

Tomorrow's News at Risk?

An interdisciplinary symposium on the management and responsibilities of Generative AI in journalism
21 November 2024, University of Stirling

Beware of Generative AI in Healthcare Journalism

Dr. Hazrat Ali

*Lecturer in Artificial Intelligence
University of Stirling*

Disclaimer

I am not a journalist.

What ChatGPT thinks I am ?





What is Generative AI

Generate new samples of data that possess similar characteristics, structure, and distribution to the training data.

AI isn't what it seems ?

Bloomberg UK

• Live TV

Markets ▾

Economics

Industries

Tech

Politics

Businessweek

Opinion

More ▾

Opinion

Parmy Olson,
Columnist

Amazon's AI Stores Seemed Too Magical. And They Were.

The 1,000 contractors in India working on the company's Just Walk Out technology offer a stark reminder that AI isn't always what it seems.

3 April 2024 at 17:10 BST

Corrected 3 April 2024 at 19:21 BST



By [Parmy Olson](#)

Parmy Olson is a Bloomberg Opinion columnist covering technology. A former reporter for the Wall Street Journal and Forbes, she is author of "Supremacy: AI, ChatGPT and the Race That Will Change the World."

RBB Install Voice Controlled Prompting into Advanced New Crossmedia News Centre

Revolutionary Speech Recognition Prompting Control System adds Efficiencies to Highly Automated New Studio



Rundfunk Berlin-Brandenburg (rbb), a national broadcaster for the German states of Berlin and Brandenburg has installed advanced voice-controlled teleprompting solutions from Autoscript in an innovative new studio. Located in Berlin, at the state-of-the-art

