

# Machine-learning models that can help doctors more efficiently find information in a patient's health record

July 14 2022, by Adam Zewe

---



Credit: CC0 Public Domain

Physicians often query a patient's electronic health record for information that helps them make treatment decisions, but the cumbersome nature of these records hampers the process. Research has

shown that even when a doctor has been trained to use an electronic health record (EHR), finding an answer to just one question can take, on average, more than eight minutes.

The more time physicians must spend navigating an oftentimes clunky EHR interface, the less time they have to interact with patients and provide treatment.

Researchers have begun developing machine-learning models that can streamline the process by automatically finding information physicians need in an EHR. However, training effective models requires huge datasets of relevant medical questions, which are often hard to come by due to privacy restrictions. Existing models struggle to generate authentic questions—those that would be asked by a human doctor—and are often unable to successfully find [correct answers](#).

To overcome this data shortage, researchers at MIT partnered with medical experts to study the questions physicians ask when reviewing EHRs. Then, they built a [publicly available dataset](#) of more than 2,000 clinically relevant questions written by these medical experts.

When they used their dataset to train a machine-learning model to generate clinical questions, they found that the model asked high-quality and authentic questions, as compared to real questions from medical experts, more than 60% of the time.

With this dataset, they plan to generate vast numbers of authentic medical questions and then use those questions to train a machine-learning model which would help doctors find sought-after information in a patient's record more efficiently.

"Two thousand questions may sound like a lot, but when you look at machine-learning models being trained nowadays, they have so much

data, maybe billions of [data points](#). When you train machine-learning models to work in health care settings, you have to be really creative because there is such a lack of data," says lead author Eric Lehman, a graduate student in the Computer Science and Artificial Intelligence Laboratory (CSAIL).

The senior author is Peter Szolovits, a professor in the Department of Electrical Engineering and Computer Science (EECS) who heads the Clinical Decision-Making Group in CSAIL and is also a member of the MIT-IBM Watson AI Lab. The research paper, a collaboration between co-authors at MIT, the MIT-IBM Watson AI Lab, IBM Research, and the doctors and medical experts who helped create questions and participated in the study, will be presented at the annual conference of the North American Chapter of the Association for Computational Linguistics.

"Realistic data is critical for training models that are relevant to the task yet difficult to find or create," Szolovits says. "The value of this work is in carefully collecting questions asked by clinicians about patient cases, from which we are able to develop methods that use these data and general language models to ask further plausible questions."

## **Data deficiency**

The few large datasets of clinical questions the researchers were able to find had a host of issues, Lehman explains. Some were composed of medical questions asked by patients on web forums, which are a far cry from physician questions. Other datasets contained questions produced from templates, so they are mostly identical in structure, making many questions unrealistic.

"Collecting high-quality data is really important for doing machine-learning tasks, especially in a health care context, and we've shown that it

can be done," Lehman says.

To build their dataset, the MIT researchers worked with practicing physicians and medical students in their last year of training. They gave these medical experts more than 100 EHR discharge summaries and told them to read through a summary and ask any questions they might have. The researchers didn't put any restrictions on question types or structures in an effort to gather natural questions. They also asked the medical experts to identify the "trigger text" in the EHR that led them to ask each question.

For instance, a medical expert might read a note in the EHR that says a patient's past medical history is significant for prostate cancer and hypothyroidism. The trigger text "prostate cancer" could lead the expert to ask questions like "date of diagnosis?" or "any interventions done?"

They found that most questions focused on symptoms, treatments, or the patient's test results. While these findings weren't unexpected, quantifying the number of questions about each broad topic will help them build an effective dataset for use in a real, clinical setting, says Lehman.

Once they had compiled their dataset of questions and accompanying trigger text, they used it to train [machine-learning models](#) to ask new questions based on the trigger text.

Then the medical experts determined whether those questions were "good" using four metrics: understandability (Does the question make sense to a human physician?), triviality (Is the question too easily answerable from the trigger text?), medical relevance (Does it make sense to ask this question based on the context?), and relevancy to the trigger (Is the trigger related to the question?).

## Cause for concern

The researchers found that when a model was given trigger text, it was able to generate a good question 63% of the time, whereas a human physician would ask a good question 80% of the time.

They also trained models to recover answers to clinical questions using the publicly available datasets they had found at the outset of this project. Then they tested these trained models to see if they could find answers to "good" questions asked by human [medical experts](#).

The models were only able to recover about 25% of answers to physician-generated questions.

"That result is really concerning. What people thought were good-performing models were, in practice, just awful because the evaluation questions they were testing on were not good to begin with," Lehman says.

The team is now applying this work toward their initial goal: building a model that can automatically answer physicians' questions in an EHR. For the next step, they will use their dataset to train a machine-learning model that can automatically generate thousands or millions of good clinical questions, which can then be used to train a new model for automatic question answering.

While there is still much work to do before that model could be a reality, Lehman is encouraged by the strong initial results the team demonstrated with this [dataset](#).

**More information:** Eric Lehman et al, Learning to Ask Like a Physician. arXiv:2206.02696v1 [cs.CL], [arxiv.org/abs/2206.02696](https://arxiv.org/abs/2206.02696)

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

Provided by Massachusetts Institute of Technology

Citation: Machine-learning models that can help doctors more efficiently find information in a patient's health record (2022, July 14) retrieved 19 April 2024 from <https://techxplore.com/news/2022-07-machine-learning-doctors-efficiently-patient-health.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.