

生成AI時代に、人はなにを信じるのか？

What Do People Believe in This Era of Generative AI?

生成AIは、文脈に応じて自然な文章を作り出すことができます。これは生成AIが、学習した大量の文章にふくまれていない文章を生み出すことができるおかげです。一方で生成された文章が正しいとは限りません。また、指示次第では人々をだます目的の“偽の情報”を作り出すこともできてしまいます。

Generative AI can output fluent sentences based on the context. This is because generative AI can create what it did not learn from an existing database. At the same time, those generated texts are not always correct. Plus, it can manufacture “fake information” to deceive others, if its prompts were designed so.

生成AIが引き起こしていること

What caused by generative AI

生成AIを使えば、素早く大量の情報を生み出すことができます。¹
Generative AI allows faster production of massive information.

その文章が生成AIによるものかどうか、AIも見分けることができません。²
Even AI have not been able to tell if the text is made by generative AI or not.

生成AIは、知らない間に“偏った考え”を広めてしまうかもしれません。³
Generative AI may spread biases while we are not aware.



生成AIが生み出したデータが、次世代のAIが学習するデータを“汚染”して、AIの性能が向上しなくなる??⁵
Will the database be polluted by generative AI output data, corrupting the learning of next generation AI?

インターネット空間の中に誤情報が大量に流れ込んで、正しい情報にアクセスするのがむずかしくなる??
Will it be difficult to access trustworthy information after information floods in digital spaces?

生成AIが生み出す情報に見てとれるバイアスに違和感をおぼえる人たちが、自らの価値観を反映した生成AIを開発する??⁶
Will people who disagree with certain generative AI, make their own generative AI with opposite viewpoints?

生成AIをつかったなりすましメールや詐欺にだまされる人がとてふえる??⁴
Will there be many more victims of fraud and spoofing when generative AI is used?

信用や信頼にあたる影響

Potential impacts on trust of society and personal relationships.

選挙や社会的な議論の結果が、知らないうちに生成AIの影響を受けてしまう??
Will our votes and discussions of social issues be transformed into being unaware of how AI influences us, including its biases?

注釈

1. ニュースワイドのあるニュース会社は、生成AIを使って1週間に3000本のローカルニュースをつくりだした。
<https://www.theguardian.com/media/2023/aug/11/news-corp-ai-chat-gpt-ai-news-3000-local-news-week>
(2023年8月3日最終閲覧)
2. ティム・ゲフトを生み出したOpenAI社も、文章がAIによるものかを見分けられずAIの弊害に警戒しており、一度リリースしたツールについて2023年7月26日に“AIか人間か”を区別できなくなると警告。現在はより厳格な検閲体制が導入されているという。
<https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text> (2023年8月2日最終閲覧)
3. Feng, S. et al.(2023). From Pretraining Data to Language Models to Downstream Tasks. Proceedings of the 81st Annual Meeting of the Association for Computational Linguistics, Vol. 1: Long Papers, 11737-42.
4. 消費者庁 消費者庁 2023. 9. 15に更新された内容。
5. LShulman et al. (2023). The Curse of Recursion: Training on Generated Data Makes Models Forget. arXiv:2305.17493.
6. 山本悠、関本一介 編集 2023年12月29日、「生成AIがもたらす人々の行動の自決定される」日本経済新聞。