This paper explores the question "Can Computers Think?" by examining historical perspectives, discussing Large Language Models (LLMs) based on the Transformer Architecture, and considering future possibilities. This question holds both theoretical and practical interest in computer science and carries significant ethical implications.

Historical Perspectives

In 1950, Alan Turing proposed the famous Turing Test to determine whether machines could exhibit intelligent behavior indistinguishable from that of humans [1]. The test involves an evaluator communicating through text with two unseen agents, one human and one machine. If the evaluator cannot reliably distinguish between the two, the machine is considered to have "intelligence." This test remains a foundational benchmark for the question of this paper.

Philosophical counter arguments often focus on consciousness and intentionality. John Searle's Chinese Room argument [2] is a notable critique, asserting that even if a machine convincingly simulates language understanding, it does not genuinely understand the language. Instead, it follows a rulebook (set of instructions) to produce appropriate responses. Searle argues that genuine thinking requires semantic understanding and consciousness.

Large Language Models (LLMs)

Recent advancements in AI, particularly LLMs with the Transformer architecture and the attention mechanism [3], have been significant, particularly in the past years. These models are trained to predict the next token with the highest probability in a given context window, the attention mechanism allowing the model to weigh the importance of different tokens in the input sequence, regardless of their position. But, despite the complex algorithm behind, we can abstract these tools to the idea of them being advanced compression tools [4]. They compress vast amounts of internet data into a set of parameters, significantly reducing the original data size from zettabytes to a more manageable form [5]. This process can be linked to Searle’s argument: LLMs simulate understanding by following a "rulebook" (the instructions to decompress the data) during the model inference phase, navigating through their “compressed”, parameter-based representation of the data and firing the neural network.

Passing through life, all humans can recall defining moments in it. Most of these moments, especially ones caused by posttraumatic growth, were also ones of great teaching value, on which we have built entire traits of our personality and ideologies. On the other hand, an LLM as mentioned before only functions if it has been fed a large part of the internet. Even though it can pinpoint some information as having more weight than other, it cannot feel emotions. The only reason why it recalls such information as more important is that, in the training phase, it led to the minimization of the used loss function. They cannot experience epiphanies from individual events, as they are a product of optimizing over the whole training data. With the ideas presented above we might argue that in some ways, machines cannot think.

Future Possibilities

Efforts to develop artificial general intelligence (AGI), aiming to create machines with cognitive abilities surpassing human intelligence, are ongoing [6]. Additionally, research into the role of quantum mechanics in the brain may uncover new aspects of consciousness that could inform the development of truly thinking machines [7]. Roger Penrose has suggested that consciousness cannot be modeled by a conventional Turing machine, which includes digital computers [8]. This also applies to modern neural network architectures, including Transformers, as they are Turing Complete [9].

Conclusion

While computers have made significant strides in mimicking aspects of human language and vision, they lack genuine perception akin to human experience. Turing cited Professor Jefferson's Lister Oration in 1949, stating that a computer must not only think but also feel itself thinking to be truly considered a thinking entity [1]. Even though unfortunately the only way to test this would be to be the machine itself, it provides an abstraction to the idea of consciousness and intentionality. Current AI systems excel at information processing and retrieval tasks but lack the conscious experience and deep semantic understanding characteristic of human thought.

Future advancements in AI and physics may provide new insights, akin to the groundbreaking "Attention is All You Need" paper [10] that introduced Transformers. However, the question remains open. In the case of LLMs based on the Transformer architecture, a plateau may soon be reached due to the finite nature of available data [11].

The state of the art of current AI models can predict the next most likely token but struggle to generate genuinely new ideas [12]. Additionally, they are constrained by the norms and conventions learned during training [13], lacking the benefit of not knowing what is impossible [14]: they have the downside of accessing more information than any human ever could.

References

1. The Turing Test (Stanford Encyclopedia of Philosophy). (2021, October 4). https://plato.stanford.edu/entries/turing-test/
2. The Chinese Room Argument (Stanford Encyclopedia of Philosophy). (2020, February 20). https://plato.stanford.edu/entries/chinese-room/
3. Toews, R. (2024, February 20). Transformers revolutionized AI. What will replace them? Forbes. https://www.forbes.com/sites/robtoews/2023/09/03/transformers-revolutionized-ai-what-will-replace-them/
4. Deletang, G., Ruoss, A., Duquenne, P., Catt, E., Genewein, T., Mattern, C., Grau-Moya, J., Wenliang, L. K., Aitchison, M., Orseau, L., Hutter, M., & Veness, J. (n.d.). Language modeling is compression. OpenReview. https://openreview.net/forum?id=jznbgiynus
5. Wikipedia contributors. (2024, May 24). GPT-4. Wikipedia. https://en.wikipedia.org/wiki/GPT-4
6. OpenAI. (n.d.). Planning for AGI and beyond. Retrieved April 20, 2024, from https://openai.com/index/planning-for-agi-and-beyond/

7. Schwartz, J. M., Stapp, H. P., & Beauregard, M. (2005). Quantum physics in neuroscience and psychology: a neurophysical model of mind-brain interaction. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1458), 1309-1327. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1569494/
8. Penrose, R. (1989). The emperor's new mind: Concerning computers, minds, and the laws of physics. Oxford University Press.
9. Pérez, J., Marinković, J., & Barceló, P. (2019, January 10). On the Turing Completeness of Modern Neural Network Architectures. arXiv.org. https://arxiv.org/abs/1901.03429
10. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017, June 12). Attention is all you need. arXiv.org. https://arxiv.org/abs/1706.03762
11. Microland. (2023, October 11). When Generative AI begins to plateau, what do we do? Medium. https://microland.medium.com/when-generative-ai-begins-to-plateau-what-do-we-do-fe2944027c70
12. Franceschelli, G., & Musolesi, M. (2023, March 27). On the Creativity of Large Language Models. arXiv.org. https://arxiv.org/abs/2304.00008
13. Can computers surprise us? (n.d.). https://www.adalovelaceinstitute.org/blog/can-computers-surprise-us/
14. Why are children so much more creative than adults? (2013, December 14). Destination Innovation. https://www.destination-innovation.com/why-are-children-so-much-more-creative-than-adults/