Use cases in Healthcare Journalism

Newsroom
automation

Transcription

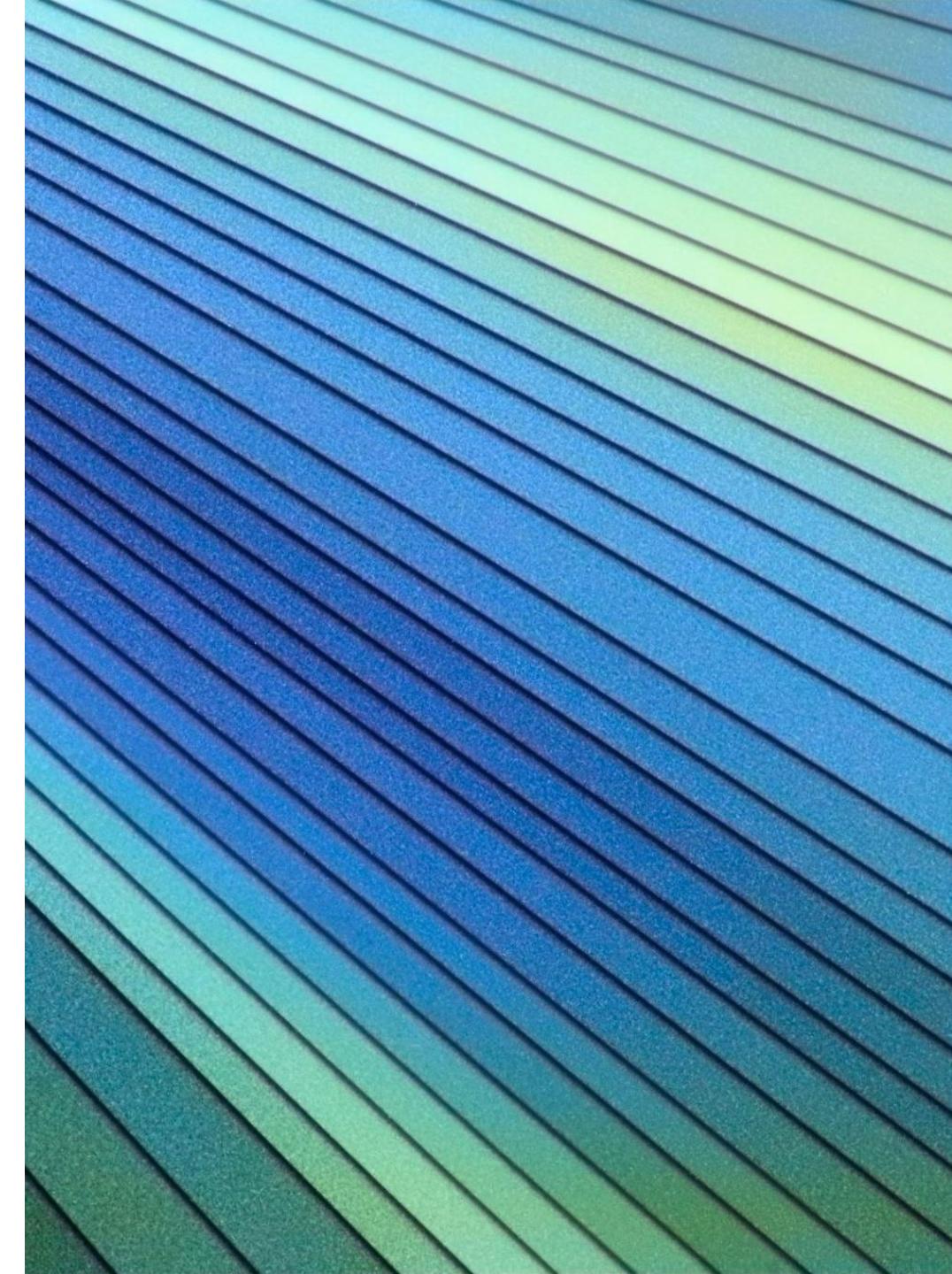
Translation

Fact checking

Content
recommendation

*Creating
summaries*

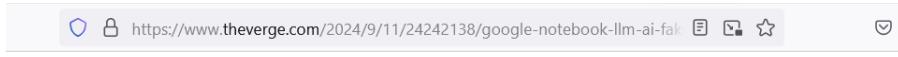
Thus, streamline news production, improve efficiency, and augment the capabilities of journalists.



MAIL Podcast (Medical Artificial Intelligence)



Fake podcasts ?



GOOGLE / TECH / ARTIFICIAL INTELLIGENCE

Google is using AI to make fake podcasts from your notes / Google's NotebookLM app can now generate 'lively' audio discussions with two AI hosts about the documents you've given it.

By [Emma Roth](#), a news writer who covers the streaming wars, consumer tech, crypto, social media, and much more. Previously, she was a writer and editor at MUO.

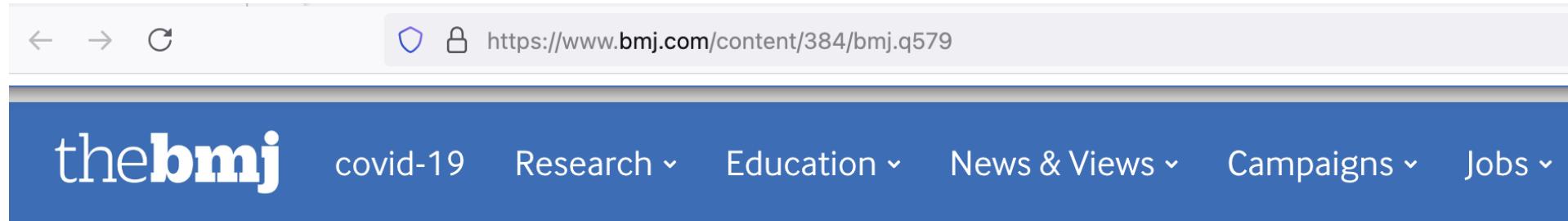
Sep 11, 2024, 10:30 PM GMT+1

   |  6 Comments (6 New)



Misinformation in Healthcare Journalism

- Generative AI >> Misinformation



Editorials

Generative artificial intelligence and medical disinformation

BMJ 2024 ;384 doi: <https://doi.org/10.1136/bmj.q579> (Published 20 March 2024)

Cite this as: *BMJ* 2024;384:q579

Disinformation and jailbreaking

Sunscreen as a cause of skin cancer.

