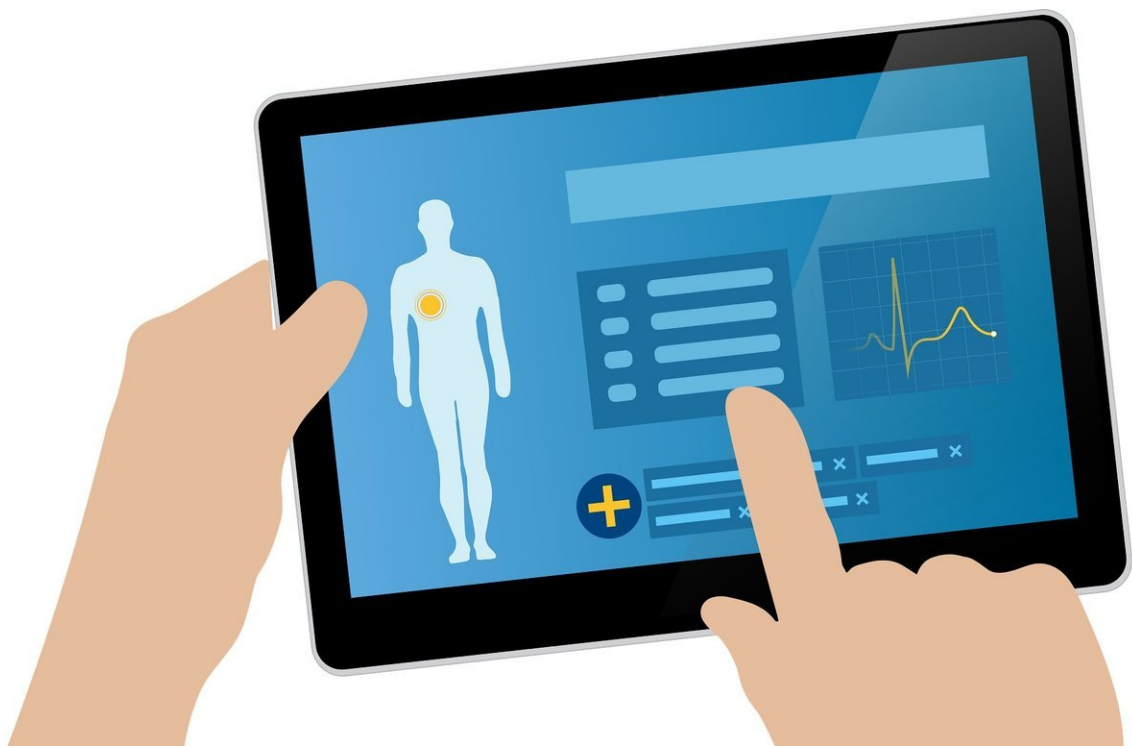# Deep-learning model extracts important data from health records to assist with personalized medicine

December 1 2022, by Rachel Gordon

Credit: CC0 Public Domain

Electronic health records (EHRs) need a new public relations manager. Ten years ago, the U.S. government passed a law that required hospitals to digitize their health records with the intent of improving and

streamlining care. The enormous amount of information in these now-digital records could be used to answer very specific questions beyond the scope of clinical trials: What's the right dose of this medication for patients with this height and weight? What about patients with a specific genomic profile?

Unfortunately, most of the data that could answer these questions is trapped in doctor's notes, full of jargon and abbreviations. These notes are hard for computers to understand using current techniques—extracting information requires training multiple machine learning models. Models trained for one hospital, also, don't work well at others, and training each model requires domain experts to label lots of data, a time-consuming and expensive process.

An ideal system would use a single model that can extract many types of information, work well at multiple hospitals, and learn from a small amount of labeled data. But how? Researchers from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) believed that to disentangle the data, they needed to call on something bigger: large language models. To pull that important medical information, they used a very big, GPT-3 style model to do tasks like expand overloaded jargon and acronyms and extract medication regimens.

For example, the system takes an input, which in this case is a clinical note, "prompts" the model with a question about the note, such as "expand this abbreviation, C-T-A." The system returns an output such as "clear to auscultation," as opposed to say, a CT angiography. The objective of extracting this clean data, the team says, is to eventually enable more personalized clinical recommendations.

Medical data is, understandably, a pretty tricky resource to navigate freely. There's plenty of red tape around using public resources for testing the performance of large models because of data use restrictions,

so the team decided to scrape together their own. Using a set of short, publicly available clinical snippets, they cobbled together a small dataset to enable evaluation of the extraction performance of large language models.

