

Generative AI can not yet reliably read and extract information from clinical notes in medical records, finds study

August 19 2024



Credit: Unsplash/CC0 Public Domain

It may someday be possible to use Large Language Models (LLM) to automatically read clinical notes in medical records and reliably and

efficiently extract relevant information to support patient care or research. But recent research from Columbia University Mailman School of Public Health using ChatGPT-4 to read medical notes from Emergency Department admissions to determine whether injured scooter and bicycle riders were wearing a helmet finds that LLM can't yet do this reliably. The findings are [published](#) in *JAMA Network Open*.

In a study of 54,569 emergency department visits among patients injured while riding a bicycle, scooter or other micromobility conveyance from 2019 to 2022, the AI LLM had difficulty replicating results of a text string–search based approach for extracting helmet status from clinical notes.

The LLM only performed well when the prompt included all of the text used in the text string search-based approach. The LLM also had difficulty replicating its work across trials on each of five successive days, it did better at replicating its hallucinations than its accurate work. It particularly struggled when phrases were negated, such as reading "w/o helmet" or "unhelmeted" and reporting that the patient wore a helmet.

Large amounts of medically relevant data are included in electronic [medical records](#) in the form of written clinical notes, a type of unstructured data. Efficient ways to read and extract information from these notes would be extremely useful for research.

Currently, information from these clinical notes can be extracted using simple string-matching text search approaches or through more sophisticated artificial intelligence (AI)–based approaches such as natural language processing. The hope was that a new LLM, such as ChatGPT-4, could extract information faster and more reliably.

"While we see potential efficiency gains in using the generative AI LLM for information extraction tasks, issues of reliability and hallucinations

currently limit its utility," said Andrew Rundle, DrPH, professor of Epidemiology at Columbia Mailman School and senior author.

"When we used highly detailed prompts that included all of the text strings related to helmets, on some days ChatGPT-4 could extract [accurate data](#) from the clinical notes. But the time required to define and test all of the text that had to be included in the prompt and ChatGPT-4's inability to replicate its work, day after day, indicates to us that ChatGPT-4 was not yet up to this task."

Using publicly available 2019 to 2022 data from the U.S. Consumer Product Safety Commission's National Electronic Injury Surveillance System, a sample of 96 U.S. hospitals, Rundle and colleagues analyzed emergency department records of patients injured in e-bike, bicycle, hoverboard, and powered scooter accidents. They compared the results of ChatGPT-4's analyses of the records to data generated using more traditional text-string-based searches, and for 400 records, they compared ChatGPT's analyses to their own reading of the clinical notes in the records.

This research builds on their work studying how to prevent injuries among micromobility users (i.e. bicyclists, e-bike riders, scooter riders). "Helmet use is a key factor in injury severity, yet in most [emergency department](#) medical records and incident reports information on helmet use is buried in the clinical notes written by the physician or EMS respondent. There is a significant research need to be able to reliably and efficiently access this information," said Kathryn Burford, the lead author on the paper and a post-doctoral fellow in the Department of Epidemiology at the Mailman School.

"Our study examined the potential of an LLM for extracting information from [clinical notes](#), a rich source of information for [health professionals](#) and researchers," said Rundle. "But at the time we used ChatGPT-4 it

could not reliably provide us with data."

Co-authors are Nicole G. Itzkowitz, Columbia Mailman School of Public Health; Ashley G. Ortega, Columbia Population Research Center; and Julien O. Teitler, Columbia School of Social Work.

More information: Kathryn G. Burford et al, Use of Generative AI to Identify Helmet Status Among Patients With Micromobility-Related Injuries From Unstructured Clinical Notes, *JAMA Network Open* (2024). DOI: [10.1001/jamanetworkopen.2024.25981](https://doi.org/10.1001/jamanetworkopen.2024.25981)

Provided by Columbia University's Mailman School of Public Health

Citation: Generative AI can not yet reliably read and extract information from clinical notes in medical records, finds study (2024, August 19) retrieved 19 August 2024 from <https://medicalxpress.com/news/2024-08-generative-ai-reliably-clinical-medical.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.