustworthy way.

The authors end with a plea for literature on the topic to adopt the more correct "bullshit" term instead of the vacuous, anthropomorphizing "hallucination." Of course, being a word already loaded with a negative meaning, it is an unlikely request.

This is a great paper that mixes together computer science and philosophy, and can shed some light on a topic that many find hard to grasp.

Reviewer: Gunnar Wolf   Review #: CR147892

1) Frankfurt, H. G. *On bullshit*. Princeton University Press, Princeton, NJ, 2005.

2) Zoglauer, T. *Constructed truths: truth and knowledge in a post-truth world*. Springer, New York, NY, 2023.

3) Harari, Y. N. *Nexus: a brief history of information networks from the Stone Age to AI*. Random House, New York, NY, 2024.

Would you recommend this review?   ○ yes   ○ no   (Enter)

Other reviews under "**Ethics**":   **Date**

<u>Making ethical decisions</u>   Apr 1 2001
Kreie J., Cronan T. Communications of the ACM 43(12): 66-71, 2000. Type: Article

<u>Ethical and social issues in the information age</u>   May 6 2003
Kizza J., Springer-Verlag New York, Inc., Secaucus, NJ, 2002. 232, Type: Book (9780387954219)

<u>The error of futurism: prediction and computer ethics</u>   Apr 30 2004
Horner D. ACM SIGCAS Computers and Society 32(7): 42004. Type: Article

<u>*more...*</u>

✉ <u>E-Mail This</u>   ⎙ <u>Printer-Friendly</u>

REVIEWER'S AREA   **MASTHEAD**   **SUBSCRIBE**   **NEWS**   **TIPS**   **HELP**   **CONTACT US**