MIT CSAIL Alliances | Yoon_Kim_Project_4

Welcome to MIT's Computer Science and Artificial Intelligence Labs Alliances podcast. I'm Kara Miller.

[MUSIC PLAYING]

On today's show, a look at the promise of large language models to transform business and the economy.

Many professions will be-- certainly productivity will be enhanced by learning to use these technologies effectively as was the case when Search came about, for example.

Yoon Kim, an assistant professor in MIT's Department of Electrical Engineering and Computer Science, explores the huge potential of models that can predict what's to come.

We certainly have to think about what are the ethical implications of these technologies, but also what are the ethical implications of not making these technologies available if, for example, in the field of law, they're able to level the playing field.

But with great promise, Kim says there can be hype and we may have arrived at a moment when it makes sense to pay attention to that hype.

In the start up landscape, there does seem to be a lot of excitement around these models potentially without clear applications in mind.

So up next, how large language models could transform different sectors of the economy with a dose of skepticism thrown in.

In the early 2000s, a young engineer, just a couple of years out of school, was riding the bus to work at Google. A bus with Wi-Fi, so he was able to get work done. And the engineer, Kevin Gibbs, was working on a side project.

He wanted to be able to type a few letters of a website address and get suggestions for the full website address. It didn't take a whole lot of time for the folks at Google to realize, whoa, this could be used for more than just completing website addresses.

What if you could type in the name of a violin player or a soccer star that you have remembered, and then you were offered a suggestion for the rest of that person's name. Very quietly, Google's approach to an autocomplete system changed technology.

In some sense, large language models are nothing more than autocomplete systems. There are some things one does after training these to be autocomplete systems so that you can get interactive chat bot agents. But at its core, these models are just autocomplete systems.

But remember, says Yoon Kim, that was a long time ago before ChatGPT became all the rage. Long before we heard that large language models, which predicted what came next, well, they would change everything.

So one question is like, how is this different from Google Search Autocomplete that's been there since 2010? Well, it turns out they're essentially philosophically the same system that completes stuff for you. But this function that completes stuff for you it's a much, much larger model. It's a larger sort of a large transformer neural network architecture than what was happening before.

The AI that people are going gaga over today embodied by transformer neural networks, well, it's part of a long evolution. But since about 2020, a few key things have happened that have radically changed the game.

The big change that's happened over the past, let's say, three or four years has been this idea that traditionally when folks were working in natural language processing, we were generally focused on solving a particular task.

And this task is something like machine translation or question answering where often we had researchers saying, I want to build a machine translation system for translating from English to Korean, or I want to build a question answering system that will do question answering over Wikipedia.

Then folks doing different research and implementing different systems at companies and universities, well, they started to merge as it became clear that training large language models could encode knowledge about language and the world.

And you can take this now pre-trained language model and essentially adapt it for any task involving language. So whereas, for example, in the past folks were building question answering specific systems, the pipeline now is you take this pre-trained language model and then you adapt it to become a question answering system.

But the revolution in the last few years has not just been about what has happened at places like MIT. There has also been a revolution in how business and the public think about this technology. And that became clear in November of 2022 when ChatGPT debuted.

What ChatGPT showed was that there is an incredible commercial demand for these types of good enough assistant type systems that can interact with you. And moreover, you can tweak these pre-trained large language models to become these good enough interactive systems. And I think this resulted in the shift of like what things could be done with language technologies.

So what will be done with these technologies? Well, Kim says it's still hard to know. He believes there may be clinical applications, for example, and there are whole professions that may be changed.

It'd be very interesting to see what the occupation of software engineering looks like five years from now or even like two years from now.

In some ways, Kim comes at this transition with a unique set of skills. He went into finance initially and says he didn't even do very well in the one computer science course that he took in college. But in grad school, he took a machine learning course and then later a natural language processing course and he was hooked.

It fascinated me. I could see myself being in this field for the rest of my life, which resulted in going back to grad school and then sort of pursuing a career in research. Could be in academia or an industry, and then I sort of chose academia.

So perhaps it's not surprising that when he tries to explain the brilliance, if you can call it that, of today's large language models, he reaches outside computer science to explain how we got to where we are. And he notes that in lots of domains, simple goals can spur incredible complexity.

A good example of this is sort of evolution what our genes are optimizing is just simply at the gene level, we want to have maximize the population of similar genes in the next generation. So from a mathematical standpoint, it's not that complex of an objective, right? But of course, incredible complexity arises, Homo sapiens arise, the incredible richness of our lives arise from this simple objective.

So this is one example of a setting in which simple objective results in incredible complexity. Now importing this analogy to the next word prediction objective it still is in some sense incredible that these models are able to essentially display signs of intelligence and world knowledge and linguistic knowledge through just predicting the next word.

But if you think about it, if you actually are able to perfectly predict a distribution over the next word or what I'm going to say next and if you think about what's required for that, that does require a lot of world knowledge, a lot of intelligence, a lot of knowledge about linguistics.

So I think in some sense, philosophically, a lot of people did realize that this next word prediction objective is in some sense, in a very loose sense, what's called AI complete, i.e. it does require intelligence.

But that this actually happens or this seems to happen in practice when you train very large language models on billions or trillions of tokens. That's what's been incredibly surprising to certainly me and I think a lot of people.

I know one of the things you've noted is that ChatGPT is very expensive to run. How do you think about that and how do you deal with it ultimately?

Great question. So certainly and a lot of other folks are working on making these models efficient to train and deploy with the idea that if we view these technologies as being a net good, which I think it is-- so certainly there are again like benefits and risks to these technologies. But I think overall, there will be benefits.

Then we want to democratize these technologies and make them accessible to as many people as possible. And that requires techniques and research into making these models more efficient. And that's one of the big areas of my research program that I focus on.

You talked about democratizing the technology and I wonder how much you worry about the cost of running these models. Judy Estrin, who was the CTO of Cisco who also worked with Vint Cerf at Stanford back in the 1970s, wrote not that long ago in *Time Magazine* and this caught my eye. This is a quote.

"A small number of tech titans are busy designing our collective future, presenting their societal vision and specific beliefs about our humanity as the only possible path." Like, the overall article was, I think, concerned that though it feels like a kind of ubiquitous technology, really the people in the driver's seat are relatively few and relatively-- they have a lot of money at their disposal.

Yeah, I'm certainly worried by that, not just with large language models, but it's an example of a technology where it results in potential even more concentration of power, which seems sort of counter to this general idea of democracy.

I do think-- and this is not to say we shouldn't think about this, but this is not different from, let's say, Search or any other technology, like let's say like smart phones, which, again, is concentrated across folks, which, again, isn't to say we shouldn't be thinking about this. But also like we shouldn't be treating these technologies as something especially special in my opinion.

And I think this also speaks to some folks talking about the potential risks of these technologies being on par with nuclear war or something and I just don't see it. I think we should remain humble to the potential risks of these technologies.

In particular, they can be a multiplier of risks. And certainly the existence of these technologies could reduce the barrier of entry for malicious actors to, for example, make bombs, to hack into systems.

One scenario that folks talk about is this emergence of superintelligence that continually improves itself and decides that humanity is like not helpful. These types of scenarios seem very, very far fetched to me and I think in some sense distract from the real risks that we should be worrying about as it pertains to these technologies.

Got it. Let's get back to a thread that you had started on before, which is like the question of applicability, how this is going to change the economic landscape, the business landscape. I wonder in your mind what the most exciting uses are right now in large language models. And then the other side of that question is like, what might be something that they would be good for, but like they're not ready for prime time on that yet?

Yeah, I think with all the applications, still there should be careful vetting, scoping, red teaming of whether applications of language models are appropriate. So I will qualify everything I'm going to say, saying maybe like still these areas are not ready to be-- at least not yet ready to be combined or ready to be fields on which these models are applied.

Certainly I think large language models as it pertains to giving advice about legal aspects, in particular, if you wanted legal advice about a situation, that automatically resulted in like needing to book an appointment with a lawyer, which is expensive and is only a channel that's available to a small segment of society.

