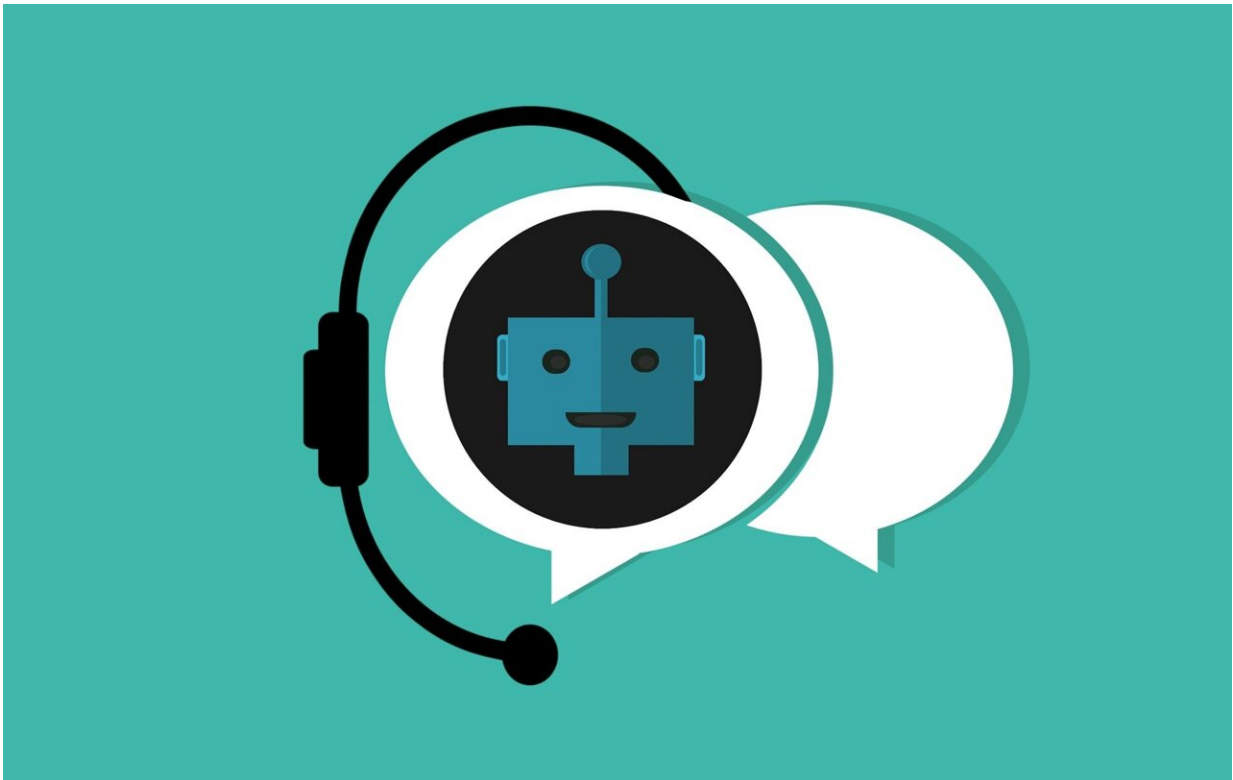# Researchers outline how AI chatbots could be approved as medical devices

June 30 2023

Credit: Pixabay/CC0 Public Domain

LLM-based generative chat tools, such as ChatGPT or Google's MedPaLM have great medical potential, but there are inherent risks associated with their unregulated use in health care. A new *Nature Medicine* paper by Prof. Stephen Gilbert and colleagues addresses one of

the most pressing international issues of our time: How to regulate Large Language Models (LLMs) in general and specifically in health.

"Large Language Models are neural network language models with remarkable conversational skills. They generate human-like responses and engage in interactive conversations. However, they often generate highly convincing statements that are verifiably wrong or provide inappropriate responses," said Prof. Stephen Gilbert, Professor for Medical Device Regulatory Science at Else Kröner Fresenius Center for Digital Health at TU Dresden.

"Today there is no way to be certain about the quality, evidence level, or consistency of clinical information or supporting evidence for any response. These chatbots are unsafe tools when it comes to medical advice and it is necessary to develop new frameworks that ensure patient safety."

Challenges in the regulatory approval of large language models

Most people research their symptoms online before seeking medical advice. Search engines play a role in decision-making process. The forthcoming integration of LLM-chatbots into search engines may increase users' confidence in the answers given by a chatbot that mimics conversation. It has been demonstrated that LLMs can provide profoundly dangerous information when prompted with medical questions. LLM's underlying approach has no model of medical "ground truth," which is dangerous.

Chat interfaced LLMs have already provided harmful medical responses and have already been used unethically in "experiments" on patients without consent. Almost every medical LLM use case requires regulatory control in the EU and U.S. In the U.S. the lack of

explainability disqualifies them from being "non devices." LLMs with explainability, low bias, predictability, correctness, and verifiable outputs do not currently exist and they are not exempted from current (or future) governance approaches.

In this paper the authors describe the limited scenarios in which LLMs could find application under current frameworks, they describe how developers can seek to create LLM-based tools that could be approved as medical devices, and they explore the development of new frameworks that preserve patient safety.

"Current LLM-chatbots do not meet key principles for AI in health care, like bias control, explainability, systems of oversight, validation and transparency. To earn their place in medical armamentarium, chatbots must be designed for better accuracy, with safety and clinical efficacy demonstrated and approved by regulators," concludes Prof. Gilbert.