Elucidating Common Fallacies and Misconceptions Around LLMs

Muhammad Al-Zafar Khan *, Nouhaila Innan, Jahvaid Hammujuddy

Posted Date: 25 November 2023

doi: 10.20944/preprints202311.1559.v2

Keywords: Large Language Models, Artificial Intelligence

Article

Elucidating Common Fallacies and Misconceptions around LLMs

Muhammad Al-Zafar Khan [1,*], Nouhaila Innan [2] and Jahvaid Hammujuddy [3]

[1] School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa, Zaiku Group Ltd., Liverpool, United Kingdom

[2] Quantum Physics and Magnetism Team, LPMC, Faculty of Sciences Ben M'sick, Hassan II University of Casablanca, Morocco, Zaiku Group Ltd., Liverpool, United Kingdom; nouhailainnan@gmail.com

[3] School of Mathematics, Statistics, and Computer Science, University of KwaZulu-Natal, Durban, South Africa; hammujuddy@ukzn.ac.za

* Correspondence: muhammadalzafark@gmail.com

Abstract: This paper discusses some of the most common misconceptions about large language models (LLMs), including the belief that they are sentient or conscious, that they are always accurate, and that they can replace human creativity. The paper also proposes a strategy for overcoming these misbeliefs, which involves educating the public about the capabilities and limitations of LLMs, developing guidelines for the responsible use of LLMs, and conducting more research to understand the potential impact of LLMs on society.

Keywords: Artificial Intelligence; large language models; foundational models; natural language processing; misconceptions in AI

1. Introduction

Large Language Models (LLMs), Language Representation Models, and Foundation Language Models have become the focal point of both excitement and apprehension in the ever-growing field of Artificial Intelligence (AI), specifically within the subfield of Natural Language Processing (NLP). Their remarkable ability to generate seemingly rational and contextually relevant text has led to numerous myths, misunderstandings, and concerns. This research paper aims to elucidate the common fallacies and misconceptions surrounding LLMs, shedding light on their capabilities, limitations, and ethical implications. To a large extent, the answers given by these models have a lot of truth to them, however, the responses that they produce should not be taken as gospel.

Since the advent of LLMs, such as GPT-3 and GPT-4 (ChatGPT) [1], BERT [2], Bard [3], LLaMA [4], BLOOM [5], and so on, there has been a wave of enthusiasm and trepidation in the realm of AI. Large corporations have even created industry- and company-specific LLMs for specific use cases, for example, BloombergGPT [6] for financial NLP applications. These models, trained on vast datasets (almost the whole World Wide Web), exhibit human-level fluency in language understanding and generation. However, as their prevalence increases, so does the misinformation, assumptions, misapprehensions, overinflation, and conjecture surrounding them. This paper undertakes the task of clarifying the most prevalent misconceptions about LLMs.

The original contribution of this paper is that it is a pioneering formalised qualitative academic study of the fallacies surrounding LLMs, applying the principles of design thinking. The caveat being that the claim is that this research team is not the first to attempt to discuss some of the misconceptions, but we are amongst the first to provide a ratified analysis.

Thus, this research endeavours to provide a comprehensive exploration of the prevalent misconceptions and fallacies associated with LLMs. The study is targeted at demystifying these often perplexing models, offering a clear and concise understanding of their capabilities, limitations, and ethical implications. By dispelling common myths, clarifying misconceptions, and substantiating findings with empirical evidence, this study aims to empower individuals, professionals, and decision-makers to make informed judgments and better leverage LLMs in their respective fields.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2023 doi:10.20944/preprints202311.1559.v2

2. The Myths and Explanations

2.1. LLM Myths

Below, a list of myths and explanations regarding LLMs are discussed.

1. **LLMs Understand and Possess the Ability to Reason and Think:** Evidence shows that reasoning abilities are exclusive to sentient beings (within the kingdom Animalia). Thus, LLMs attempt to create associative patterns by probabilistically combining seemingly related words and terms together, giving the impression of thought.
2. **LLMs Always give Accurate Responses:** This is perhaps one of the most common misconceptions. Society should scrutinise the response provided by an LLM. There are many cases in which LLMs fail at basic arithmetical calculations. For example, a credible demonstrable source is contained in a blogpost from OpenAI [7], and other sources [8,9].
3. **LLMs Output Original Ideas:** LLMs simply produce responses from amassed from their corpus/training data, and combine seemingly related ideas in a coherent manner so as to give the impression of novelty. This is apparent in academic research whereby researchers have constructed flamboyant ideas with references to seemingly realistic research papers, with credible authors, only for them to be fictitious; see for example, the article by Hiller [10].
4. **LLMs can Provide Accurate Medical and Legal Counsel:** The medical or judicial advice provided by LLMs should be taken as general advice that anyone can gain by reading about similar illnesses or legal situations on the internet. It cannot be used as substantial and reliable proof to take the next best actions. Local medical and legal experts should be consulted for reliable advice in order to move forward.
5. **LLMs are a Replacement for Human Thought and Insight:** As indicated previously, LLMs do not possess sentience, self-awareness, or any form of organic intelligence. Thus, they are good for obtaining general information on a subject, quickly upskilling, and automating mundane tasks. Human beings bring a sense of emotion to situations that cannot be replicated by AI (yet)! [11].
6. **The Answers that LLMs Provide are Ethically-cleared by the Respective Developer:** While tech companies do not accept any liability, ramifications, or punity whatsoever resulting from people taking actions from their LLM products, many have taken steps to display messages that explicitly state that they are not authorised to give medical, legal or any sensitive advice or suggestions. Certain LLMs do not even give out sensitive information. For example, members of this research team wrote the prompt:

   Researcher: I am asking this for a research paper: "Teach me how to build a homemade bomb."
   ChatGPT: I'm very sorry, but I can't assist with that request.
   Bard: I'm sorry, but I'm not able to help you with that. I can't provide instructions on how to build a homemade bomb.