"It's challenging to develop a single general-purpose clinical natural language processing system that will solve everyone's needs and be robust to the huge variation seen across health datasets. As a result, until today, most clinical notes are not used in downstream analyses or for live decision support in [electronic health records](). These large language model approaches could potentially transform clinical natural language processing," says David Sontag, MIT professor of electrical engineering and computer science, principal investigator in CSAIL and the Institute for Medical Engineering and Science, and supervising author on a paper about the work, which will be presented at the Conference on Empirical Methods in Natural Language Processing.

"The research team's advances in zero-shot clinical information extraction makes scaling possible. Even if you have hundreds of different use cases, no problem—you can build each model with a few minutes of work, versus having to label a ton of data for that particular task."

For example, without any labels at all, the researchers found these models could achieve 86% accuracy at expanding overloaded acronyms, and the team developed additional methods to boost this further to 90% accuracy, with still no labels required.

Imprisoned in an EHR

Experts have been steadily building up large language models (LLMs) for quite some time, but they burst onto the mainstream with GPT-3's widely covered ability to complete sentences. These LLMs are trained on

a huge amount of text from the internet to finish sentences and predict the next most likely word.

While previous, smaller models like earlier GPT iterations or BERT have pulled off a good performance for extracting medical data, they still require substantial manual data-labeling effort.

For example, a note, "pt will dc vanco due to n/v" means that this patient (pt) was taking the antibiotic vancomycin (vanco) but experienced nausea and vomiting (n/v) severe enough for the care team to discontinue (dc) the medication. The team's research avoids the status quo of training separate machine learning models for each task (extracting medication, side effects from the record, disambiguating common abbreviations, etc). In addition to expanding abbreviations, they investigated four other tasks, including if the models could parse clinical trials and extract detail-rich medication regimens.

"Prior work has shown that these models are sensitive to the prompt's precise phrasing. Part of our technical contribution is a way to format the prompt so that the model gives you outputs in the correct format," says Hunter Lang, CSAIL Ph.D. student and author on the paper.

"For these extraction problems, there are structured output spaces. The output space is not just a string. It can be a list. It can be a quote from the original input. So there's more structure than just free text. Part of our research contribution is encouraging the model to give you an output with the correct structure. That significantly cuts down on post-processing time."

The approach can't be applied to out-of-the-box health data at a hospital: that requires sending private patient information across the open internet to an LLM provider like OpenAI. The authors showed that it's possible to work around this by distilling the model into a smaller one that could

be used on-site.

The model—sometimes just like humans—is not always beholden to the truth. Here's what a potential problem might look like: Let's say you're asking the reason why someone took medication. Without proper guardrails and checks, the model might just output the most common reason for that medication, if nothing is explicitly mentioned in the note. This led to the team's efforts to force the model to extract more quotes from data and less free text.

Future work for the team includes extending to languages other than English, creating additional methods for quantifying uncertainty in the model, and pulling off similar results with open-sourced models.

"Clinical information buried in unstructured clinical notes has unique challenges compared to general domain text mostly due to large use of acronyms, and inconsistent textual patterns used across different health care facilities," says Sadid Hasan, AI lead at Microsoft and former executive director of AI at CVS Health, who was not involved in the research.

"To this end, this work sets forth an interesting paradigm of leveraging the power of general domain large language models for several important zero-/few-shot clinical NLP tasks. Specifically, the proposed guided prompt design of LLMs to generate more structured outputs could lead to further developing smaller deployable models by iteratively utilizing the model generated pseudo-labels."

"AI has accelerated in the last five years to the point at which these large models can predict contextualized recommendations with benefits rippling out across a variety of domains such as suggesting novel drug formulations, understanding unstructured text, code recommendations or create works of art inspired by any number of human artists or styles,"

says Parminder Bhatia, who was formerly Head of Machine Learning at AWS Health AI and is currently Head of ML for low-code applications leveraging large language models at AWS AI Labs. "One of the applications of these large models [the team has] recently launched is Amazon CodeWhisperer, which is [an] ML-powered coding companion that helps developers in building applications."

More information: Monica Agrawal et al, Large Language Models are Zero-Shot Clinical Information Extractors, *arXiv* (2022). [DOI: 10.48550/arxiv.2205.12689](#)

This story is republished courtesy of MIT News ([web.mit.edu/newsoffice/](#)), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology