https://doi.org/10.1038/s41591-023-02700-1

Six ways large language models are changing healthcare

Nature Medicine asks six leading AI researchers to explain how LLM-powered chatbots are having an impact on health, from virtual nurses to detecting cancer progression.

The public debut in late 2022 of ChatGPT, a large language model (LLM) chatbot able to write well-informed and precise texts on just about any topic, galvanized a huge surge of interest in LLMs — including in many areas of healthcare research. Investment in LLMs is now exploding, as is research into other innovations for artificial intelligence (AI) in health.

Nature Medicine asked six leading researchers to explain how LLMs are beginning to transform healthcare.

Virtual nurses

68 million adults in the USA have two or more chronic diseases. To meet their demands for future healthcare, care capacity will have to increase at least tenfold, estimates Munjal Shah, co-founder and CEO of Hippocratic AI, based in Palo Alto, California. At present, he says, "we only have the resources to provide chronic care support to the top 20% of the sickest of the sick. Put simply, we don't have enough nurses."

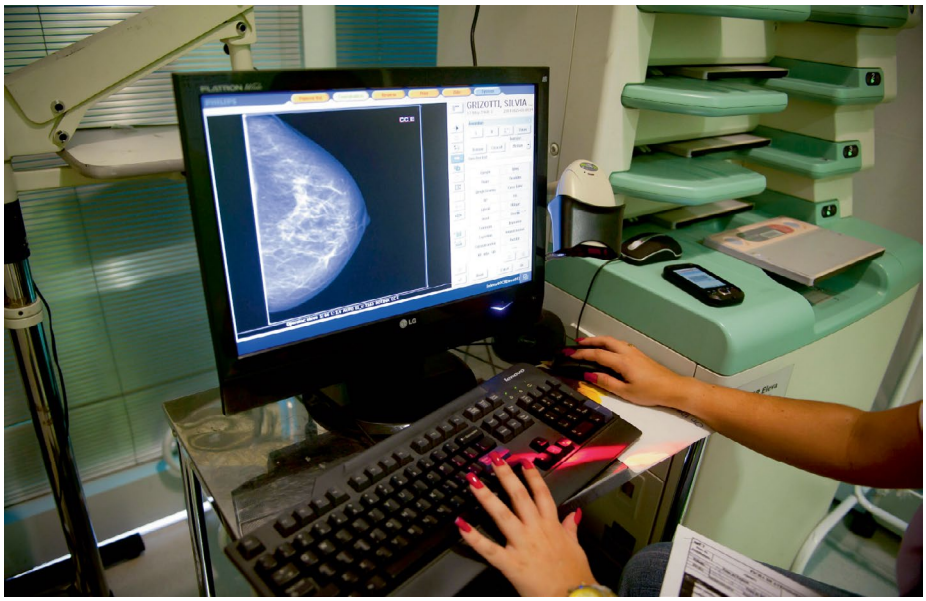
With this in mind, Shah's company (which has received US$50 million in venture capital support) is using LLMs to help nurses do their work. Such administrative support is "a safer use case for LLMs than diagnostic use cases," Shah explains, and he believes that AI tools "can be scaled to truly solve the healthcare staffing crisis."

The aim is to create virtual nurses for chronic care that use an automated 'voice' to speak to, and listen to, patients. Although the virtual nurses will not be involved in clinical diagnosis, says Shah, they could remind patients to take their medicine, follow through with care plans, schedule follow-up appointments, review medication issues and help patients navigate care-access issues.

An LLM-backed chronic-care nurse can pass the NCLEX licensing exam for nurses and the NAPLEX licensing exam for pharmacists, says Shah. An LLM nurse can "speak every language, and remember every conversation with each patient," he says.

To train its LLM nurse, Hippocratic AI built its own model using text from care plans, regulations and other medical manuals. The company then trained the model on how to speak like a chronic-care nurse using conversations between registered chronic-care nurses and patient actors.

CREDIT: ANDREY SUSLOV / ALAMY STOCK VECTOR



AI can analyze hundreds of mammogram images to predict metastatic cancer.

To safety-check its LLM nurse, Hippocratic AI has set up a safety council, and it has recruited more than a thousand nurses to evaluate the model in a double-blind randomized trial. The product, says Shah, will be released one sub-condition or procedure at a time – for example, for patients with chronic kidney disease, then for patients with arthritis, and so on.

Clinical note-taking

LLMs are poised to create huge efficiencies for clinicians by liberating them from their keyboards, says David Bates, a professor of medicine at Harvard Medical School.

"One problem we have is the inbox," says Bates. "Most physicians are now using electronic health records, and these generate a lot of email and most physicians don't have good strategies for managing this." An LLM can help categorize emails and can be trained to respond to basic messages, Bates says. "The trick is to ensure it doesn't say anything stupid."

LLMs could also prove good at checking in with patients with chronic conditions during the time between conventional encounters with caregivers, writing notes in patients' records and summarizing patients' issues. All of this "could make care much faster," says Bates.

Bates's musings are far from academic. In September, Oracle Computer (based in Austin, Texas) announced that it is poised to market an AI-powered voice- and screen-based Clinical Digital Assistant that can conduct administrative tasks in response to conversational voice commands. Oracle says it aims to make its digital assistant available by the end of 2024.

