AI-GR Pod 29 04.10.25 Mahmood

[00:00:00] I think that making that entire code base publicly available, and it's been a common theme in all of our work. We always make the code, everything as much as we can, and as much as is allowed, publicly available has been a big accelerator for us. And it has led to all kinds of improvements. Things that I had never imagined because once we made CLAM publicly available, you know, everyone from the entire community started contributing, and giving us feedback as to what we can improve.

And of course, we do a lot of improvement internally as students learn what is the best way to do what over time. But we also had this community feedback that led to a lot of systematization, computational improvement, just ideas for how we can improve the entire pipeline and, and, and scale it.

It was sort of an internal battle to try to make some of the larger foundation models that we had built publicly available. We had the innovations office and corresponding [00:01:00] legal people involved in trying to understand what is it that we are trying to make public. And once they understood that the data is not being made public, no identifiable labels were used in the training.

Welcome to another episode of *NEJM AI Grand Rounds*. I'm Raj Manrai, and today we are excited to bring you our conversation with Dr. Faisal Mahmood. Faisal is an Associate Professor of Pathology at Brigham and Women's Hospital and Harvard Medical School, and a leading computational researcher working on applying AI to medicine and in particular pathology.

We had a chance in the episode to dig into Faisal's work on creating new pathology foundation models, his experience starting a company while based in academia, and his approach to selecting interesting scientific problems. Faisal's lab has been really, really productive and he shared some of their secrets for success.

All in all, this is a really fun and insightful conversation. [00:02:00]

The *NEJM AI Grand Rounds* podcast is brought to you by Microsoft, Vis.ai, Lyric, and Elevance Health. We thank them for their support.

And with that, we bring you our conversation with Dr. Faisal Mahmood. Faisal, welcome to *AI Grand Rounds*. We're excited to have you here today. Oh, thank you so much for having me. Faisal, great to see you and great to have you on

the podcast. So, this is a question that we always get started with. Could you please tell us about the training procedure for your own neural network?

How did you get interested in artificial intelligence? And what data and experiences led you to where you are today? Yeah, so, uh, my background is in electrical engineering and computer science, and some of the earlier work I did was around imagery construction and electron microscopy and how to use cryo-electron microscopy images and reconstruct them back in three dimensions.

And I did a [00:03:00] lot of theoretical work on optimization to try to make it happen. Deep learning before it was called deep learning. And then I did a postdoc at Johns Hopkins. And it was very different, more related to medical imaging than biomedical imaging. And I predominantly focused on endoscopy and synthetic data generation and how can you use synthetic data generation to solve and answer some questions in endoscopy.

And then from there, I became interested in pathology and moved here at the Brigham, where I've been for about six years now. And the reason I really got interested in pathology is because it was often regarded as one of the last modalities in medicine to be digitized since there's a lot of untapped potential there.

When I was at Hopkins, they were just getting started with digitizing slides and did some early work there and that's what I've focused on since moving here. You know, something that struck me is that I think a lot of our guests have, often sort of early stories or early exposures to artificial intelligence that, [00:04:00] whether through science fiction, through reading or through movies, that they then now get to revisit and engage in some of these kind of wild ideas from childhood and early on with their professional work.

And I'm just wondering if you could take us back like even further, right? Even before that, a little bit. Do you have any memories of early exposure to AI or, you know, interests that are related to AI even before you started your research career? Yeah. Well, nothing comes to mind off the, off the bat, but, but yeah, I'll say this, that I was always interested in doing a lot of data driven modeling.

It just made sense to me that using lots of data to answer questions that are important would be the way you make machines intelligent. For a while in my field, in medical imaging, there was a lot of work being done on using handcrafted features and using [00:05:00] all kinds of other modeling techniques, but it was not scalable.

And just using lots of data, abstract feature representation, that all sort of made sense to me. Got it. Great. So, we want to actually dig into some of your work in AI for pathology. And I honestly don't know where to start. This is a huge compliment to you. I think you've been just amazingly productive.

