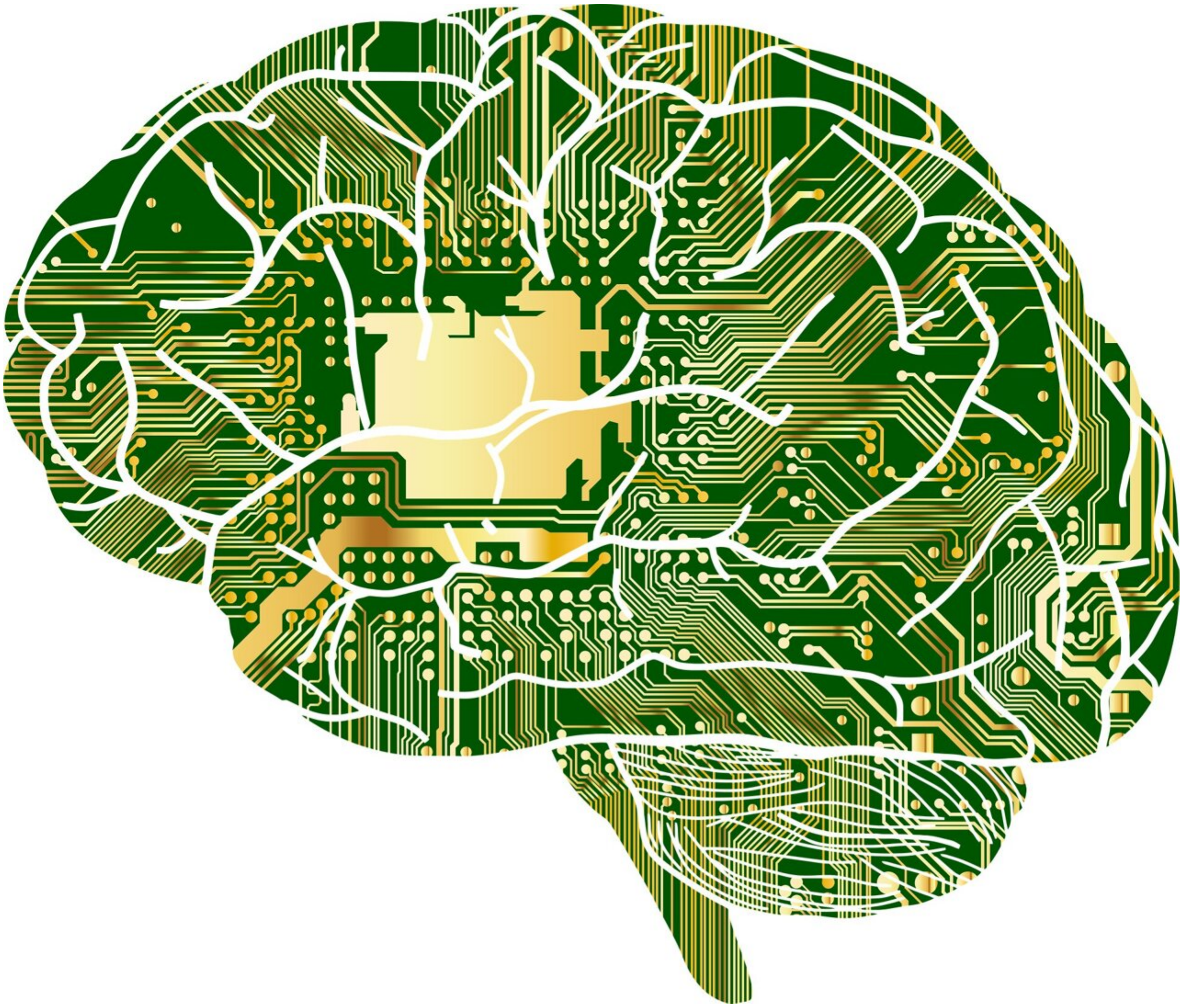
Medical Xpress

New paper introduces ethics framework for use of generative AI in health care

May 16 2023



Credit: Pixabay/CC0 Public Domain

A new paper published by leading Australian AI ethicist Stefan Harrer Ph.D. proposes for the first time a comprehensive ethical framework for the responsible use, design, and governance of Generative AI applications in health care and medicine.

The study, published in *eBioMedicine*, details how Large Language Models (LLMs) have the potential to fundamentally transform information management, education, and communication workflows in health care and medicine but equally remain one of the most dangerous and misunderstood types of AI.

"LLMs used to be boring and safe. They have become exciting and dangerous," said Dr. Harrer who is also the Chief Innovation Officer of major Australian funder of digital health research and development, the Digital Health Cooperative Research Centre (DHCRC) and a member of the Coalition for Health AI (CHAI).

"This study is a plea for regulation of generative AI technology in health care and medicine and provides technical and governance guidance to all stakeholders of the digital health ecosystem: developers, users, and regulators. Because generative AI should be both exciting and safe."

LLMs are a key component of generative AI applications for creating new content including text, imagery, audio, code, and videos in response to textual instructions. Prominent examples scrutinized in the study against ethical design, release and use principles and performance include OpenAI's chatbot ChatGPT, Google's chatbot Med-PALM, Stability AI's imagery generator Stable Diffusion, and Microsoft's BioGPT bot.

