

Gamifying medical data labeling to advance AI

June 28 2023, by Zach Winn



Credit: CC0 Public Domain

When Erik Duhaime, Ph.D. was working on his thesis in MIT's Center for Collective Intelligence, he noticed his wife, then a medical student, spending hours studying on apps that offered flash cards and quizzes. His research had shown that as a group, medical students could classify skin lesions more accurately than professional dermatologists; the trick was to continually measure each student's performance on cases with known answers, throw out the opinions of people who were bad at the task, and intelligently pool the opinions of people who were good.

Combining his wife's studying habits with his research, Duhaime

founded Centaur Labs, a company that created a mobile app called DiagnosUs to gather the opinions of medical experts on real-world scientific and [biomedical data](#). Through the app, users review anything from images of potentially cancerous skin lesions or audio clips of heart and lung sounds that could indicate a problem. If the users are accurate, Centaur uses their opinions and awards them small cash prizes. Those opinions, in turn, help medical AI companies train and improve their algorithms.

The approach combines the desire of medical experts to hone their skills with the desperate need for well-labeled [medical data](#) by companies using AI for biotech, developing pharmaceuticals, or commercializing medical devices.

"I realized my wife's studying could be productive work for AI developers," Duhaime recalls. "Today we have tens of thousands of people using our app, and about half are medical students who are blown away that they win money in the process of studying. So, we have this gamified platform where people are competing with each other to train data and winning money if they're good and improving their skills at the same time—and by doing that they are labeling data for teams building life saving AI."

Gamifying medical labeling

Duhaime completed his Ph.D. under Thomas Malone, the Patrick J. McGovern Professor of Management and founding director of the Center for Collective Intelligence at MIT.

"What interested me was the wisdom of crowds phenomenon," Duhaime says. "Ask a bunch of people how many jelly beans are in a jar, and the average of everybody's answer is pretty close. I was interested in how you navigate that problem in a task that requires skill or expertise."

Obviously you don't just want to ask a bunch of random people if you have cancer, but at the same time, we know that second opinions in health care can be extremely valuable. You can think of our platform as a supercharged way of getting a second opinion."

Duhaime began exploring ways to leverage [collective intelligence](#) to improve medical diagnoses. In one experiment, he trained groups of lay people and medical school students that he describes as "semi-experts" to classify skin conditions, finding that by combining the opinions of the highest performers he could outperform professional dermatologists. He also found that by combining algorithms trained to detect skin cancer with the opinions of experts, he could outperform either method on its own.

"The core insight was you do two things," Duhaime explains. "The first thing is to measure people's performance—which sounds obvious, but even in the medical domain it isn't done much. If you ask a dermatologist if they're good, they say, 'Yeah of course, I'm a dermatologist.' They don't necessarily know how good they are at specific tasks. The second thing is that when you get multiple opinions, you need to identify complementarities between the different people. You need to recognize that expertise is multidimensional, so it's a little more like putting together the optimal trivia team than it is getting the five people who are all the best at the same thing. For example, one dermatologist might be better at identifying melanoma, whereas another might be better at classifying the severity of psoriasis."

While still pursuing his Ph.D., Duhaime founded Centaur and began using MIT's entrepreneurial ecosystem to further develop the idea. He participated in the delta v startup accelerator run by the Martin Trust Center for MIT Entrepreneurship over the summer of 2018. The experience helped him get into the prestigious Y Combinator accelerator later that year.

The DiagnosUs app, which Duhaime developed with Centaur co-founders Zach Rausnitz and Tom Gellatly, is designed to help users test and improve their skills. Duhaime says about half of users are medical school students and the other half are mostly doctors, nurses, and other medical professionals.

"It's better than studying for exams, where you might have multiple choice questions," Duhaime says. "They get to see actual cases and practice."

Centaur gathers millions of opinions every week from tens of thousands of people around the world. Duhaime says most people earn coffee money, although the person who's earned the most from the platform is a doctor in eastern Europe who's made around \$10,000.

"People can do it on the couch, they can do it on the T," Duhaime says. "It doesn't feel like work—it's fun."

The approach stands in sharp contrast to traditional data labeling and AI content moderation, which are typically outsourced to low-resource countries.

Centaur's approach produces accurate results, too. In a paper published on the pre-print server *arXiv* with researchers from Brigham and Women's Hospital, Massachusetts General Hospital (MGH), and Eindhoven University of Technology, Centaur showed its crowdsourced opinions labeled lung ultrasounds as reliably as experts did. Another study with researchers at Memorial Sloan Kettering showed crowdsourced labeling of dermoscopic images was more accurate than that of highly experienced dermatologists. Beyond images, Centaur's platform also works with video, audio, text from sources like research papers or anonymized conversations between doctors and patients, and waves from electroencephalograms (EEGs) and electrocardiographs

(ECGs).

Finding the experts

Centaur has found that the best performers come from surprising places. In 2021, to collect expert opinions on EEG patterns, researchers held a contest through the DiagnosUs app at a conference featuring about 50 epileptologists, each with more than 10 years of experience. The organizers made a custom shirt to give to the contest's winner, whom they assumed would be in attendance at the conference.

But when the results came in, a pair of medical students in Ghana, Jeffery Danquah and Andrews Gyabaah, had beaten everyone in attendance. The highest-ranked conference attendee had come in ninth.

"I started by doing it for the money, but I realized it actually started helping me a lot," Gyabaah told Centaur's team later. "There were times in the clinic where I realized that I was doing better than others because of what I learned on the DiagnosUs app."

As AI continues to change the nature of work, Duhaime believes Centaur Labs will be used as an ongoing check on AI models.

"Right now, we're helping people train algorithms primarily, but increasingly I think we'll be used for monitoring algorithms and in conjunction with algorithms, basically serving as the humans in the loop for a range of tasks," Duhaime says. "You might think of us less as a way to train AI and more as a part of the full life cycle, where we're providing feedback on models' outputs or monitoring the model."

Duhaime sees the work of humans and AI algorithms becoming increasingly integrated and believes Centaur Labs has an important role to play in that future.

"It's not just train algorithm, deploy [algorithm](#)," Duhaime says. "Instead, there will be these digital assembly lines all throughout the economy, and you need on-demand expert human judgment infused in different places along the value chain."

More information: Nicole M Duggan et al, Gamified Crowdsourcing as a Novel Approach to Lung Ultrasound Dataset Labeling, *arXiv* (2023). [DOI: 10.48550/arxiv.2306.06773](https://doi.org/10.48550/arxiv.2306.06773)

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

Citation: Gamifying medical data labeling to advance AI (2023, June 28) retrieved 13 May 2024 from <https://medicalxpress.com/news/2023-06-gamifying-medical-advance-ai.html>

<p>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.</p>
--