Future of Interacting with Computers and Large Language Models

Migul Jain, Vels Vidyashram, July 2023

Introduction

Over the past several years, the technology sector has been buzzing with enthusiasm for large language models (LLMs) and the nearly limitless tools that can be powered by artificial intelligence. Emerging around 2018, LLMs have grown exponentially and tools such as GPT-4 contain over 100 trillion parameters that can process text and images that fuel AI applications (Birhane, Kasirzadeh, Leslie, and Watcher, 2023). The future of LLMs is nearly mind-boggling and the technology has the potential to change human existence and interaction as we know it. While the ethical and moral question regarding LLMs are many, the technology promises to be the future of computing if the few hurdles to its use can be worked out.

Current State of Large Language Models

The current state of LLMs arrived from a technology introduced in 2017 by Google called the Transformer model architecture (Harrer, 2023). Transformers are programs that learn contextual information about data that is presented in a sequence, data such as words, audio signals, or even videos. Transforms can store trillions of parameters, algorithms, and data that take the information presented and contextualize it and turn it into other data modalities such as text-to-speech, text to image, or text to video schemes (p.2). The models that serve as the basis for contextualizing the data are called large language models.

To understand how the LLMs work, one can look at the most well-known version of technology, which is the phenomenally popular ChatGPT. Using ChatGPT begins with entering a prompt in text. ChatGPT then scraps data from the internet and uses that data to find statistical correlations based on the position of words and phrases in the prompt. ChatGPT then returns the model response. This response is fine-tuned by the user by giving another prompt in which ChatGPT can then narrow its response even further to exact the results. By narrowing the results, the ChatGPT is also "learning" better correlations of data and adding them to its trillions of parameters. This ability to "learn" is essentially the adding of statistical information to a database, yet it is fueling the development of artificial intelligence applications the world over and stirring controversy along with it (Harrer, 2023).

The Future of Large Language Models

If one stops to imagine the myriad of possibilities with programs such as ChatGPT, the uses seem compellingly opportunistic. Perhaps medical bills could be reduced by not needing to see a doctor, just input symptoms into a medical model and receive a diagnosis. Programmers currently use ChatGPT with

limited success to assist with creating new programming solutions (Perkel, 2023). Scientists With only having a solid five years in development in their current form, large language models have a tremendous amount of refinement ahead of them for them to become trustworthy sources of information. The hurdles that need to be overcome are primarily three limitations to their usability:

1. The use of unreliable Internet data
2. No self-checking method for reliability of data
3. Algorithms that can be tweaked for inaccuracy

For ChatGPT in particular, which solely uses the Internet for its answers, it doesn't take much imagination to see the problems in this data field. In an unprecedented era of fake news and misinformation, ChatGPT has no way of discerning valid information from invalid information. Furthermore, algorithms can be maliciously tweaked by the user to contain misinformation, a serious flaw in its core design. What the best and the brightest minds in the technology sector are working on currently is the crux of the problem with LLMs, which is ensuring that the LLM is "behaving in a way that is aligned with human values and ethical principles" (Angelis, et al, 2023). It was Freud, in *Civilization And Its Discontents,* who coined the term "prosthetic God" to describe what man has become through his creation of tools. Technology has indeed given marvelous powers over the material aspects of our lives within the last several decades. Advances such as cloning, stem-cell research, organ transplantation, and mapping the human genome are now giving humans tantalizing hints with respect to its potential to both eliminate enormous problems and to increase human the human lifespan. To look at the ethical implications of LLMs those in charge of the technology have to decide which ethical system to use. I would suggest that utilitarianism would make a good choice here because it is, *de facto*, the way in which most arguments are framed today. John Mill's statement regarding utilitarianism, "…actions are right in proportion as they tend to promote happiness… (407)" may not be sound in terms of ethical theory but, it is the doctrine which, most likely, will determine the future of LLMs. This is because utilitarianism has one great advantage: it is how most people make up their minds on ethical issues.