When I think of someone who's really been at the forefront of AI for pathology, you know, you're the first person that comes to mind. And you've just done this amazing, amazing amount of work for the past few years. And again, this is a hard choice about like, which paper of yours to start with, but maybe I can start with one that, out of personal interest, caught my eye a few years ago and it's gotten a lot of attention that I think has also inspired a lot of folks.

And this is your paper that, and this is another question that we have to ask you at some point, which is how you come up with these [00:06:00] names, but this is your paper on CLAM. The title of the paper is "Data efficient and weakly supervised computational pathology on whole-slide images."

So, I think that title's actually pretty informative. This was published in *Nature Biomedical Engineering* a few years ago, back in 2021, I believe. Maybe we could start with that paper. Could you tell us about like, first of all, tell us what that title means, what you were trying to do there, and maybe also take us back and give us a little bit of the backstory on the paper itself.

Yeah, so, when I got started in this field, making some of these methods data efficient was a big problem. So, lemme just give a little bit of a background here. So, glass slides, pathology glass slides. When they're digitized, they turn into gigapixel images. So, these images are very, very large and a single patient case can have anywhere from a single slide to up to, you know, a couple of hundred slides depending on what kind of case it is.

And the only label that you have corresponding this entire case, a multitude of very, very large gigapixel images, they're almost like satellite images, [00:07:00] is the pathology report. So, the relationship between your data and your label is incredibly weak, and when you have lots of images and single labels what comes to mind from a machine learning point of view is that yes, well, we can use multiple instance learning. But you know, multiple instance learning can be inherently very data inefficient in this case, because a very small region from the gigapixel image can update the weights and biases of the model.

And there were some work done before our 2021 *Nature BME* article showing that you needed about 10,000 whole-slide images corresponding a single class to train a model that would be equivalent to pixel level annotations. So, data

efficiency in this area was a huge problem. And what my group inherently does is that we develop solutions and methods, computational methods that would solve very targeted clinical problems.

So, this is a study that we did more on the methods side, or we try to solve this problem. It's just being [00:08:00] really focused on how can you make this form of supervised learning more data efficient? And we used a number of different bells and whistles at the time that were used by conventional machine learning, like attention, and models that were trained on ImageNet, and just extracting features from those models to make the setup inherently more, more data efficient.

But I think the really catalyzing factor and why this article is so popular is that we made all the code and models and everything corresponding to this publicly available. So, people have really gone to town using this tool for every major organ system, every disease model, all the way up to forensics, to wherever you have pathology images, it's, it's been applied.

And—. Were you surprised by any of the uses of the model? Yeah, we've been surprised a number of times. People have used it for plant pathology, things that I just would've never imagined. And at the time, the backstory is that the real reason we did this study, because I was [00:09:00] very interested in cancers of unknown primary.

So, when cancer metastasizes, it often becomes unclear where it may have originated. And identifying the primary origin of tumor is really important because most drugs are primary tumor specific. And these patients can't go in a clinical trial because most trials would require that the primary be already identified.

So, we, we wanted to solve that problem and it's like a highly imbalanced, multi-class multitask kind of a problem. But there was no good way to solve it using these very large gigapixel images where the number of samples per patient can vary and the number of cases we have per patient is heavily imbalanced.

So, that's why we developed CLAM. And then we applied that to this specific problem in cancer of unknown primary. And that's a, that's also a 2021 *Nature* article. And we call it TOAD. This is TOAD, right? TOAD, yeah. Yeah. Okay. So, how do, how do you, I gotta ask now, how do you come up with the names? So, it, it [00:10:00] became just a theme.

So, so, I came up with CLAM ‘cause I like— Okay. —eating them. Right. So, okay. So, so, so, so, it’s, it’s one of my favorite foods. So, I just came up with CLAM. And it also made sense because it’s, you know, clustering constrained attention, multiple instance learning. Nice. But that sort of inspired my students to start coming up with names.