Adverse-event detection

Vivek Rudrapatna is a gastroenterologist at the University of California, San Francisco, who has been working with the US Food and Drug Administration on unearthing adverse-event data from clinical notes within electronic health record systems. "Clinical notes have been underutilized because we lack tools to effectively examine them," says Rudrapatna.

Thanks to LLMs, this is no longer the case, Rudrapatna enthuses. There have been impressive advances in understanding natural language processing, initially following the release of context-aware deep learning models such as BERT (bidirectional encoder representations from transformers) and now using GPT-4. However, the presence of identifiable patient information in clinical notes has hindered research in this area.

Rudrapatna tackled this challenge with the automated redaction of protected health information from 75 million clinical notes, together with the extraction of serious adverse events[1]. The BERT model that Rudrapatna and his team used was fine-tuned to extract only events that resulted in hospitalization and were also associated with the use of biologic immunosuppressants for inflammatory bowel disease – Rudrapatna's specialty.

Rudrapatna is now hopeful that LLMs can automate the detection of adverse events from electronic health records and "may someday support the surveillance of drug safety in the post-marketing setting using new data sources," he says.

Predicting cancer metastasis

LLMs are being used to predict metastatic cancer and to assist in devising clinical treatment responses, says Amber Simpson, Canada Research Chair in Biomedical Computing and Informatics at Queens University, Canada, who collaborates with a research team at Memorial Sloan Kettering Cancer Center in New York City.

Her research on LLM-assisted cancer diagnosis includes a set of studies in which cancer-progression patterns were extracted from vast numbers of computed tomography reports (714,454 structured radiology reports were used in the most recent study[2]) and were analyzed by a natural language processing model that was then used to predict metastatic disease in multiple organs. The models used features from consecutive structured text radiology reports to predict the presence of metastatic disease.

By offering oncologists a predictive pathway of how cancer will progress, the LLM shifts treatment strategies toward a more targeted, deliberate approach.

"Currently, we treat everyone to help a few," Simpson explains. "With this type of model, if you add the treatment data, you can start saying the next treatment this patient should receive is this specific therapy – and that's really the holy grail of precision medicine."

Such an approach helps to elucidate cancer drugs beyond clinical trials, in real-world settings, Simpson adds. "This type of analysis allows us to understand population-level analysis of cancer patients and how drugs work in the wild."

Unusually for an AI model, this is already in clinical use. "We're already showing this approach can be used clinically. We're already targeting patients," says Simpson.

Social determinants of health

LLMs can deliver useful information on the social determinants of health to clinicians, says Maxim Topaz, a professor of nursing at Columbia University. He has spent years refining LLM-based methods that deliver hard-to-find, unstructured, non-medical information that has an impact on health.

His team is embedded within VNS Health, a New York City homecare services company with almost 50,000 patients. "We work with clinicians, nurses, occupational therapists, social workers, speech therapists," he explains. His team wants to understand how information on the social determinants of health will be used and what will be helpful for different specialists.

In a 2022 study, Topaz developed and validated a natural language processing algorithm to identify information related to the social determinants of health that was not routinely captured in structured electronic health records[3]. The tool was used to identify patients who lacked social support and who were likely to become incapacitated. The algorithm analyzed electronic clinical notes on case management, consultations, discharge summaries, nursing, nutrition, rehabilitation, social work, admissions, deaths, procedures and progress notes.

Using LLMs to analyze social determinants of health is likely to be applied in healthcare fields beyond nursing. A February 2023 National Institutes of Health Workshop Report emphasized that "computational approaches to atrial fibrillation must include social determinants of health. These non-medical factors must be tackled in order to eliminate inequities in healthcare delivery and outcomes."

Topaz and his research team are poised to begin implementing LLMs in clinical settings "within a year," he says.

Conversational AI diagnostics

LLMs will soon be built into predictive systems that are seamlessly integrated into clinical practice and "offer extremely high accuracy in the medical imaging domain," says Greg Corrado, a research director at Google AI. Many doctors yearn for easy access to a user-friendly system that can examine and explain medical images, he believes.

Several retrospective studies (including one authored by Corrado and colleagues) and a large prospective study published in 2023 (ref. 4) have shown that AI can analyze mammogram images for breast cancer with diagnostic accuracy. However, Corrado believes AI can go further, by integrating screening and diagnosis with an LLM-powered chatbot.

This integration of AI tools would allow the clinician to "have a conversation with the system, ask it questions, ask it for help." Such a tool could also be used to draft a report, or improve an existing report, he says.

This approach, he explains, would take the representation that the medical imaging foundation model is using to make its determinations and its recommendations and "re-express that representation as a sequence of vectors that the language model naturally understands and is able to be conversant about," he says.

The technical means of welding LLMs into clinical systems such as those that accurately read mammograms have become much more doable over the past year, says Corrado.

"I would say that I think we're not there yet," he muses, "but you can see the early footprints on this path."

Paul Webster
Science journalist

References

1. Sushil, M., Ludwig, D., Butte, A. J. & Rudrapatna, V. A. Preprint at https://doi.org/10.48550/arXiv.2210.06566 (2022).
2. Batch, K. E. et al. *Front. Artif. Intell.* **5**, 826402 (2022).
3. Song, J. et al. *PLoS ONE* **17**, e0270220 (2022).
4. Dembrower, K. et al. *Lancet Digit. Health* **5**, E703–E711 (2023).