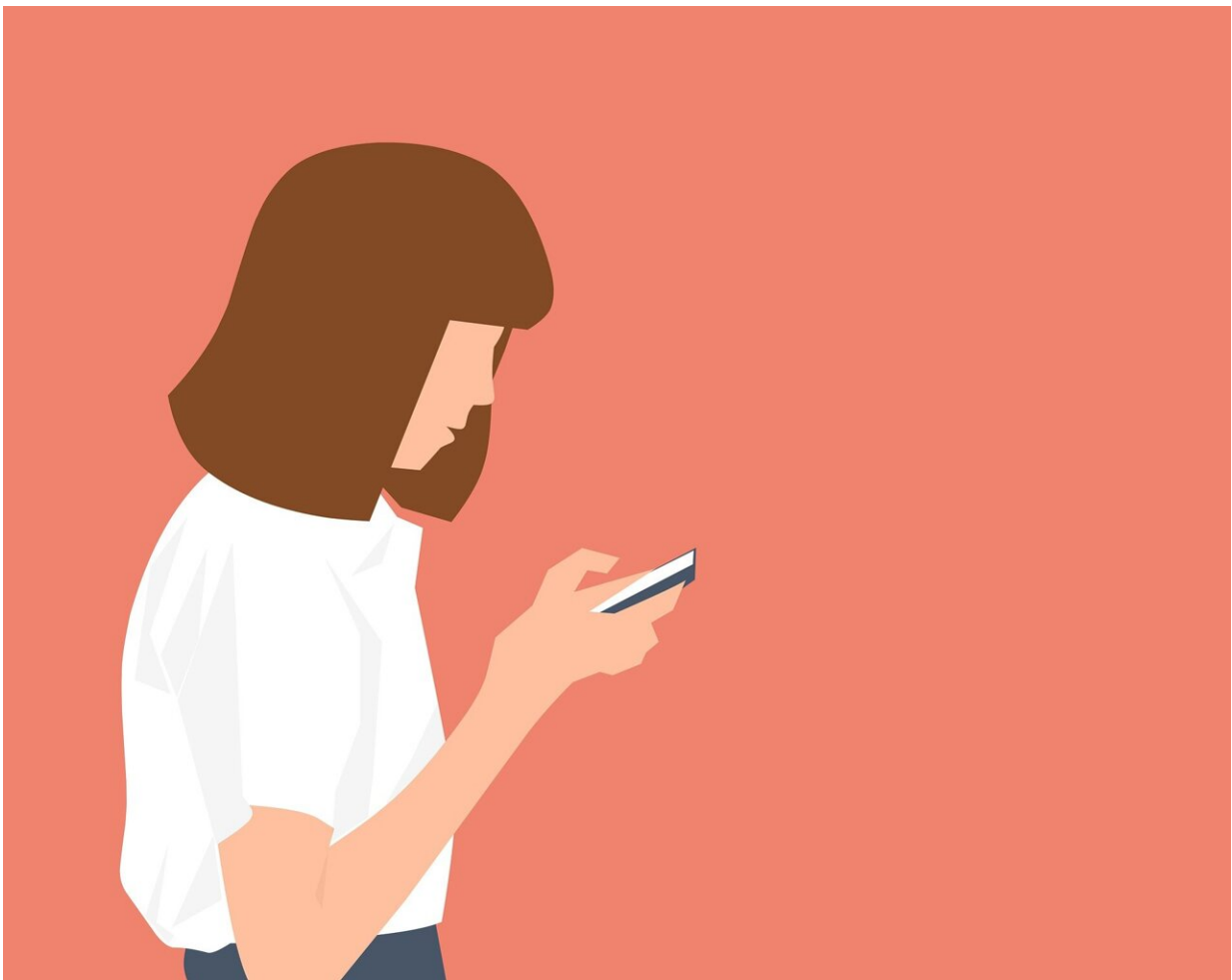


Research identifies pitfalls and opportunities for generative AI in patient messaging systems

April 24 2024



Credit: Pixabay/CC0 Public Domain

A new study by investigators from Mass General Brigham demonstrates that large language models (LLMs), a type of generative AI, may help reduce physician workload and improve patient education when used to draft replies to patient messages.