To determine whether LLMs will offer "the greatest good for the greatest number" means that one must have an accurate assessment of the risks involved. ChatGPT has already garnered many articles in the popular press of frightening future dystopias (Pompeo, 2023); how realistic such visions are is very difficult to determine. A utilitarian standard requires that one make a systematic, thorough, and accurate assessment of the risks and possible rewards offered by the new technologies. This should be done, I submit, before further research is allowed to proceed. I also submit that to date there has been a lot of dithering and bickering in the scholarly

communities, but that a great deal of what has emerged has been polemical in nature, a function not of rational study, but of ethical and religious bias. An authoritative body composed of "the best and brightest" needs to be empaneled. This group should be given the resources to make a thorough investigation of what the probable gains and risks are. The conclusions that this group might make would not end the controversy, but they might shed some light on issues now enshrouded in rhetoric, hyperbole, and disinformation.

The ethical and moral dichotomies associated with LLMs have been widely explored by several researchers in the context of Artificial Intelligence models, or AI, including the renowned Joseph Weizenbaum. Weizenbaum, a professor at MIT in the late 1960s, developed an AI program called ELIZA. ELIZA was a highly advanced AI software program that could engage in conversation with an individual. The program was so effective that when compared in a phone test with three humans, scientists could not tell which voice was that of the computer and which were those of the human participants. When developing this software, Weizenbaum outlined several constraints, which he believed should be placed on all AI programs developed. Among these the most pertinent are listed below (Weizenbaum, 1976):

1. Computers should not attempt to mimic emotions such as respect, compassion and love.
2. Scientists must not give into the idea that humans function exactly like machines.

Although Weizenbaum's ideas were not well received by an audience who believe that AI held the potential to solve the problems of the world, much of what Weizenbaum espoused was taken with some degree of analysis by scholars and academics. All over the world, scientists, researchers, and philosophers began to examine the impact that creating intelligent beings would have on society. Undeniably, the technological benefit of such machines could be readily seen; but what would the outcome be 50 or 100 years from the time this technology was first utilized in mainstream culture? Would society create beings that were so advanced that they would eventually take over the word—much like in the movie *The Terminator*? Or could science and man come together to find a way to utilize this technology without exploiting it? The time to answer these questions has come.

Conclusion

While the ethical considerations are being debated, the technology will undoubtedly be honed to include solutions to the issues and concerns. For example, currently models are being developed that use data strictly from certain vertical applications that can be monitored and mediated. Fields such as the medical and legal fields are focusing on the technology through training it with only legal and medical data with monitored input (Scholtes and Vance, 2023). Reinforcement learning is

also used to better LLMs. Reinforcement learning is better known as active learning, something that is done with strict adherence to human values and human-like dialogue to create the best platform for machine learning. Thus, one can witness that the solutions to the problems introduced by LLMs are being worked through and the future for the technology is stunning and bright.

References:

Birhane, A., Kasirzadeh, A., Leslie, D. *et al.* Science in the age of large language models. *Nat Rev Phys* **5**, 277–280 (2023). https://doi.org/10.1038/s42254-023-00581-4

De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., & Rizzo, C. (2023). ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Frontiers in public health*, *11*, 1166120.https://doi.org/10.3389/fpubh.2023.1166120

Freud, S. (1989). Civilization and Its Discontents. New York: W.W. Norton & Company.

Herrer, S. (2023). Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. eBioMedicine, 23. doi: doi.org/10.1016/j.ebiom.2023.104512

Mill, J. (1973). Utilitarianism. In *The Utilitarians*. Garden City, New York: Doubleday.

Perkel, J.M (2023). Six tips for better coding with ChatGPT, Nature, 618, 422-423. *doi:* https://doi.org/10.1038/d41586-023-01833-0

Pompeo, J. (2023). ChatGPT's mind-boggling, possibly dystopian impact on the media world. Vanity Fair, January, Available from https://www.vanityfair.com/news/2023/01/chatgpt-journalism-ai-media

Scholtes, J. & Vance, G. (2023). Future of generative large language models and potential applications in LegalTech. IPRO, 14 April, Available from IPRO.com

Weizenbaum, J. (1976). *Computer Power andHuman Reason: From Judgment to Calculation*. W.H. Freeman and Company.