MIT CSAIL Alliances | John Guttag Export 3

Welcome to MIT's Computer Science and Artificial Intelligence Labs, Alliance's podcast series. My name is Steve Lewis. I'm the assistant director of Global Strategic Alliances for CSAIL at MIT. In this podcast series, I will interview principal researchers at CSAIL to discover what they're working on and how it will impact society.

Professor John Guttag leads CSAIL Clinical and Applied Machine Learning Group. The group uses advanced machine learning and computer vision to improve outcomes in medicine, finance, and sports. Current research projects include prediction and reduction of adverse medical events, matching patients to therapies and providers, and medical imaging.

Professor Guttag has also done research, published, and lectured in the areas of sports analytics, financial analytics, software defined radios, software engineering, mechanical theorem proving, and hardware verification. In addition to his academic activities, Professor Guttag has extensive industrial experience.

He is currently chairman of the board of directors and chief scientists of Health at Scale Technologies and on the board of directors of Frontier. Professor Guttag received a bachelor's degree in English from Brown University in 1971 and a master's degree in applied mathematics from Brown in 1972. In 1975 he received his doctorate in computer science from the University of Toronto.

Professor Guttag joined the Faculty of MIT in 1979 and he is a fellow of the ACM and a member of the American Academy of Arts and Sciences. In recent years, his classroom teaching has centered around helping students apply computational modes of thought to frame problems and guide the process of extracting useful information from data.

His textbook on the topic introduction, to computation and programming using Python with application to understanding data is used in online courses that have been taken by over a million students. All right, John. Thank you very much for your time. It's a pleasure meeting you. We have a few questions that we want you to talk about. There's been a lot of buzz around AI and health care. How much of it is hype?

Wow. That's an interesting place to start. How much of it is hype? Maybe 98.7%. Mostly hype, I would say. It's been disappointing to me. It's not all hype, there have been some successes. Hospitals have done a good job of using AI to up code and build more. Biotech has made good progress with it. But for the most part, it's been kind of discouraging. There's been a lot less impact than if you'd asked me 10 years ago, we would have than we actually had.

And why do you think that is? I mean, why haven't the big players in health care been able to do more with the AI?

Yeah. So Watson it's a case in point, right? A lot of hype, a lot of money. Now they're selling it off. I think part of it. There are a lot of different things. So the big players are for the most part, really good at handling big data.

And I think naively some of them thought they could just move that expertise to health care. When in fact, most of the important problems in health care end up, I think being small data problems because they have to do with making good decisions for sub-population, sometimes for individuals.

And the techniques that are work on fine on average don't necessarily work very well for individuals. I had the same by the way, misconception when I started working in this area. I thought it was going to be all about big data and finding clever ways to handle large amounts of data.

And one of my early experiences I was working with chronic kidney disease and I got a data set of, I think, 3 million people, I thought this is going to be great. And then I discovered well, most of them didn't actually have chronic kidney disease.

So 3 million got a lot smaller very quickly. And then I discovered OK, it actually manifests pretty differently in African-Americans and in others. And so you then have to start partitioning the data. And by the time I finished partitioning every set was too small.

And so the technical problems I had to deal with were about sample efficiency, not about managing large amounts of data. And that's very different from what most of-- what Google say has to worry about. And so I think that's a piece of it.

And then their economics. The big technology players, health care at least for now it doesn't move the needle on their businesses. The big health care players have a different problem, which is their people.

If you look at the number of PhDs in computer science at the large health care companies. And I mean health care writ large, ranging from providers to insurers. They're very few people who really have PhDs and say machine learning from good institutions.

Partly they can't compete with the Google's and the Facebooks and the Amazons in the hiring market. And so they just don't have the talent they need to do this. And yet they have very strong vested interests in not invented here. And so it's a bad mismatch that you need to invent it yourself and you don't have the people capable of inventing it.

This is quite a dilemma. But let me just step back a little bit. You said the companies like Google, for example are not good at doing-- they're good at big data but not good at, I guess finite data. But they're so good it would seem like personalization.