The study also found limitations to LLMs that may affect [patient safety](#), suggesting that vigilant oversight of LLM-generated communications is essential for safe usage. Findings, published in *The Lancet Digital Health*, emphasize the need for a measured approach to LLM implementation.

Rising administrative and documentation responsibilities have contributed to increases in physician burnout. To help streamline and automate physician workflows, electronic health record (EHR) vendors have adopted generative AI algorithms to aid clinicians in drafting messages to patients; however, the efficiency, safety and clinical impact of their use had been unknown.

"Generative AI has the potential to provide a 'best of both worlds' scenario of reducing burden on the clinician and better educating the patient in the process," said corresponding author Danielle Bitterman, MD, a faculty member in the Artificial Intelligence in Medicine (AIM) Program at Mass General Brigham and a physician in the Department of Radiation Oncology at Brigham and Women's Hospital.

"However, based on our team's experience working with LLMs, we have concerns about the potential risks associated with integrating LLMs into messaging systems. With LLM-integration into EHRs becoming increasingly common, our goal in this study was to identify relevant benefits and shortcomings."

For the study, the researchers used OpenAI's GPT-4, a foundational LLM, to generate 100 scenarios about patients with cancer and an accompanying patient question. No questions from actual patients were

used for the study. Six [radiation oncologists](#) manually responded to the queries; then, GPT-4 generated responses to the questions.

Finally, the same radiation oncologists were provided with the LLM-generated responses for review and editing. The radiation oncologists did not know whether GPT-4 or a human had written the responses, and in 31% of cases, believed that an LLM-generated response had been written by a human.

On average, physician-drafted responses were shorter than the LLM-generated responses. GPT-4 tended to include more educational background for patients but was less directive in its instructions. The physicians reported that LLM-assistance improved their perceived efficiency and deemed the LLM-generated responses to be safe in 82.1% of cases and acceptable to send to a patient without any further editing in 58.3% of cases.

The researchers also identified some shortcomings: If left unedited, 7.1% of LLM-generated responses could pose a risk to the patient and 0.6% of responses could pose a risk of death, most often because GPT-4's response failed to urgently instruct the patient to seek immediate medical care.

Notably, LLM-generated/physician-edited responses were more similar in length and content to LLM-generated responses versus the manual responses. In many cases, physicians retained LLM-generated educational content, suggesting that they perceived it to be valuable. While this may promote patient education, the researchers emphasize that overreliance on LLMs may also pose risks, given their demonstrated shortcomings.

Going forward, the study's authors are investigating how patients perceive LLM-based communications and how patients' racial and

demographic characteristics influence LLM-generated responses, based on known [algorithmic biases](#) in LLMs.

"Keeping a human in the loop is an essential safety step when it comes to using AI in medicine, but it isn't a single solution," Bitterman said.

"As providers rely more on LLMs, we could miss errors that could lead to patient harm. This study demonstrates the need for systems to monitor the quality of LLMs, training for clinicians to appropriately supervise LLM output, more AI literacy for both patients and clinicians, and on a fundamental level, a better understanding of how to address the errors that LLMs make."

More information: Chen, S et al. The effect of using a large language model to respond to patient messages, *The Lancet Digital Health* (2024). DOI: [10.1016/S2589-7500\(24\)00060-8/](https://doi.org/10.1016/S2589-7500(24)00060-8/)

Provided by Mass General Brigham

Citation: Research identifies pitfalls and opportunities for generative AI in patient messaging systems (2024, April 24) retrieved 5 May 2024 from <https://medicalxpress.com/news/2024-04-pitfalls-opportunities-generative-ai-patient.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.