The study highlights and explains many key applications for health care:

- assisting clinicians with the generation of medical reports or

preauthorization letters;
- helping medical students to study more efficiently;
- simplifying medical jargon in clinician-patient communication;
- increasing the efficiency of clinical trial design;
- helping to overcome interoperability and standardization hurdles in EHR mining;
- making drug discovery and design processes more efficient.

However, the paper also highlights that the inherent danger of LLM-driven generative AI arising from the ability of LLMs to authoritatively and convincingly produce and disseminate false, inappropriate, and dangerous content at unprecedented scale is increasingly being marginalized in an ongoing hype around the recently released latest generation of powerful LLM chatbots.

A framework for mitigating risks of AI in health care

As part of the study, Dr. Harrer identified a comprehensive set of risk factors which are of special relevance to using LLM technology as part of generative AI systems in health and medicine, and proposes risk mitigation pathways for each of them. The study highlights and analyzes real life use cases of both, ethical and unethical development of LLM technology.

"Good actors chose to follow an ethical path to building safe generative AI applications. Bad actors, however, are getting away with doing the opposite: hastily productizing and releasing LLM-powered generative AI tools into a fast-growing commercial market they gamble with the well-being of users and the integrity of AI and knowledge databases at scale. This dynamic needs to change," said Dr. Harrer.

Dr. Harrer argues that the limitations of LLMs are systemic and rooted in their lack of language comprehension.

"The essence of efficient knowledge retrieval is to ask the right questions, and the art of critical thinking rests on one's ability to probe responses by assessing their validity against models of the world. LLMs can perform none of these tasks. They are in-betweeners which can narrow down the vastness of all possible responses to a prompt to the most likely ones but are unable to assess whether prompt or response made sense or were contextually appropriate," Dr. Harrer said.

Therefore, he suggests that boosting training data sizes and building ever more complex LLMs will not mitigate risks but rather amplify them. The study proposes alternative approaches to ethically (re-) designing generative AI applications, to shaping regulatory frameworks, and to directing technical research efforts towards exploring methods for implementation and enforcement of ethical design and use principles.

Dr. Harrer proposes a regulatory framework with 10 principles for mitigating the risks of generative AI in health:

1. design AI as an assistive tool for augmenting the capabilities of human decision makers, not for replacing them;
2. design AI to produce performance, usage and impact metrics explaining when and how AI is used to assist decision making and scan for potential bias,
3. study the value systems of target user groups and design AI to adhere to them;
4. declare the purpose of designing and using AI at the outset of any conceptual or development work,
5. disclose all training data sources and data features;
6. design AI systems to clearly and transparently label any AI-generated content as such;
7. ongoingly audit AI against data privacy, safety, and performance standards;
8. maintain databases for documenting and sharing the results of AI

audits, educate users about model capabilities, limitations and risks, and improve performance and trustworthiness of AI systems by retraining and redeploying updated algorithms;

9. apply fair-work and safe-work standards when employing human developers;
10. establish legal precedence to define under which circumstances data may be used for training AI, and establish copyright, liability and accountability frameworks for governing the legal dependencies of training data, AI-generated content, and the impact of decisions humans make using such data.

"Without human oversight, guidance and responsible design and operation, LLM-powered generative AI applications will remain a party trick with substantial potential for creating and spreading misinformation or harmful and inaccurate content at unprecedented scale," said Dr. Harrer.

He predicts that the field will move from the current competitive LLM arms race to a phase of more nuanced and risk-conscious experimentation with research-grade generative AI applications in health, medicine and biotech which will deliver first commercial product offerings for niche applications in digital health data management within the next 2 years.

"I am inspired by thinking about the transformative role generative AI and LLMs could one day play in health care and medicine, But I am also acutely aware that we are by no means there yet and that, despite the prevailing hype, LLM-powered generative AI may only gain the trust and endorsement of clinicians and patients if the research and development community aims for equal levels of ethical and technical integrity as it progresses this transformative technology to market maturity."

"The DHCRC has a critical role in translating ethical AI into practice," said DHCRC CEO Annette Schmiede. "There is a newfound enthusiasm for the role of generative AI in transforming health care and we are at a tipping point where AI will start to become ever more integrated into the digital health ecosystem. We are on the frontline and frameworks like the one outlined in this paper will become critical to ensure an ethical and safe use of AI."

"Ethical AI requires a lifecycle approach from data curation to model testing, to ongoing monitoring. Only with the right guidelines and guardrails can we ensure our patients benefit from emerging technologies while minimizing bias and unintended consequences," said John Halamka, M.D., M.S, President of Mayo Clinic Platform and a co-founder of CHAI.

"This study provides important ethical and technical guidance to users, developers, providers, and regulators of generative AI and incentivizes them to responsibly and collectively prepare for the transformational role this technology could play in health and medicine," said Brian Anderson, M.D., Chief Digital Health Physician at MITRE.

More information: Stefan Harrer, Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine, *eBioMedicine* (2023). DOI: 10.1016/j.ebiom.2023.104512

Provided by Digital Health Cooperative Research Centre

Citation: New paper introduces ethics framework for use of generative AI in health care (2023, May 16) retrieved 30 June 2024 from https://medicalxpress.com/news/2023-05-paper-ethics-framework-generative-ai.html