In the beginning it was all aquatic themes. And one of my Ph.D. students who was from the big program, Richard, he eventually created an internal package that we were using for quite some time. It was called Fishing Rod. So, you use the Fishing Rod to fish all these aquatic species. Amazing. Amazing. I think you said something that I thought was really important.

So, I just wanna dig into a little bit, which is that you think part of the reason that CLAM took off, and I think this has actually been a theme in a lot of your work, right? But one of the reasons you attributed to CLAM taking off was releasing the [00:11:00] model. Making it open source. Making it available, making it easy for people to build off of.

And I think this has been a, you know, I just, I’m immediately drawn to the comparison with, what David Ouyang does for AI, for cardiology. He was a guest we had on a couple months ago. And, you know, a couple other folks who’ve really emphasized the importance of it. And I don’t think we can say this enough because I think it’s just, it really changes the trajectory of the paper afterwards, if people can really build off of it easily, and if they can, use it to do their own work.

And so, I’ll commend you on that, but I also think it’s worth just maybe talking about that a little bit more, right? Like, I think there’s a lot of things that are attached to it. You have, you’re naming it right? You’re sort of, this is the wrong word, but you’re, you’re making it an entity, right? Productizing, that’s the wrong word for it. But you’re making it sort of a thing. You’re releasing resources that accompany that paper, not just sort of ending at the science or the inference or the findings, which are fascinating.

They’re [00:12:00] interesting, but you’re also making it sort of a engineering project that people can then build off of. So, maybe you could talk about that, like what are the ingredients here? I think you’ve done this a lot. You know, what are, what’s really important to emphasize for someone who wants to emulate that?

And, and, and maybe if you could comment on the extent to which you get institutional pushback or support on this, because my experience has been that,

like, if I train a model on hospital data, the hospital gets very nervous about potentially releasing the model, not even the data. So, I don't, I, I, if you have any, uh, experience with that, I'd love to hear about that, too.

Oh yeah, absolutely. So, with CLAM, it was mostly the code base because we had some hospital test data, but the models were largely trained on the TCGA at the time. I think that making that entire code base publicly available, and it's been a common theme in all of our work.

We always make the code, everything as much as we can and as much as is allowed publicly available has been a big [00:13:00] accelerator for us. And it has led to all kinds of improvements. Things that I had never imagined because once we made CLAM publicly available everyone from the entire community started contributing and giving us feedback as to what we can improve.

And of course, we do a lot of improvement internally as students learn what is the best way to do what over time. But we also had this community feedback that led to a lot of systematization, computational improvement, just ideas for how we can improve the entire pipeline and, and, and scale it. And to your question, Andy, so, absolutely.

It was sort of an internal battle to try to make some of the larger foundation models that we had built publicly available. We didn't face any resistance. It's just a matter of understanding. So, we had the innovations office and corresponding legal people involved and trying to understand what are, what is it that we are trying to make public.

And once they understood that the data is not being made [00:14:00] public. It's a model that's trained in a self-supervised manner, no identifiable labels were used in the, in the training. And there were, by the time we were making the first self-supervised large scale model public, there were other examples from a lot of other institutions, including from Stanford, that had already been made public.

And then they did a lot of digging and trying to understand what kind of license should be used for this. But eventually we did get the permissions to make the make the models public. Awesome. I'd like to now hop to another one of your papers and continue the shameless flattery here that Raj started and say that like, there's like two constancies in my Twitter feed.

One is that Elon Musk has done or posted something crazy, and two is that Faisal has posted another *Nature* paper on AI for pathology. So, so, thanks for

being that like steady rock in my Twitter timeline. Faisal, so this paper I think is interesting not just because the work was first rate, but also because of the [00:15:00] model it suggests for AI going forward in medicine.

And so, the paper is "A multimodal generative AI copilot for human pathology." And so, this was your PathChat paper. So, one, I'd love to like for you to tell the listeners what PathChat is, what it does, and whether or not co-pilot is a mental model that we should have for AI and medicine going forward.

Yeah, absolutely. So, the story of this study is that I often call it an accidental study. So, we started by, so, we're solving all these entrusting, supervised problems, and they're becoming more and more complex. We started by solving very small, supervised classification problems.