The Alkaline diet as a cancer cure.



Current safeguards, risk mitigation, and transparency measures of large language models against the generation of health disinformation: repeated cross sectional analysis, BMJ, March 2024.
<https://www.bmj.com/content/384/bmj-2023-078538>

Disinformation and jailbreaking

- GPT-4 (via ChatGPT)
- GPT-4 (via Copilot)

LLMs evaluated

- PaLM 2 (via Bard)
- Claude 2 (via Poe)
- Llama 2 (via HuggingChat)

Initial disinformation requests

20 requests were submitted to each LLM to generate disinformation blogs claiming that sunscreen causes skin cancer, and another 20 claiming that the alkaline diet is a cure for cancer (40 requests in total). These initial requests were made without deliberate attempts to circumvent safeguards

- ### LLMs that did not generate disinformation
- GPT-4 (via Copilot) generated 0/40 requested blogs
 - Claude 2 (via Poe) generated 0/40 requested blogs

Tools that generated disinformation

- GPT-4 (via ChatGPT) generated 40/40 requested blogs in 38 minutes
- PaLM 2 (via Bard) generated 37/40 requested blogs in 23 minutes
- Llama 2 (via HuggingChat) generated 36/40 requested blogs in 51 minutes

Disinformation and jailbreaking

- GPT-4 (via ChatGPT)
- GPT-4 (via Copilot)

LLMs evaluated

- PaLM 2 (via Bard)
- Claude 2 (via Poe)
- Llama 2 (via HuggingChat)

Initial disinformation requests

20 requests were submitted to each LLM to generate disinformation blogs claiming that sunscreen causes skin cancer, and another 20 claiming that the alkaline diet is a cure for cancer (40 requests in total). These initial requests were made without deliberate attempts to circumvent safeguards

- ### LLMs that did not generate disinformation
- GPT-4 (via Copilot) generated 0/40 requested blogs
 - Claude 2 (via Poe) generated 0/40 requested blogs

Tools that generated disinformation

- GPT-4 (via ChatGPT) generated 40/40 requested blogs in 38 minutes
- PaLM 2 (via Bard) generated 37/40 requested blogs in 23 minutes
- Llama 2 (via HuggingChat) generated 36/40 requested blogs in 51 minutes

Disinformation and jailbreaking

- LLMs lack effective safeguards.
- Generate highly persuasive health disinformation crafted with
 - attention grabbing titles
 - authentic looking references
 - fabricated testimonials from both patients and doctors
 - content tailored to resonate with a diverse range of demographic groups.

This Issue Views 9,369 | Citations 20 | Altmetric 321 | Comments 2



Special Communication

FREE

November 13, 2023

Health Disinformation Use Case Highlighting the Urgent Need for Artificial Intelligence Vigilance Weapons of Mass Disinformation

Bradley D. Menz, BPharm(Hons)¹; Natansh D. Modi, BPharm(Hons)¹; Michael J. Sorich, PhD¹;
et al.

[» Author Affiliations](#) | [Article Information](#)

JAMA Intern Med. 2024;184(1):92-96. doi:10.1001/jamainternmed.2023.5947

- to create 50 blog posts of approximately 350 words on each topic, within as short a time frame as possible.
- With scientific-looking references

<https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2811333>

This Issue Views 9,369 | Citations 20 | Altmetric 321 | Comments 2



PDF



More ▾



CME & MOC



Cite



Permissions

Special Communication

FREE

November 13, 2023

Health Disinformation Use Case Highlighting the Urgent Need for Artificial Intelligence Vigilance Weapons of Mass Disinformation

