Can Computers Think?

David Nistor

Introduction

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 [Sta21]. 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 counterarguments often focus on consciousness and intentionality. John Searle's Chinese Room argument [Sta20] 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.

The presented perspectives highlight two contrasting historical views on whether computers can think. On one side, Alan Turing, while acknowledging the complexities and limits of this question [Sta21], proposes the Turing Test as a practical benchmark for determining machine intelligence. On the other hand, John Searle's Chinese Room argument delves into a deeper philosophical critique. While Turing focuses on behavioral indistinguishability in human communication as a measure of intelligence, Searle emphasizes the necessity of semantic understanding and consciousness, arguing that machines, even if they simulate understanding, do not truly "think" in the human sense.

Large Language Models (LLMs)

Recent advancements in AI, particularly LLMs with the Transformer architecture and the attention mechanism [Toe24], 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, they still run on computers. We can abstract these tools to the idea of them being advanced compression tools [Del+nd]. 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 [Wik24]. 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 others, 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. This emphasizes a critical difference between human cognition and machine learning, showcasing the limitations of current AI models. 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 [Opend]. 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 [SSB05]. Roger Penrose has suggested that consciousness cannot be modeled by a conventional Turing machine, which includes digital computers [Pen89]. This also applies to modern neural network architectures, including Transformers, as they are Turing Complete [PMB19].

Another promising area is the development of neuromorphic computing, out of which the Spiking Neural Network seems the most promising. It mimics the neural structure and functioning of the human brain with dedicated hardware and software, potentially leading to more human-like processing and understanding. Neuromorphic hardware is designed to emulate the synaptic connections found in biological brains. Spiking neurons measure and encode analog signal changes, not encoding data through a binary system as most computers, enabling more efficient and adaptable learning processes [BL23]. This approach might significantly advance our understanding of machine cognition, paving the way for machines that not only process data more naturally but also develop a form of understanding closer to human cognition. As research and technology progress, neuromorphic computing could bridge the gap between current AI capabilities and the goal of genuine machine thinking, potentially even modeling the entire human brain.

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 [Sta21]. 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 [Vas+17] 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 [Mic23].

The state of the art of current AI models can predict the next most likely token but struggle to generate genuinely new ideas [FM23]. Additionally, they are constrained by the norms and conventions learned during training[Com], which limits their creativity and ability to think outside the box. Unlike humans, who can sometimes innovate by not knowing what is impossible [Cre], these models are restricted by the vast amount of information they have been trained on, thus what gives these models their "intelligence" also limits it.

References

[BL23] N. Barney and B. Lutkevich. *neuromorphic computing.* [Online]. Available: `https://www.techtarget.com/searchenterpriseai/definition/neuromorphic-computing`. 2023.

[Com] *Can computers surprise us?* [Online]. Available: `https://www.adalovelaceinstitute.org/blog/can-computers-surprise-us/`. n.d.

[Cre] *Why are children so much more creative than adults?* [Online]. Available: https://www.destination-innovation.com/why-are-children-so-much-more-creative-than-adults/. 2013.

[Del+nd] G. Deletang et al. *Language modeling is compression.* [Online]. Available: https://openreview.net/forum?id=jznbgiynus. n.d.

[FM23] G. Franceschelli and M. Musolesi. *On the Creativity of Large Language Models.* [Online]. Available: https://arxiv.org/abs/2304.00008. 2023.

[Mic23] Microland. *When Generative AI begins to plateau, what do we do?* [Online]. Available: https://microland.medium.com/when-generative-ai-begins-to-plateau-what-do-we-do-fe2944027c70. 2023.

[Opend] OpenAI. *Planning for AGI and beyond.* [Online]. Available: https://openai.com/index/planning-for-agi-and-beyond/. n.d.

[Pen89] R. Penrose. *The emperor's new mind: Concerning computers, minds, and the laws of physics.* Oxford University Press, 1989.

[PMB19] J. Pérez, J. Marinković, and P. Barceló. *On the Turing Completeness of Modern Neural Network Architectures.* [Online]. Available: https://arxiv.org/abs/1901.03429. 2019.

[SSB05] J. M. Schwartz, H. P. Stapp, and M. Beauregard. "Quantum physics in neuroscience and psychology: a neurophysical model of mind-brain interaction". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 360.1458 (2005). [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1569494/, pp. 1309–1327.

[Sta20] Stanford Encyclopedia of Philosophy. *The Chinese Room Argument.* [Online]. Available: https://plato.stanford.edu/entries/chinese-room/. 2020.

[Sta21] Stanford Encyclopedia of Philosophy. *The Turing Test.* [Online]. Available: https://plato.stanford.edu/entries/turing-test/. 2021.

[Toe24] R. Toews. "Transformers revolutionized AI. What will replace them?" In: *Forbes* (2024). [Online]. Available: https://www.forbes.com/sites/robtoews/2023/09/03/transformers-revolutionized-ai-what-will-replace-them/.

[Vas+17] A. Vaswani et al. *Attention is all you need.* [Online]. Available: https://arxiv.org/abs/1706.03762. 2017.

[Wik24] Wikipedia contributors. *GPT-4.* [Online]. Available: https://en.wikipedia.org/wiki/GPT-4. 2024.