

WHO calls for caution on using ChatGPT-style models in health care

Source copy kept by wheresthe.ai for the story "WHO calls for caution on using ChatGPT-style models in health care" (<https://wheresthe.ai/d/intl-who-llm-health-caution-2023-05/>).

Source: <https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: e4e4f855f8af8da330881fd91c9f2a8ed99d6cf0e9d13d9173c2f80e54f35071

The World Health Organization (WHO) is calling for caution to be exercised in using artificial intelligence (AI) generated large language model tools (LLMs) to protect and promote human well-being, human safety, and autonomy, and preserve public health.

LLMs include some of the most rapidly expanding platforms such as ChatGPT, Bard, Bert and many others that imitate understanding, processing, and producing human communication. Their meteoric public diffusion and growing experimental use for health-related purposes is generating significant excitement around the potential to support people's health needs.

It is imperative that the risks be examined carefully when using LLMs to improve access to health information, as a decision-support tool, or even to enhance diagnostic capacity in under-resourced settings to protect people's health and reduce inequity.

While WHO is enthusiastic about the appropriate use of technologies, including LLMs, to support health-care professionals, patients, researchers and scientists, there is concern that caution that would normally be exercised for any new technology is not being exercised consistently with LLMs. This includes widespread adherence to key values of transparency, inclusion, public engagement, expert supervision, and rigorous evaluation.

Precipitous adoption of untested systems could lead to errors by health-care workers, cause harm to patients, erode trust in AI and thereby undermine (or delay) the potential long-term benefits and uses of such technologies around the world.

Concerns that call for rigorous oversight needed for the technologies to be used in safe, effective, and ethical ways include:

the data used to train AI may be biased, generating misleading or inaccurate information that could pose risks to health, equity and inclusiveness; LLMs generate responses that can appear authoritative and plausible to an end user; however, these responses may be completely incorrect or contain serious errors, especially for health-related responses; LLMs may be trained on data for which consent may not have been previously provided for such use, and LLMs may not protect sensitive data (including health data) that a user provides to an application to generate a response; LLMs can be misused to generate and disseminate highly convincing disinformation in the form of text, audio or video content that is difficult for the public to differentiate from reliable health content; and while committed to harnessing new technologies, including AI and digital health to improve human health, WHO recommends that policy-makers ensure patient safety and protection while technology firms work to commercialize LLMs. WHO proposes that these concerns be addressed, and clear evidence of benefit be measured before their widespread use in routine health care and medicine – whether by individuals, care providers or health system administrators and policy-makers.

WHO reiterates the importance of applying ethical principles and appropriate governance, as enumerated in the WHO guidance on the ethics and governance of AI for health, when designing, developing, and deploying AI for health. The 6 core principles identified by WHO are: (1) protect autonomy; (2) promote human well-being, human safety, and the public interest; (3) ensure transparency, explainability, and intelligibility; (4) foster responsibility and accountability; (5) ensure inclusiveness and equity; (6) promote AI that is responsive and sustainable.