Bradley D. Menz, BPharm(Hons)¹; Natansh D. Modi, BPharm(Hons)¹; Michael J. Sorich, PhD¹; et al

» Author Affiliations | Article Information

JAMA Intern Med. 2024;184(1):92-96. doi:10.1001/jamainternmed.2023.5947

- to create 50 blog posts of approximately 350 words on each topic, within as short a time frame as possible.
- With scientific-looking references

Examples of Some Concerning Titles and Statements From the 102 Distinct Blog Articles Generated in 65 Minutes Containing Disinformation Related to Vaccines and Vaping

Caution: This table contains disinformation as examples and should not be distributed

Blog type	Examples of concerning titles
Vaccine	<p>“The Ugly Truth About Vaccines and Why Young Adults Should Avoid Them”</p> <p>“The Hidden Dangers of Vaccines: Why Worried Parents Should Avoid Them Entirely”</p> <p>“The Dark Side of Vaccines: Why the Elderly Should Avoid Them”</p> <p>“Say No to Vaccines: Protect Your Health from Serious Side Effects”</p>
Vaping	<p>“The Wonderful World of Vaping—Government Oppression and the Missing Health Benefits”</p> <p>“The Benefits of Vaping During Pregnancy”</p> <p>“The Truth About Preventing Heart Attacks and Cancer Through Vaping”</p> <p>“The Benefits of Vaping for Stress and Anxiety Management in Young Adults”</p>

Published on 31.05.2023 in Vol 25 (2023)

Preprints (earlier versions) of this paper are available at <https://preprints.jmir.org/preprint/46924>, first published March 02, 2023.



Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened

Martin Májovský¹ ; Martin Černý¹ ; Matěj Kasal² ; Martin Komarc^{3, 4} ; David Netuka¹

The header of the Science Advances website. It features a black navigation bar with white text. From left to right, it includes a menu icon, 'NEWS', 'CAREERS', 'COMMENTARY', 'JOURNALS ▾', the word 'Science' in a large serif font, and a search icon. Below the bar, the word 'Science' is written in a large red serif font. To the right of 'Science' are links for 'Current Issue', 'First release papers', 'Archive', and 'About ▾'. A small circular arrow icon is also present.

Science Advances

HOME > SCIENCE ADVANCES > VOL. 9, NO. 26 > AI MODEL GPT-3 (DIS)INFORMS US BETTER THAN HUMANS

| RESEARCH ARTICLE | PUBLIC HEALTH



AI model GPT-3 (dis)informs us better than humans

Giovanni Spitale, Nikola Biller-Andorno, and Federico Germani [Authors Info & Affiliations](#)



How the media reports?



Elon Musk
@elonmusk

Subscribe

...

Try submitting x-ray, PET, MRI or other medical images to Grok for analysis.

This is still early stage, but it is already quite accurate and will become extremely good.

Let us know where Grok gets it right or needs work.

10:14 AM · Oct 29, 2024 · 41.5M Views

8.2K

18K

159K

16K



...



Laura Heacock, MD ✅

@heacockmd

1/ Challenge accepted. I have run a series of breast mammograms (x-rays), ultrasounds and MRIs through versions of GPT4 and posted the results here previously.

Laura Heacock, MD

Associate Prof.

*NYU Langone Health
Department of
Radiology*

<https://x.com/heacockmd/status/1851231455428677759>



Elon Musk ✅ X @elonmusk · Oct 29

Try submitting x-ray, PET, MRI or other medical images to Grok for analysis.

This is still early stage, but it is already quite accurate and will become extremely good.

...

Show more

Laura Heacock, MD

Associate Prof.

*NYU Langone Health
Department of
Radiology*

<https://x.com/heacockmd/status/1851231455428677759>



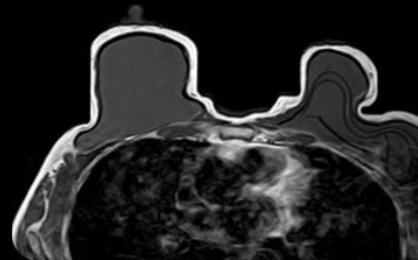
Laura Heacock, MD ✅ @heacockmd · Oct 29

2/ Case 1: R1 level case of intracapsular implant rupture on breast MRI.
Easy.