And I think that is an area in which these models will play a huge role in terms of giving some legal advice. It's not going to be like perfect advice. They're still going to make errors. But this in some sense does level the playing field between those who can afford legal advice and those who cannot.

Some people have talked about this in the context of health care. I think we should be especially vigilant about having these models give health advice. Maybe we'll get there, but that seems especially premature to me.

I think these models reduce the barrier to entry for folks who may not have a traditional, let's say, programming background or experience into-- and these models will enable, for example, folks with backgrounds in fashion or art to easily create applications that they're dreaming of rather than having to pay app developers.

And more generally, as with other technologies, many professions will be-- certainly productivity will be enhanced by learning to use these technologies effectively as was the case when Search came about, for example.

Mm-hmm. I feel like you're saying in large part this unlocks certain pools of specialized knowledge, right? Like, a lawyer might charge you $600 an hour if you wanted to set up a trust or a will or whatever. And maybe you could gain access to some of that knowledge for considerably less because as you said, like, how many people can afford very specialized knowledge from a very specialized lawyer?

And similarly with coders, you want to create a website or you want to do certain kinds of things, like only a few people really know that stuff. And so you have to pay them a bunch of money to get access to it, right? And it sounds like you're saying this helps to kind of break down some of those walls.

Yeah, and when you consider these types of benefits, which I think do exist, then like we certainly have to think about what are the ethical implications of these technologies. But also what are the ethical implications of not making these technologies available if, for example, in the field of law, they're able to level the playing field.

You talked before about how software engineers, things may be completely different in two years and five years. Tell me what may shift.

Yeah, so I think for many languages or many applications, I can imagine a world where now instead of having our software engineers create programs or code from scratch, initially it would be the template or the draft will be given by, let's say, a large language model and then the task of the software engineer is to, one, verify the correctness of the program generated by these models, and then, two, improve upon it.

And in some sense, this is the case to an extent for many aspects of software engineering, right? So when you're creating something in a language you don't know, you might get similar code snippets from Stack Overflow and then modify it and then play around with it to make it do what you want.

And then you want to, of course, verify that it is doing what you want. But that sort of initial draft process I can see being done largely by these types of models in the near future.

So humans are turned more into editors than into the initial generators of the stuff.

That's right.

When you talk to folks in business, I wonder if you feel like this stuff is already being applied and if you feel like there are misperceptions about how it could or should be applied.

Yeah, so certainly many folks are looking to apply these types of technologies to either improve the productivity of employees or enable new products, for example, like customer service chat bots that actually work.

Right.

And I get the sense there are a lot of startups that are deploying this types of technologies essentially using like API access to these models and providing these as service to other companies.

And I think the larger corporations are a little bit more conservative about immediately adopting these technologies, unless there's a clear sort of almost existential risk in not adopting these technologies, like what happened with Google, for example, and Bard. Whether their applications are missing something, that I think the applications are too myriad for me to comment on.

I will say, though, in the startup landscape, there does seem to be a lot of excitement around these models potentially without clear applications in mind. And maybe this is just how the startup space is.

You have an exciting new technology, you have a bunch of startups that say we're doing this technology. And that results in tons of funding and excitement. And hopefully some of these startups will become sort of society-changing companies. But a lot of them-- yeah, certainly there are startups that I don't know what they do and why they wonder.

It sounds like the word that comes to mind here is hype.

Hype. Yup, yup. But also like, yeah, don't want to discourage that. And that I think there should be-- again, not an economist, they should be funding given to startups that are exploring the space, and then just don't necessarily have a clear core application in mind. So I do want to qualify my potentially pessimistic statement before.

[CHUCKLES] So do you feel like there are companies that are being more conservative and being like, well, we're not going to start integrating this stuff until we can figure out what the heck we're doing-- like, why we would do that? Do you feel like the conservatism makes sense?

I do think so, especially in high stakes applications. And there needs to be more research, but also sort of a clearer like application-specific investigations of the capabilities and limitations of these models. And such risks are obviously greater for established companies.