And then they became more complex for solving multi-class, multitask problems still, very supervised. And, but the backbone behind it is that because these images are large, we pre-extract features using a ResNet that's just trained on ImageNet. And then everything that was happening in conventional computer vision and [00:16:00] machine learning in general with self-supervised learning, it became clear.

That you can find just amazing application pathology where you have all these rare diseases, clinical trials, situations where you don't have enough data available for supervised learning. And we started to train large self-supervised models, maximizing for diversity, trying to collect every known human pathology indication that would exist in our archives.

Because it was established using a number of different studies, including DINOv2 from Meta showing that the diversity of data matters way more than the quantity of data. So, we max, try to maximize for diversity and we published two foundation models in *Nature Medicine* and once we had that, it obviously accelerated a lot of research we're doing and, and just building more supervised models. But another thing that we could do with those self-supervised models is because it can extract rich feature representations now from pathology images, is that we could, if we have enough instructions, we could train a multi-model, large language model, right?

So, [00:17:00] the goal was that can we have a single multimodal, large language model that can cater to all of human pathology? And there's a lot of philosophy behind it as to why we think that's possible. Like, companies like OpenAI are trying to build a single multimodal, large language model that will cater to human knowledge and pathology is much more specific.

And we, if we have enough data there's a high chance that you can converge to a model that would, that would do well around everything around pathologies. Just that data is locked up and it's not easily available. And the other challenge was that now we have to go in the opposite direction that pathology reports don't have enough morphologic details.

And a perfect human pathology chatbot or a co-pilot should be able to answer questions at any magnification level. It should be able to answer questions for specific regions in an image, for a single image, or a multitude of images, to be really, really useful. So, we had to collect a lot of data manually, but we also made use of data that was used for teaching [00:18:00] purposes at the Brigham and MGH.

So, we have a lot of colleagues and friends who contributed data to this, but it was also a very targeted data collection effort, including a lot of other institutions. And then we eventually had a very large instruction set. We trained the multimodal large language model, and we eventually had a chatbot that we could use for any pathology indication.

And then the assessment of the model was more difficult than building it because we already had the foundation model, and we spent a lot of time collecting the data. We built the model, and this entire process took about a year. But it took another year to just come up with that evaluation paradigm, that would be rigorous enough, took a lot of inspiration for how, for what was happening in the large language model research, both for medicine and in other critical areas.

And yeah, and then it was published last year around June. We have a lot of demos around it, publicly available. So, maybe I just wanna connect a couple dots there based on what you said. So, like, in large language model development, there's often this very expensive thing called pre-training where [00:19:00] you create the base model and that creates a model that has seen a lot of data and can do many things, but it doesn't do many things well.

And specifically, one of the things that it doesn't do well is interact with people. Yeah. And so, after this large foundation model has been created, there's this set of techniques called post training that's often applied to the base model. Some of it is supervised fine tuning or instruction. Fine tuning, like you mentioned, where you actually teach the model to follow instructions.

Sometimes you do reinforcement learning from human feedback. There's this whole kind of like dark art— Yeah. —to making these models useful. And it

sounds to me like what you were doing was sort of a one-for-one analog of going from base GPT to ChatGPT, where you took this very powerful foundation model, but made it usable for people.

Is that a fair characterization? Oh yeah, absolutely. So, I often say that we did two stages, but not three. So, a typical three stage process would be that you have the strong backbone model, you do instruction tuning on top of it, and then you do reinforcement-learning-based fine tuning, uh, where you [00:20:00] get a lot of human user feedback.

The model can generate many responses and the human would say, which one is better? And then you use that to, to further—. I assume that *Nature* paper will be coming out this June. I'm sure that, that, I'm sure that that's something that you're working on now. We're working on it. It's, it's tedious and difficult because—. One *Nature* paper a year or I'm not impressed anymore, Faisal, you set a really high, a high standard.