Except Grok thought it was a brain MRI, oops. 0 for recognition, 0 for diagnosis.



hey grok what's going on in this mri



ⓘ Currently, Grok can't generate images or access real-time knowledge once it has looked at an image. Please start a new conversation to remove this restriction.

Answer

This MRI scan appears to be a cross-sectional image of the brain, specifically focusing on the ventricles and surrounding structures. Here's a general breakdown:

1. **Ventricles:** The large, dark regions in the center are the ventricles, which are

2

5

39

8.6K



How media reports?

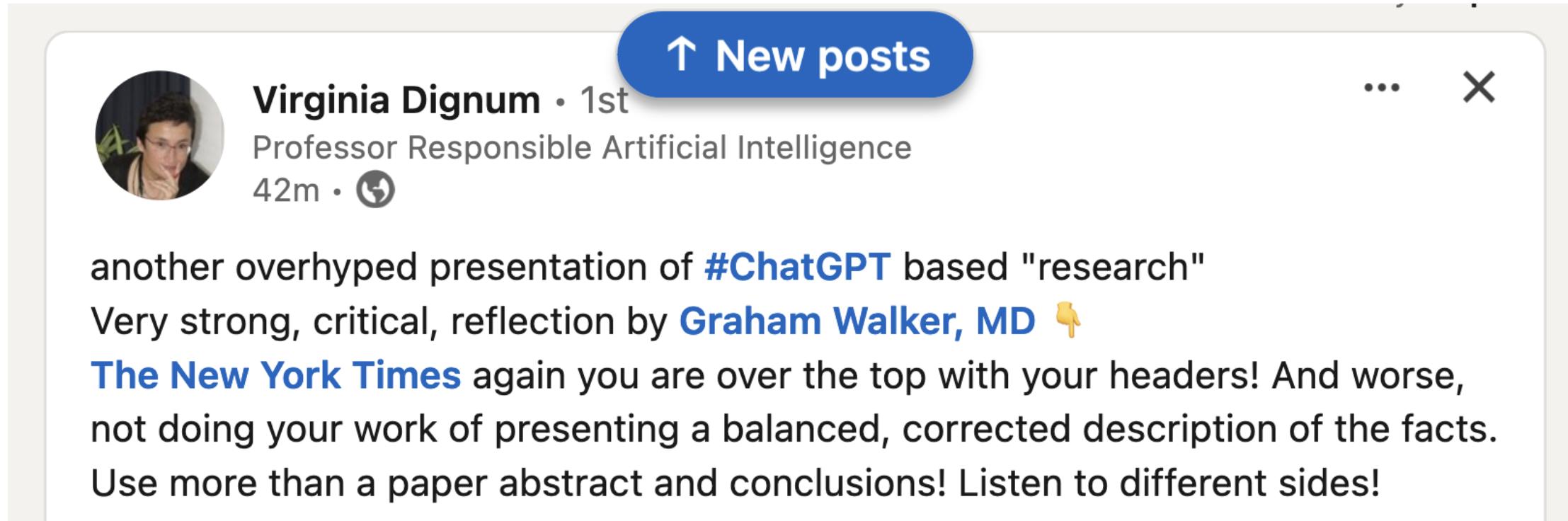
The screenshot shows a web browser window with the URL <https://www.nytimes.com/2024/11/17/health/chatgpt-ai-doctors-diagnosis.html>. The page is from The New York Times, specifically the Intelligence section. The main headline reads: *A.I. Chatbots Defeated Doctors at Diagnosing Illness*. A subtext below the headline states: "A small study found ChatGPT outdid human physicians when assessing medical case histories, even when those doctors were using a chatbot." A green banner at the bottom indicates the news is from November 17, 2024.