Where it looks like they have these tools that they have but they're catering to our individual needs, be it recommending products we buy or some type of personalization engine. I guess it's not applicable in health care?

I wouldn't say it's not applicable. It might be applicable. They do of course from, say, our search history have a lot of information about each of us. And I think they do a remarkably good job of personalization. But they've somehow been unable despite a lot of efforts to convert it to anything that has had a really meaningful impact in the health care space. Do I know why? Not really. Maybe it's that they make so much money serving ads, that's where they put their bandwidth doing that better.

Could be for sure. And when you talk about big data in general, most of the topic is around health care records, for example. Is that what you're focusing on as a more big data, meaning the data sets that you're training for machine learning algorithms that are able to identify certain pathologies?

Yeah. So I have to answer that question in two different ways, because I wear two different hats. My hat I wear most of the time is a Professor at MIT EECS and CSAIL. And the other hat I wear is chief technology officer of a health care company called Health at Scale. And I do different things with those two hats. So I'm going to give you two separate answers.

With the MIT hat on, where of course, the hat there is long range things. I do a lot of work with medical images. And I would say most of what we do is about medicine. We work with medical images, we work with signals, for example ECG, we work with electronic health records.

We've done a lot of work, say with MGH is infection control department. Trying to understand the spread of infections. And so there it's been something very Catholic in the data. At the company, it's been quite different. So there the company is really about health care, not medicine.

And I think a lot of people-- there are three things. There's biology, there's medicine, there's health care. Health care involves biology and medicine, but it's not the same. When we work in the company, we're working almost entirely with billing data.

I was very skeptical about that because it's course, it's noisy. The advantage of it is it's copious. We have billing data literally complete. Many years of billing data on well over 100 million Americans. And billing data on every provider in the country.

And you can't get the other data modalities that way. And we work on things that you can do with that. Now, interestingly, from the billing data, if you do your machine learning correctly, you can get a pretty good model of someone's health. Just for what has been billed.

Some of it simple, some of it, it's not simple. If you see someone has been filling prescriptions for insulin, make a pretty good guess they have diabetes. Other things are much more complex, but still there's enough data you can do a pretty good job.

And now the question is, what do you do? So one of the things we do is we match individuals to physicians. If you think about it, most people will spend more time choosing the restaurant where they're going to go to dinner than they will choosing their doctor. And it's shocking when you think about it, which is more which is a more important decision.

Now why is that? Well, mostly they. They don't have enough information to make a good choice about a doctor. Yet if you look at the data, it makes an enormous difference. The first step in good health care outcomes is the right physicians, or the right facilities. What nursing facility do you go to, for example for skilled nursing care? What physical therapy? All of those things.

And interestingly, one might have thought that there was a nice ordering on doctors, for example. That OK, these doctors are good, and these doctors are not, or these doctors are good at this procedure. Turns out it isn't. It's actually the patient-doctor pairing that matters.

So you could think about it, it's like match.com. The idea is this a great person, are these two good people a good match. And doctors do well or not well with different classes of patients. We've looked, for example, very carefully at orthopedic surgeries and oncology and actually about 30 different specialties.

I'm particularly interested in orthopedics because I've had more than my fair share of orthopedic procedures. And will find that say some doctors do very well with 60-year-old men and terribly with 18-year-old women. And some are vise versa.

Some doctors will do well with obese patients, other doctors are disaster with obese patients, but good with non-obese patients. Some doctors for some unknown reason will struggle with people with a history of cardiac problems.

And so this is a place where health care and medicine are different. So the health care, you can dramatically reduce bad outcomes by matching people to the right doctors. And you can do that using billing history. But it's a small data problem. How many knee replacements does an individual surgeon do every year, not that many. How many do they do on 17-year-old girls, maybe zero, maybe one.

And so you have to use fairly sophisticated methods of transfer learning and multi-task learning to be able to build reliable models in just getting back to the small data question, in the absence of copious amounts of data. It can make a huge difference in outcomes. You can reduce bad outcomes, sometimes by as much as 18% by choosing the right doctor.