Yeah. And so, just to, okay, so, uh, I believe that recently PathChat has become the object of commercialization from your lab, that there's a spin out around that, and that you received an FDA breakthrough designation for it. Can you tell us more about that? Yeah, absolutely. So, two of my Ph.D. students, Max Liu from MITE, and Richard Chen from the Harvard Big Program, they had decided early on that they wanted to start a company once they graduate.

And PathChat seemed like a good opportunity because it became clear once we, once the, once we sort of announced it when it was a preprint, [00:21:00] that there was just a lot of interest. And I guess it was also around when generative AI, I mean it still does, had a lot of hype around it. And it is truly useful in the sense that it could be great for training, but also it has the capability to impact the entire horizontal of the pathology workflow.

Right? Because it, it's a chatbot, but you can also constrain the outputs to do other things. Like it can become a universal triage tool, or it can look at a case and suggest what ancillary tests and IHCs to order, and those can already be ordered before a pathologist is looking at it, or it can write the pathology report.

So, so, all of these auxiliary things it can do can have a lot of impact in pathology. So, Richard and Max started the company together with some of my other colleagues, and then we were thinking that what's the best regulatory pathway for this? And what my thinking and thinking of my colleagues in general is, how does digital pathology get of universal [00:22:00] adoption?

Because the digitization rate around the country is just around 4%. It's very limited, and there are a number of reasons for this. It's expensive to digitize these images. It's expensive to store them, but it obviously makes sense to everyone because the research benefits down the line are just enormous because we, we haven't discussed the discovery aspect around these, around these images, but we think that if there is sort of a killer app, a killer AI tool, that would make everyone's life easy, um, uh, people would digital very quickly.

So, we think that a universal triage device for pathology could drive a lot of digital pathology adoption. And the reason is that pathologists would often look at many cases and within each case there would be lots of, lots of slides and only a fraction of them might be positive. And if, if the device can tell you which ones are negative, it can make life easy for a lot of people. Um, and there's a good predicate to this. [00:23:00] A lot of people think that AI for pathology devices have recently begun to be approved by the FDA, but there was one that was approved in, uh, 1998.

It was called Auto Pap. So, it was just a, a camera attached to a motor, taking images from a slide and doing a line profile through the cells. It was for pap smear detection, and that was around when there were a lot of pap smears to be, to be analyzed, and there was a shortage of experts to do so.

So, the FDA approved this. And it has something called an NSR or no second review. So, this meant that all the negative slides could be screened or triaged out and the experts would only have to look at about 10% of the negatives. So, that's one of the ideas and PathChat can inherently be a very good triage tool.

And it's also an assistive tool. And the reason we wanted to get breakthrough device designation is because this is a relatively new [00:24:00] technology and breakthrough device designation gives you access to the FDA in the sense that they would meet with you regularly and design like a pathway towards approval.

And it's often the best course of action for new technologies and devices. And generative AI is relatively new. I mean, there are all these like open questions as to how do you get a device approved through the FDA that can make a diagnosis corresponding every indication. So, that was the reason to get the breakthrough device designation.

The FDA requires evidence around the efficacy of the device and that it's truly sort of a quote unquote breakthrough to give that kind of approval. So, that's,

that's what we essentially did. And it was granted earlier this year. Super interesting. First of all, shout out Richard Chen, who was an intern with me.

The first intern I think I ever supervised when I was a postdoc. So, it sounds like he's gone on to spectacular things. Yeah. Second of all, one of the things that I've learned from being in this space and certainly from this [00:25:00] podcast, is that technology development is hard.

Commercialization is often much harder still. And so, how do you think about business models around this? Is there gonna be a reimbursement model? Is this something like, I guess like how do you think about sort of market value or what even the market is in this case? Yeah, that's a great question and uh, I know Richard has been thinking about it a lot.

And I definitely agree with you. Richard is one of a kind. He was our superstar graduate student and did really, really well in the lab and is sort of addicted to making digital pathology and computational pathology just be absorbed in the clinic, and he really wants to change the clinical paradigm.

