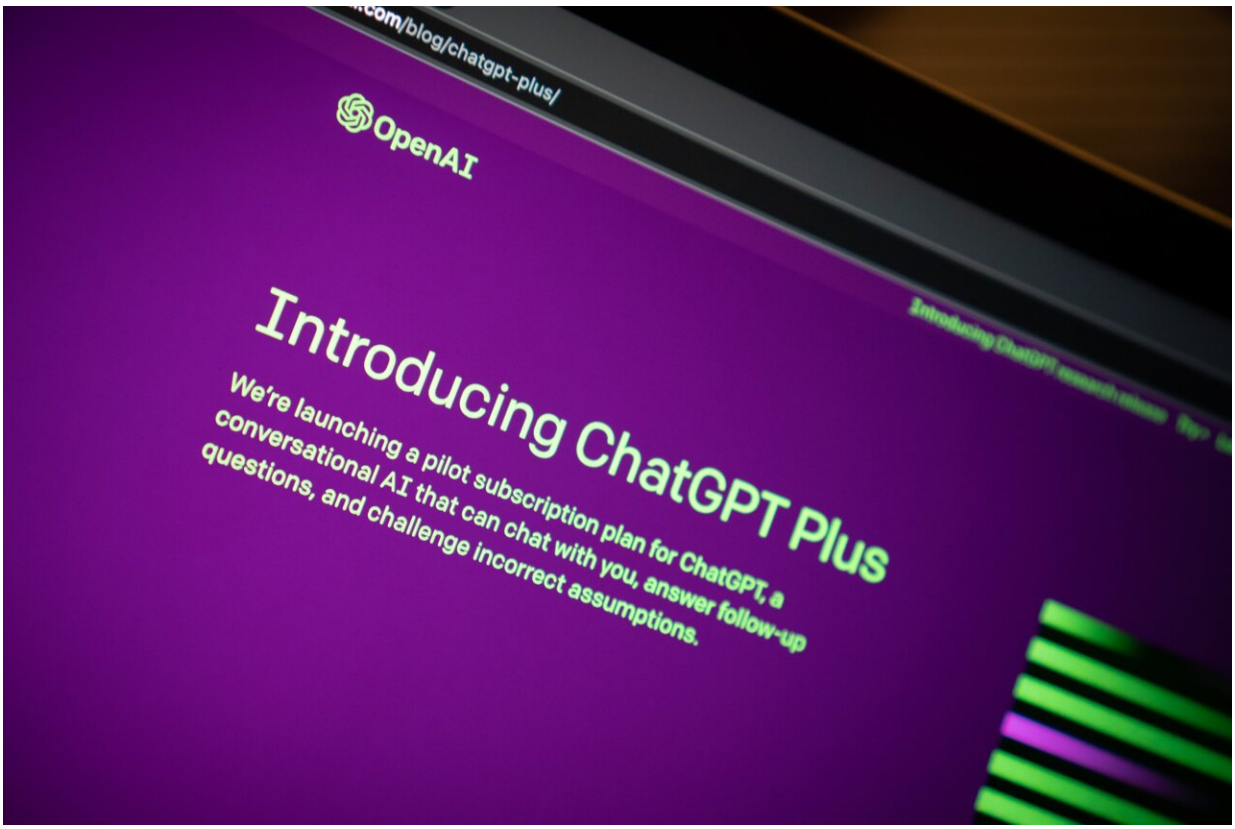
GPT-3 (dis)informs us better than humans, study finds

June 28 2023



Credit: Unsplash/CC0 Public Domain

A recent study conducted by researchers at the University of Zurich has delved into the capabilities of AI models, specifically focusing on OpenAI's GPT-3, to determine their potential risks and benefits in

generating and disseminating (dis)information.

Led by postdoctoral researchers Giovanni Spitale and Federico Germani, alongside Nikola Biller-Andorno, director of the Institute of Biomedical Ethics and History of Medicine (IBME), University of Zurich, the study involving 697 participants sought to evaluate whether individuals could differentiate between disinformation and accurate information presented in the form of tweets.

Furthermore, the researchers aimed to determine if participants could discern whether a tweet was written by a genuine Twitter user or generated by GPT-3, an advanced AI language model. The topics covered included climate change, vaccine safety, the COVID-19 pandemic, flat earth theory, and homeopathic treatments for cancer.

The paper is published in *Science Advances*.

Examples of fake news created by AI. Credit: UZH

AI-powered systems could generate large-scale disinformation campaigns

On the one hand, GPT-3 demonstrated the ability to generate accurate and—compared to tweets from real Twitter users—more easily comprehensible information. However, the researchers also discovered that the AI language model had an unsettling knack for producing highly persuasive disinformation. In a concerning twist, participants were unable to reliably differentiate between tweets created by GPT-3 and those written by real Twitter users.

"Our study reveals the power of AI to both inform and mislead, raising critical questions about the future of information ecosystems," says Federico Germani.

These findings suggest that information campaigns created by GPT-3, based on well-structured prompts and evaluated by trained humans, would prove more effective for instance in a public health crisis which requires fast and clear communication to the public. The findings also raise significant concerns regarding the threat of AI perpetuating disinformation, particularly in the context of the rapid and widespread dissemination of misinformation and disinformation during a crisis or public health event.

The study reveals that AI-powered systems could be exploited to generate large-scale disinformation campaigns on potentially any topic, jeopardizing not only public health but also the integrity of information ecosystems vital for functioning democracies.

Examples of true news created by AI. Credit: UZH

Proactive regulation highly recommended

As the impact of AI on information creation and evaluation becomes increasingly pronounced, the researchers call on policymakers to respond with stringent, evidence-based and ethically informed regulations to address the potential threats posed by these disruptive technologies and ensure the responsible use of AI in shaping our collective knowledge and well-being.

"The findings underscore the critical importance of proactive regulation to mitigate the potential harm caused by AI-driven disinformation campaigns," says Nikola Biller-Andorno. "Recognizing the risks associated with AI-generated disinformation is crucial for safeguarding public health and maintaining a robust and trustworthy information ecosystem in the digital age."

More information: Giovanni Spitale et al, AI model GPT-3 (dis)informs us better than humans, *Science Advances* (2023). DOI: 10.1126/sciadv.adh1850

Provided by University of Zurich

Citation: GPT-3 (dis)informs us better than humans, study finds (2023, June 28) retrieved 8 May 2024 from https://techxplore.com/news/2023-06-gpt-disinforms-humans.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.