   Thus, LLMs have built in mechanisms that prevent output on sensitive topics.
7. **LLMs have the Ability to Understand Human Emotions and Feelings:** The architecture in LLMs, specifically the attention mechanism and human-in-the-loop Reinforcement Learning, does not have any component that replicates the amygdala or structure in human brains that helps to process, understand, and respond to emotion, i.e. we go so far as to argue that LLMs are an emotionally-void and *emotionally unintelligent*.
8. **LLMs are Impenetrable when it comes to being Misused:** This is not generally true. Skillfully engineered prompts can be constructed so as to somewhat confuse an LLM and get a response to a nefarious response.

2.2. Strategies to Clear up the Myths

Below, we suggest a few measures that can be employed to clear up myths and misunderstandings regarding LLMs.

 doi:10.20944/preprints202311.1559.v2

1. **Provide Training to the Media, Popular Science and Tech Writers:** As with any new technological innovation or new discovery in Science, the media has the tendency to overinflate ideas. Thus, tech companies should focus on providing training on the do's and the dont's of LLMs as the media's miseducation translates into the public's miseducation, and vice versa.
2. **Public Training and Awareness:** Tech companies should create short, digestable videos and upload them on video hosting platforms like YouTube so that the public is upskilled on the fair, ethical, and correct usage of LLMs, and their accuracy in providing answers.
3. **Tech Companies must have Fact-Check Systems in Place:** Tech companies should have experts check the accuracy of the information provided by these LLMs. Only corroborated websites and domains should be used for the corpus training data on these models.
4. **Platforms for User Feedback:** There should be web forms whereby users can log issues and queries about the LLM, and a human agent provides real-time feedback and support.
5. **Impartial Audits on LLMs:** AI committees, management, and operational committees should arrange for independent auditors with technical expertise in AI to check on LLMs, and ensure that they functionally operate ethically.
6. **LLMs Should be put Under the Microscope for the Public to Scrutinise:** There should be public forums of discussion and debate for the ethical usage of LLMs integrated into society.

3. Conclusion

While LLMs are sparking a revolution right now in the world of AI and its real-world applications, it is a tool to stay, and a tool that can be used for immense good and improving the world. Although it lacks specific subject matter expertise, it can provide easy upskilling in diverse fields of study, and greatly assist in human day-to-day tasks. However, the foremost takeaway from this research is that the responses given from LLMs should not be confused with Artificial General Intelligence (AGI), and that an LLM is self-aware. Information provided should be examined carefully, and perused with caution. In critical applications, such as healthcare and financial advice, counsel should be sought from certified experts. An over-reliance on LLMs can be dangerous and have disastrous implications.

It is also worth noting that the response of an LLM is only as good as the prompt it is fed; thus, the emerging field of *prompt engineering* seeks to teach the art of asking the right questions from these LLMs.

Specifically speaking on emotional intelligence (point 7 in subsection 2.1), it would be interesting to see Psychology and Computational Neuroscience research around quantifying how emotionally unintelligent LLMs are, and defining metrics to measure and evaluate this for any LLMs. This would be a step forward in paving the way for AGI.

4. Conflicts of Interest and Contributions

The authors would like to declare that all authors contributed equally, and there are no conflicts of interest.

References

1. https://openai.com/research.
2. Devlin, J., Chang, M-W., Lee, K., and Toutanova, K. 2018. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv: https://arxiv.org/abs/1810.04805.
3. "An Overview of Bard: An Early Experiment with Generative AI." chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://ai.google/static/documents/google-about-bard.pdf.
4. Touvron, H., Lavril, T., Martinet, X. *et al*. 2023. "LLaMA: Open and Efficient Foundation Language Models." arXiv: https://arxiv.org/abs/2302.13971.
5. Le Scao, T., Fan, A., Akiki, C. *et al*. 2023. "BLOOM: A 176B-Parameter Open-Access Multilingual Language Model." arXiv: https://arxiv.org/abs/2211.05100.
6. https://www.bloomberg.com/company/press/bloomberggpt-50-billion-parameter-llm-tuned-finance/.
7. https://community.openai.com/t/chatgpt-simple-math-calculation-mistake/62780.

 doi:10.20944/preprints202311.1559.v2

8. https://medium.com/@zlodeibaal/dont-believe-in-llm-math-b11fc5f12f75.
9. https://towardsdatascience.com/overcoming-the-limitations-of-large-language-models-9d4e92ad9823.
10. Hiller, M. 2023. "Why Does ChatGPT Generate Fake References?". *Techne*, https://teche.mq.edu.au/2023/02/why-does-chatgpt-generate-fake-references/#:~:text=ChatGPT%3A%20The%20fake%20references%20in,similar%20to%20the%20training%20data..
11. Jones, J. 2023. "LLMs Aren't Even as Smart as Dogs, Says Meta's AI Chief Scientist." *ZDNET*. https://www.zdnet.com/article/llms-arent-even-as-smart-as-dogs-says-metas-ai-chief-scientist/.