Is this something that Health at Scale does differently than other AI health companies? This personalization is matching of the patient to the doctor and--

I don't know of any other company that really does it in a personalized way. There are rankings. So there are things like these are the good doctors, but they're just rankings. And they're based on population averages, if anything.

And they don't correlate very well with actual outcomes on patients. And it's hard to do. You need the data, you need the technology. Fortunately, we've been doing it long enough that-- we've been in and we figured it out. It works a lot better now than it did four years ago.

Without giving away any trade secrets but I mean, how are you able to do this without a survey with additional data points of, do you like your doctor?

That's a great question. So this is not about patient satisfaction. What we decided very early on is to focus on outcomes. So if you think about such surgeries or any kind of a visit, even especially primary care. We match the primary care physicians a lot.

What you're interested in is risk adjusted outcomes. Given say your medical history, what's the probability, how long are you likely to stay in the hospital? Are you likely to go home or are you likely to go to a rehab facility, and say you're getting joint replacement? What about surgical site infections?

So there you can think about all the things that go wrong. And that's really what people should worry about. In general, what you don't want is a disaster. Tom Brady all people, gets ACL repair and gets a surgical site infection. Surely he had access to any doctor he wanted. But I'm not blaming the doctor. These things are stochastic. Maybe everything was great and he just was unlucky.

But those are the things that you worry about. And so we focus entirely on outcomes and try and do matches that reduce the probability of bad outcomes. There are other things you could look at patient satisfaction, but that's very hard to measure. you could look at cost, which of course cares matters a lot to a lot of folks.

The interesting thing there is reducing bad outcomes reduces cost. But what drives medical costs is not the first bill, it's the subsequent bills when something goes wrong. It's when you go in for which should have been a routine intestinal surgery and you get a blockage and end up spending 10 days in the hospital. And that blows things out of the water. And so that's what you really have to always worry about is avoiding the bad parts of the distribution.

Interesting. Can you talk a little bit about your work towards improving medical imaging?

Yeah. So we do a lot of that at MIT. And it's an increasingly important modality. And I shouldn't say modality, because it's a raft of modalities. MRIs and X-rays, and echocardiograms, are all images, but they're very different from each other.

They're used for a lot of things, certainly for diagnosis. We've recently been looking at using chest X-rays to predict COVID outcomes, for example, which is an interesting and, I think useful exercise. There are a couple of aspects to what we've been doing in that area. The obvious one is making better predictions.

But most of our effort has been on what I would call the blocking and tackling of imaging, the infrastructures. So there are things like if you want to do good work with images, the first thing you typically have to do is register them. That's match them up.

If I want to see how does your brain look today compared to the way it looked two years ago, I need to take those two MRIs and overlay them in a way that lets me compare them. Sounds easy. It's actually quite hard.

And one of the things we've been interested in is making it fast. A lot of the applications of medical imaging people know how to solve the problem, but it takes forever. And therefore, they don't do it sure we could do it if we had infinite computing power and infinite time but we don't.

And so a lot of what we've done is things like can we make registration much faster. Can we make segmentation much faster, segmentation involves finding the say in the brain the different structures. And so a lot of our effort has gone on providing basic tools that people can build on top of to then do specific tasks. And so we spent a lot of time on registration, a lot of time on segmentation.

We are interested in specific tasks as well, but most of our effort has been below that level. So it doesn't sound as sexy as a lot of things, but we think it will actually have a big impact in the long run. Particularly, not only on clinical care, but on research.

If it takes two hours to register a set of brain, a pair of brains, that might be OK if you only had one pair of brains. But if you've got 10,000 patients, then you're trying to do a population study, and now you have to do 10,000 squared registrations, each of which takes two hours where you end up not doing it.

And so we think that a lot of this technology will be most fruitful for advancing medical science. At the other extreme, we have been very interested in clinical images. A lot of the research on images has been done on what I'll call research images, where you can acquire a beautiful image in the lab.