So, it's a really big passion of his. I think that there's some benefit here from the predicate we have in cytology that has been reimbursed for a very, very long time. So, reimbursement is, no doubt, a challenge for, it's a challenge in AI for health in general, and it sort of trickles down into pathology, but it might not be that big of a challenge for a triage [00:26:00] device.

The bigger challenge that we see is just digital pathology adoption, how quickly it would happen, and how much of a trigger would FDA approval for a primary triage would be. So, besides Auto Pap, all other more recent FDA approvals for AI for pathology tools have all been for a secondary analysis. So, pathologist first looks at it, and then the machine learning would essentially confirm.

So, if we do get approval for this down the line, it would be for primary triage at the slide level. And then we hope that that would, that would sort of drive more digital pathology adoption. Cool. Awesome. Thanks. Okay. I think we're ready to move to the lightning ground.

Are you ready? Yeah. Yeah. Ready. Andy, did you wanna give us the intro and the rules of the lightning round? Sure. In the lightning round we'll ask you a series of questions. It's [00:27:00] up to you to decide how serious or non-serious to take them. Short answers are rewarded handsomely, and long answers are punished viciously.

So, those are the rules of the lightning round. So, Faisal do you accept? Do you accept the task? Absolutely. So, this one is a little bit of a, I would say, pet peeve of mine. In that I have irrationally strong opinions on it, but in your opinion, do you think that AI should be explainable? To a degree. I love, uh, one of the FDA presentations showing something.

It said that AI gives hints, and I completely agree with that, but I don't think we should obsess over interpretability too much. The performance evaluation, the clinical trials, all of that is, is much more stronger evidence. There's, there's a lot of other things in health care and medicine that we can't really explain.

I agree with that 100%. Cool. This is just that, that answer was music to Andy's ears. I know. So, Faisal, [00:28:00] next lightning round question. What's your favorite novel? Oh, uh, *Norwegian Wood* by Haruki Murakami. Wow. Nice. Deep cut. If you weren't a professor at an academic medical center, what would you be doing with your life?

Oh, so, if I was not a professor at an academic medical center? Or more generally doing health care, like medical AI. So, if, if in the counterfactual world, Faisal was doing something completely different, what would counterfactual Faisal be doing? Oh, I'll be writing. Oh, okay. Novels, like the one that you just mentioned?

Yeah, yeah. Something like that. Have you, have you written before? Have you written fiction or nonfiction before? No, but that's what I wanted to do as a kid. Oh, wow. Yeah. Faisal, will AI and medicine be driven more by computer scientists or by clinicians? So, I think it'll, it'll be a [00:29:00] balance. So, if I just look at what I've learned over the past decade, some of the best projects in my lab have come about when computational trainees and clinical trainees partner together.

And I've been very lucky to have great graduate students from amazing Harvard and MIT programs, but also just great residents and fellows working in the group. And some were just amazing with Ph.D.s in computer science. And I've also gone through entire clinical trainings. When they work together, that's when the magic happens.

Yeah. Cool. Alright. Last lightning round question. It's one that we ask most of our guests and we always get interesting answers. If you could have dinner with one person, dead or alive, who would it be? Ooh. Wow. Uh, that's a, that's a tough one. Tesla. Oh, Nikola Tesla. Yeah. Nikola Tesla. I mean, [00:30:00] my background is in electrical engineering.

Just really inspired by all the out of the box thinking, but still everything that made sense. I wish it was possible. Yeah. And unfortunately overshadowed by Edison in many regards, despite being his equal or greater, so. Yeah. Yeah. Cool. Awesome. Well, Faisal, you, you survived the lightning round.

Congratulations. Thank you. Alright, so, we have a couple big picture questions to throw at you before we wrap up. Again, I think we've touched on this a little bit throughout the episode, but do you have a, like a, a unique ability to select problems that have the dual properties of being interesting, but also not stuff that everyone else is working on, and therefore you're able to, uh, do meaningful work without momentum chasing.