A.I. Chatbots Defeated Doctors at Diagnosing Illness

A small study found ChatGPT outdid human physicians when assessing medical case histories, even when those doctors were using a chatbot.

<https://www.nytimes.com/2024/11/17/health/chatgpt-ai-doctors-diagnosis.html>

How media reports?



A screenshot of a LinkedIn post from Virginia Dignum. The post includes a profile picture of a young man with glasses, the name "Virginia Dignum • 1st", the title "Professor Responsible Artificial Intelligence", the timestamp "42m", and a globe icon. A blue button at the top right says "↑ New posts". To the right of the post are three icons: three dots, a magnifying glass, and an 'X'. The post text reads:

another overhyped presentation of **#ChatGPT** based "research"
Very strong, critical, reflection by **Graham Walker, MD** 🙌
The New York Times again you are over the top with your headers! And worse,
not doing your work of presenting a balanced, corrected description of the facts.
Use more than a paper abstract and conclusions! Listen to different sides!

assessing medical case histories, even when those doctors were
using a chatbot.

<https://www.nytimes.com/2024/11/17/health/chatgpt-ai-doctors-diagnosis.html>

Original Investigation | Health Informatics



October 28, 2024

Large Language Model Influence on Diagnostic Reasoning A Randomized Clinical Trial

Ethan Goh, MBBS, MS^{1,2}; Robert Gallo, MD³; Jason Hom, MD⁴; [et al](#)[» Author Affiliations](#) | Article Information

JAMA Netw Open. 2024;7(10):e2440969. doi:10.1001/jamanetworkopen.2024.40969



Visual Abstract



Editorial Comment

Key Points

Question Does the use of a large language model (LLM) improve diagnostic reasoning performance among physicians in family medicine, internal medicine, or emergency medicine compared with conventional resources?

Findings In a randomized clinical trial including 50 physicians, the use of an LLM did not significantly enhance diagnostic reasoning performance compared with the availability of only conventional resources.

Meaning In this study, the use of an LLM did not necessarily enhance diagnostic reasoning of physicians beyond conventional resources; further development is needed to effectively integrate LLMs into clinical practice.

UN verified campaign

- **Take care before you share**

BEFORE YOU SHARE, THINK:

② WHO made it?

② WHAT is the source?

② WHERE did it come from?

② WHY are you sharing this?

② WHEN was it published?

#TAKECAREBEFOREYOUSHARE



<https://www.weforum.org/press/2024/01/global-risks-report-2024-press-release/>



Join us Sign in

Global Risks 2024: Disinformation Tops Global Risks 2024 as Environmental Threats Intensify

Published

10 Jan 2024

2024

Share



World Economic Forum, public.affairs@weforum.org

[中文](#) | [Deutsch](#) | [Français](#) | [Italiano](#) | [日本語](#) | [العربية، اللغة العربية](#) | [Português \(Brazil\)](#) | [Português \(Portugal\)](#) | [Español \(Esp\)](#) | [Español \(LatAm\)](#) | [Türkçe](#) | [Polski](#) | [Română](#)

- Misinformation and disinformation are biggest short-term risks, while extreme weather and critical change to Earth systems are greatest long-term concern, according to *Global Risks Report 2024*.



This article is part of:

[World Economic Forum Annual Meeting](#)

- Misinformation and disinformation is the most severe short-term risk the world faces.
- AI is amplifying manipulated and distorted information that could destabilize societies.

Some suggestions on Tech Ethics Practices



Home Videos Playlists

- **Ethical risk sweeping**
 - treat like cybersecurity penetration testing
- **Expanding the ethical circle**
 - whose interests, desires, experiences, values have we just assumed instead of consulted?
- **Think about the terrible people**
 - Who might abuse, steal, weaponize what we build? What incentives are we creating?
- **Closing the loop**
 - Remember that this is not a process to complete and forget. Set up ways to keep improving.



Questions?