Maybe put someone in a scanner for an hour. Well, someone who's had a stroke you're not going to put them in a scanner for an hour to get a beautiful image. And so the question is what can you do with these actual clinical images. Same thing with X-rays.

Most people think you go into a lab to get an X-ray, you stand up, they say, hold your breath, they take a beautiful image. Well, if you've got COVID, you're in a hospital bed, you probably can't hold your breath. And you're not standing, you're in the bed. And so the image looks nothing like the images you get in the lab. And so a lot of our interest is looking at the actual images you get in the clinic. And what can you do with those.

Well, can you talk about your experience with bridging the lab or academia with industry?

Yeah. Let me expand upon it a little bit the question. In that I think, there are three legs in this stool at least for me. There's academia and I couldn't be in a better place but I am, so I'm very fortunate to be here.

There's also practicing clinicians. So we have over the years in my lab and we've been doing medical stuff now for over 20 years. We've worked with MGA, we've worked with the Brigham, we've worked with Children's Hospital. It's always been critical. Dealing with people who have their feet on the ground and are dealing with patients is hugely important.

So this is not the research arms of these hospitals, it's the people who are actually treating patients that I like to work with. And that has been invaluable to me and certainly to my students as well to say, OK, here's what really happens.

And then the third, which took me a long time to come to is commercialization. And the reason I started this company with former MIT PhD students out of frustration. That we've been writing a bunch of papers, which I thought were really cool and nothing was changing.

And the truth is if you really want to have an impact, you need to connect to industry. We had lots of patents. Some of them even got licensed, but none of them actually made it into the real world. And it prompted us, let's actually try and have an impact.

And I think it's really important. And there's an interplay. And I'm often talking to my students here and they'll ask something and it will prompt me to give them a little lecture about life in the real world. And say, OK, here's what we should really be thinking about. Populations are different.

It's great to have a model that works on average, but what about sub-populations? How well does it work on women, if most of the data are men? African-Americans. There are thing people present differently. And I think spending time in industry and seeing what actually happens really has helped inform research directions here.

That said, we have to be very careful to avoid conflicts of interest. I would never hire one of my current graduate students to work as an intern at the company because that would be a conflict. We do have some former students at the company, but none of them were hired directly out of MIT. They all did something else first. And I think it's just something we need to be cognizant of as faculty members to not confuse what's good for the student for what's good for our companies.

Well, is there anything else that you're excited to share with our listeners? An open ended question you could--

Yeah, I think because of the research I do, I talk to a lot of students who are interested in medicine and health and biology and things like that. The message I try and give them is that as computer scientists, we have an enormous opportunity to improve people's health. Improve the health care system.

And we shouldn't forget that. It's an opportunity, it's a responsibility. I think computer scientists will probably have more impact on healthcare than any other field over the next decade to two decades. I certainly hope so. I think it's a great opportunity. And someone, student who's looking for something to do, matching computer science and health care is a great career and I strongly recommend it.

Certainly, can change the world. That's for sure. Where can people go to find more about your research or your company?

Well, of course, the obvious answer to all of this is Google, or Bing, or whatever your favorite search engine is, I should be fair. If you search Health at Scale, you'll find lots of things. If you search my name, if you go to the CSAIL website. I wish I had the URLs in my head, but I don't.

We could always direct people to cap.csail.mit.edu and they could do a search on your name and bring up the projects that you're involved with and get a link to your personal web page.

And if they get to my personal web page, they'll find a link to the company. The company's web page is uninformative about our technology, as you would guess. And that it is mostly proprietary. But it has information about what the technology does. And that's interesting.

Great. Well, John, thank you very much for your time today. It was a pleasure talking to you.

My pleasure. Thanks.

If you're interested in learning more about the CSAIL Alliance program and the latest research at CSAIL, please visit our website at cap.csail.mit.edu. And listen to our podcast series on Spotify, Apple Music, or wherever you listen to your podcasts. Tune in next month for a brand new edition of the CSAIL Alliance's podcast and stay ahead of the curve.