When you mention the potential, the applications for these models, obviously almost everything we've talked about and a lot of the excitement over large language models has centered around text or on chat bots. But broaden our horizons a little bit. What else can these models consume and what else can they produce?

Yeah, a great question. So there's a lot of recent-- well, exciting and work in research and applications of this around having language models use tools to interact with the external world. So if you've seen ChatGPT plugins, this is one instance of this.

And fundamentally, as you noted, language models consume text and output text. So the question is, how do we get them to do stuff in the world for us, which, again, potentially increases the benefits and risks?

But let's say we have sort of beneficial applications in mind and having these models actually interact with the world and use through tools. And tools can be things like cooling calculators, cooling translation systems, or controlling robots. I think that's an incredibly exciting area of research that has potential widespread industrial implications.

Any other things that you think just don't get enough attention because of all this focus on text in, text out?

I won't say this area doesn't get enough attention, but there is a lot of vibrant work around now training multi-modal language models so they can naturally incorporate images. GPT-4 is supposed to have this capability. I don't know when it's going to be released.

But for example, when they introduced the GPT-4 manuscript, they had some very impressive applications of its being able to interpret what's going on in images. And I can see this being generalized to not only images, but embodied environments.

So you have, let's say, a robot that's sort of like taking in information about the environment and combining it with, let's say, instructions that are given by a human. And then instead of predicting text, it's predicting what action to take next. There's a lot of quite exciting research happening in this area, but actually training, pre-training language models that are embodied in this way.

So it's not just, let's say, next word prediction, but it's you take in your current surroundings and potentially language and you predict not only the next word, but what the next environment is. I think could result in more useful systems that are embodied and have more grounded knowledge of the world.

I feel like that has all sorts of applications from medical uses and like safety uses to-- I mean, like even the Roomba has been trying-- the vacuum cleaner, of course, that works on its own that moves around your house has been trying to take in data about its surrounding. That's been one of its like the big projects at iRobot is to figure out how do you know about your surroundings, right?

Yeah.

Finally, when you look ahead, what's the development that you're excited about but maybe people aren't thinking about enough who sort of don't inhabit the world you live in? But something that you're really excited about, thinking about that may lie down the road.

Yeah, so maybe I'll answer this from a research angle because these are areas that I'm focusing on. So one research question is, are transformers enough? Transformers certainly seem empirically effective, but expensive to train, expensive to deploy. Are there other types of neural architectures that are potentially more efficient, but just as performance as transformers?

And there's a lot of rich research from various groups going around in this area. And I think I'm sort of interested in revisiting this question of other neural network architectures that are potentially more efficient and as performance.

Another question that I'm interested in is sort of going back to the fundamental question of whether language modeling is enough. So this idea of having models extract intelligence from next word prediction, it certainly seems to an extent work empirically. But is that all we need?

And I think there are arguments made for both sides. Maybe philosophically, future prediction is all you need. So for example, if you know all the particles and the positions and velocities, which you can't, but let's say you do and you're able to perfectly predict the distribution over next set of particles, then you've solved physics in a sense, right?

So from a philosophical standpoint, being able to predict the future does entail knowledge of physics, in this case, and intelligence, and let's say linguistic intelligence in the context of language models. So is that all we need? Maybe, maybe not. And that's an interesting question that I'm interested in exploring more formally.

Interesting. So when you say is it all we need, it sounds like you're not just interested in the forward arrow, but maybe like different directions that the arrow points in or something.

Yeah to an extent. Yeah.

OK, OK. Yoon Kim is an assistant professor at CSAIL. Yoon, this was a great conversation. Thank you so much.

Thank you so much for having me on. Have a nice day.

[MUSIC PLAYING]

If you're interested in learning more about the CSAIL Alliances program and the latest research at CSAIL, please visit our website at cap.csail.mit.edu and catch our podcast series on Spotify, Apple Music, or wherever you listen to your podcast.

I'm Kara Miller. Our show is produced by Matt Purdy and Nate Caldwell with help from Audrey Woods. Tune in next month for a brand new edition of the CSAIL Alliances podcast and stay ahead of the curve.