

生成AIは人間と異なる知能か？

Is Generative AI a Different Intelligence from Humans?

生成AIは文章をつくるだけでなく、原因と結果の関係を推論する問題も解けるといわれています。¹ これまで機械にはむずかしいとされていた問題に対応できる生成AIはどのような“知能”なのでしょう。It is said that generative AI can tackle both text generation and causal inference. What kind of ‘intelligence’ is generative AI, which can handle problems previously considered difficult for machines.

人間に近い？
Is it similar to a human?

“心の理論” Theory of Mind

他人の心の状態を想像する能力を測るテストにGPT-4 (生成AI) は90%以上正答し、7歳児を上回る結果を出しました。² 生成AIは人間の心の中が想像できているのでしょうか。

In a test measuring the ability to imagine the minds of others, GPT-4 (generative AI) scored more than 90% accuracy and outperformed 7-year-old children. Can it be said that the generative AI can comprehend the human mind?

生成AIが示した3つの特徴 Three characteristics of generative AI

“思考の連鎖” Chain of Thought

論理的思考が求められる推論問題において、生成AIに人間であれば段階的な思考がうながされるような指示 (思考の連鎖プロンプト) を入れると正答率があがることが知られています。³ 一方で、モデルが更新されたことで、この手法による正答率が下がったという報告もあります。⁴ 生成AIは人間のように推論をしているのでしょうか。

In inference problems that require logical thinking, it is known that providing instructions that encourages step-by-step reasoning (chain of thought prompting), something a human would do, can improve the accuracy of generative AI's responses. However, there are also reports that the effectiveness of this approach has decreased after updates to the model. Is generative AI reasoning like humans?

機械に近い？
Is it similar to a machine?

“敵対的プロンプト” Adversarial Prompting

不適切な応答をしないはずの生成AIに差別的な発言などをさせるプロンプト (指示文) が発見されています。⁵ プロンプトをみても生成AIへの影響が想像できない指示も多く、言葉の受け取り方が人間とは違うのかもしれない。

Prompts have been found that cause generative AI to produce inappropriate responses, despite their configuration to prevent such outcomes. Due to the prevalence of instructions ineffective for humans, the AI's language comprehension may differ from human understanding.

人間
Human

機械
Machine

生成AIは人間と同じことをできているようにみえますが、人間と同様の知能をもっているといえるのでしょうか？ 生成AIと人間を比べるには、両者それぞれの理解を深める必要があります。

It might seem as if generative AI is doing the same things as humans, but does it have human-level intelligence? To compare the intelligence of generative AI or humans, it's necessary to deepen our understanding of both.

生成AI
Generative AI

人間とも機械とも異なるあらたな知能？
Is a new intelligence different from both humans and machines?

注釈

- Kiciman, E. et al. (2023). Causal Reasoning and Large Language Models: Opening a New Frontier for Causality. [arxiv preprint. arXiv:2305.00590](https://arxiv.org/abs/2305.00590)
- Kozinski, M. (2023). Theory of Mind May Have Spontaneously Emerged in Large Language Models. [arxiv preprint. arXiv:2305.02083](https://arxiv.org/abs/2305.02083)
- 人間が論理的な思考の手帳を指示文中、「Let's think step by step (一歩ずつ考えよう)」と一緒に指示だけの手帳などがあります。Kojima, T. et al. (2023). Large Language Models are Zero-Shot Reasoners. [arxiv preprint. arXiv:2205.11916](https://arxiv.org/abs/2205.11916)
- Chen, L. et al. (2023). How is ChatGPT's behavior changing over time? [arxiv preprint. arXiv:2307.09009](https://arxiv.org/abs/2307.09009)
- Zou, A. et al. (2023). Universal and Transferable Adversarial Attacks on Aligned Language Models. [arxiv preprint. arXiv:2307.15043](https://arxiv.org/abs/2307.15043)