Generative artificial intelligence models effectively highlight social determinants of health in doctors' notes

January 11 2024

Infographic summarizing new resarch led by Danielle Bitterman, MD, that used large language models to identify social determinants of health from doctor visit notes. Credit: Mass General Brigham

Where we live and work, our age, and the conditions we grew up in can influence our health and lead to disparities, but these factors can be difficult for clinicians and researchers to capture and address.

A new study by investigators from Mass General Brigham demonstrates that large language models (LLMs), a type of generative artificial intelligence (AI), can be trained to automatically extract information on social determinants of health (SDoH) from clinicians' notes which could augment efforts to identify patients who may benefit from resource support.

Findings published in *npj Digital Medicine* show that the finely tuned models could identify 93.8 percent of patients with adverse SDoH, whereas official diagnostic codes included this information in only 2 percent of cases. These specialized models were less prone to bias than generalist models such as GPT-4.

"Our goal is to identify patients who could benefit from resource and social work support and draw attention to the under-documented impact of social factors in health outcomes," 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.

"Algorithms that can pass major medical exams have received a lot of attention, but this is not what doctors need in the clinic to help take better care of patients each day. Algorithms that can notice what doctors may miss in the ever-increasing volume of medical records will be more clinically relevant and therefore more powerful for improving health."

Health disparities are widely linked to SDoH, including employment, housing, and other non-medical circumstances that impact medical care.

For example, the distance a [cancer patient](link) lives from a major medical center or the support they have from a partner can substantially influence outcomes. While clinicians may summarize relevant SDoH in their visit notes, this vital information is rarely systematically organized in the electronic health record (EHR).

To create LMs capable of extracting information on SDoH, the researchers manually reviewed 800 clinician notes from 770 patients with cancer who received radiotherapy at the Department of Radiation Oncology at Brigham and Women's Hospital. They tagged sentences that referred to one or more of six pre-determined SDoH: employment status, housing, transportation, parental status (if the patient has a child under 18 years old), relationships, and presence or absence of social support.

Using this "annotated" dataset, the researchers trained existing LMs to identify references to SDoH in clinician notes. They tested their models using 400 clinic notes from patients treated with immunotherapy at Dana-Farber Cancer Institute and patients admitted to the critical care units at Beth Israel Deaconess Medical Center.

The researchers found that fine-tuned LMs, especially Flan-T5 LMs, could consistently identify rare references to SDoH in clinician notes. The "learning capacity" of these models was limited by the rarity of SDoH documentation in the training set, where the researchers found that only 3 percent of sentences in clinician notes contained any mention of SDoH.

To address this issue, the researchers used ChatGPT, another LM, to produce an additional 900 synthetic examples of SDoH sentences that could be used as an extra training dataset.

A major criticism of generative AI models in health care is that they can

potentially perpetuate bias and widen health disparities. The researchers found that their fine-tuned LM was less likely than OpenAI's GPT-4, a generalist LM, to change its determination about an SDoH based on individuals' race/ethnicity and gender.

The researchers state that it is difficult to understand how biases are formed and deconstructed—both in human and computer models. Understanding the origins of algorithmic bias is an ongoing endeavor for the researchers.

"If we don't monitor algorithmic bias when we develop and implement large language models, we could make existing health disparities much worse than they currently are," Bitterman said. "This study demonstrated that fine-tuning LMs may be a strategy to reduce algorithmic bias, but more research is needed in this area."

  More information: Large Language Models to Identify Social Determinants of Health in Electronic Health Records, *npj Digital Medicine* (2024). DOI: 10.1038/s41746-023-00970-0

Provided by Mass General Brigham