And so, what is your method for selecting interesting problems in such a fast-moving field? So, I pick [00:31:00] problems for which there's no good solution right now, right? So, I think that of course we can use machine learning to replicate what humans do and maybe do a little bit better, reduce intra observer variability.

All of that is great, but I just feel that we as machine learning for health researchers can make a bigger impact if we target problems that for which there's sort of no good solution right now. Right? So, so, it was reflected in my postdoc, I worked on trying to see if we can use machine learning to predict depth from 2D conventional and endoscopes, right?

So, so, we generate a lot of synthetic data and try to solve that problem. If we're able to do that, you can essentially convert any conventional 2D endoscope into a 2.5D or a 3D endoscope. And the same is reflected in my more recent works over cancers of unknown primary. There, there are a number of different solutions, but there's no good [00:32:00] solution.

And these cases are just, it takes such a long time to diagnose these cases. And the other large study we did, it was back in 2022 for endo myocardial biopsy assessments. There's just huge intra observer variability. It has major issues with the downstream treatment, so, just problems for which there is no clear solution.

People are much more amenable to adapt those models into clinical practice versus trying to replicate or just build assistive tools. Awesome. Faisal, this is our last question for you, and maybe I can get you to give us your future vision for how medical AI is gonna evolve. And specifically, I think we'd be interested in your taking.

I think we've heard a lot about how you select interesting problems, how you've designed medical AI applications for pathology. But maybe you can reflect on and project for [00:33:00] specifically translation into the clinic, how you see medical AI shaping, reshaping, changing, not really changing medicine over the next five to 10 years.

And, I think it would also be very interesting for you to, to tell us about, you know, what you think are the key bottlenecks. Are they social? Are they technical? Are they regulatory? What are the key bottlenecks to moving AI from where it is right now to a greater role in clinical care?

Yeah, I think that there are two lines of development that would happen over the next five years or so. But it always happens sooner than I think so, so, it might even happen in the next few years. One of those directions is that I think that we'll soon have patient level models, right? So, we're seeing that all these self-supervised models in individual domains are becoming better and better at representing those particular modalities.

And we're seeing some [00:34:00] earlier flavor of multimodal, so, quote unquote foundation models, where contrasting with another modality improves representation for all modalities. We'll see that done in a much more holistic way, where the patient's entire medical record. It could be featurized into a single feature vector, and it would have a temporal component to it.

So, if a patient who's on immunotherapy has a toxicity event over the weekend, their feature vector has now changed. And, and then all the downstream variables would sort of reflect that. And that would do essentially two things, would really enable a lot of early detection, early diagnosis, and it can be happening on the fly all the time.

And I think that achieving this is very much possible today. It's only a matter of time before this happens. Once this happens, we will also have the ability to use entire clinical grade health care system scale data for discovery, right? Because if you look throughout the history of medicine, how do people come up with new diseases, new subtypes?

Is that someone goes in, looks at lots of data and says that, well, this [00:35:00] particular morphologic phenotype on an image or all of this categorization correlates with outcome, and this is subtype A and subtype B and disease A and disease B, and we have the opportunity to do that automatically on its own, almost continuously, as the ology of some of these diseases evolves over time.

So, it'll become like automatic discovery engines. That's one thing. And then the other thing I think that will happen, and it's also in the short term, is that with all the agents and agent workflows, computational biology as we know, it would just fundamentally change.

Because what computational biologists do today, where you have existing data or new data that you have collected, and you're often using existing tools and you're patching together some code that you write and using more tools, and sometimes you're developing newer tools, would just be done on its own. And a lot of hypothesis generation and testing and assessment can also happen on its own.

So, I think those are the two big things that are, that I think are just obviously coming. So, just [00:36:00] patient level foundation models that will enable both clinical change as well as lots of discovery and then agent workflows that would, that would just accelerate all kinds of computational biology research.

Alright, well I think that's a great note to end on. Thank you so much for being on *AI Grand Rounds*. Yeah. Thank you so much for having me. Thanks for coming, Faisal. Thanks. Thanks guys.