And Humanity Created Intelligence podcast: episode 4, transcript

How AI learned attention – Kyunghyun Cho

Eric Malmi:
Was it still a surprise – like how influential this concept ended up being?

Kyunghyun Cho:
Absolutely. I was hyperfocused on building this machine translation system, and then attention turned out to be the perfect way to solve this problem. It took me another, let's say, 2 to 3 years to realize that this is a pretty much universal mechanism that can help us build AI systems for the different problems.

The decision to come to Finland I still think was the best decision ever.

Eric:
Welcome to the new episode of Aalto University's AI podcast called And Humanity Created Intelligence. My name is Eric Malmi. I am an adjunct professor here at Aalto University and ELLIS Institute Finland. Today I have an honor to sit down with Professor Kyunghyun Cho from New York University. Kyunghyun completed his master's degree and PhD here at Aalto University. He has also worked as a research scientist at Facebook, co-founded the protein design company Prescient Design, which Genentech acquired in 2021. Kyunghyun's work has been cited well over 200 000 times, and he has co-authored the paper that introduced attention, the mechanism underneath every modern language model. It's an honor to have you. Welcome, Kyunghyun.

Kyunghyun:
Thanks for the invitation. Great to be back in Helsinki and in this studio, which did not exist when I was studying.

Eric:
True.

Today's episode is a bit of a special one. I would usually co-host with Professor Arno Solin and we would interview a Finnish AI pioneer. Well, now today Arno is traveling. And you're not strictly speaking Finnish, but you do have a strong connection to Finland. Actually, to the extent that more than once it has happened to me that when I introduced myself to some of my international colleagues, that I'm from Aalto, they would tell that, oh, that's the place where Kyunghyun did his PhD at.

Kyunghyun:
Very flattered.

Eric:
So you've been a great ambassador to Finland and to Aalto. I wanted to start by just asking, how did you come here in the first place?

Kyunghyun:
Yeah, it was a highly coincidental incident that I ended up here. So I was back in Korea finishing my undergrad. I was very interested in a lot of different stuff when I was in undergrad years. So I worked at different places. I was trying to major in different things. Eventually I just majored in computer science, no minors or anything like that.

I was finishing up my degree, but I couldn't really decide what I wanted to do next. And there was one friend who also was in the same boat. So he started at a similar time. And he couldn't also decide what he was going to do afterward. So we were just talking to each other a lot. Then we were taking one course that we thought was going to be extremely interesting. That was the introduction to bioinformatics. But to be frank, that course was not interesting at all. It was extremely boring.

So my friend and I were always sitting at the very back of the gigantic lecture hall and then just passing time. But one day my friend came a bit late and then gave me this brochure. The brochure looked pretty ugly, I gotta say. It was a bright yellow brochure with a red colored font, a blue colored font, and a black colored font. I was like, what is this? And my friend asked me, have you heard of the program called Macadamia? I was like, no, actually Macadamia is just nuts. Right? But anyhow, I decided to look at the brochure.

The brochure was the program brochure about the international master's program in machine learning and data mining at Helsinki University of Technology. And of course, at the end of the brochure it talked about how the Aalto University was going to be formed by merging Helsinki University of Technology with two other universities. And that really piqued my interest. I was like, oh, there are universities in Europe. Of course there are, but never thought about it. And in particular, if you are in South Korea, South Korea is an extremely U.S. centric country. The society's view about the foreign countries or the international, let's say, situation, is all centered on the United States. Almost no one, in fact, goes to study abroad in Europe or anywhere else other than the U.S. or Canada.

So I never thought about it. But because of that brochure, I started to think about it. And then the funny thing, if I recall correctly, that was the only brochure left in front of the office that my friend picked up. And as soon as I saw it, I was like, this is it. I got to actually apply to this program. And then of course, as a contingency plan, I decided to apply to a few other programs. Master's programs, like one in KTH, one in Delft and whatnot. And that's how I actually ended up here.

Eric:
In your blog, you've written that "this was one of the best decisions, if not the best one, I've ever made in my whole life". So could you elaborate a bit? What did you like here?

Kyunghyun:
Yeah, it really was. I think it still continues to be, other than marrying my wife, of course. You know, that was kind of an equally best decision I made. So what happened is that when I decided to go to Finland, in fact, quite a lot of people told me not to or discouraged me from doing so. They just didn't know what was going to happen. The

level of uncertainty was really high. But because of that heightened level of uncertainty, I decided to go. It just intrigued me too much.

So when I came here, I really didn't have any background in machine learning or anything that is related to artificial intelligence. I just got here with these skills for software engineering and then with some knowledge in the networks, because I was actually working at a networking device company back in Korea.

But then what happened is that, you know, Finland has this kind of thing that they try to provide the opportunities to as many people as possible, in a very more or less equal manner. Now, it is impossible to control the outcome. But trying to give the opportunities to people is important. But then there is a question of how do you actually ensure that the opportunities are going to be spread out as evenly as possible? And then one of the things that I noticed in Finland is there are actually a lot of things that are more or less randomly done. So the selections are not necessarily based on what you choose or volunteer. But sometimes things are just given to you as some kind of opportunities.

And one of those opportunities that I was given was what they call as an Honors Program within Macadamia. I think there were more than half, if not all of the students. Because it was a very small cohort back then. We were assigned unilaterally. So we didn't apply for anything. We didn't apply to the lab or anything like that. But all my friends in the same course were somehow assigned to the different labs so that we could spend one day a week in order to receive some stipend and do some research and then get the experience there.

I was assigned to what was called back then base group. The name changed later on, perhaps unsurprisingly. But base group. And then I was assigned to work with Tapani Raiko and Alexander Ilin who were postdocs back then. Now they are startup founders as well as successful scientists at Big Tech. And then they like literally randomly gave me a paper on third-order Boltzmann machines. They asked me to read it and then trying to implement it. That's how I got into this whole field of artificial intelligence.

So in a sense that wasn't my decision. But the decision to come to Finland and then Finland being a society where a lot of people are given opportunities as evenly as possible, I mean, that just happened. So I still think that it was the best decision ever.

Eric:
It's funny to think that random assignment to this group kind of ended up being so consequential. I mean, not only for you, but I would say kind of to the field in general, given what you've done, which we'll cover later.

You partially answered, but since the focus of this podcast is on Finnish AI pioneers, is there anyone in particular or people you'd like to give a shoutout to, that have influenced your thinking?

Kyunghyun:

Yeah, absolutely. So in that brochure that I was reading, there was one mention actually, as a prominent example of the Finnish excellence in artificial intelligence or machine learning. It was the Independent Component Analysis. So there was a textbook by Erkki Oja, Juha Karhunen and Aapo Hyvärinen. Of course, I didn't read that book back then. It was just an example. Then I came here and I got to, of course, know Juha Karhunen who was my PhD advisor during my time here. And then I got to talk to Erkki, I got to talk to Aapo.

What I noticed was their persistence in thinking about the problem already from the late 70s and early 80s in the case of Erkki. Then Juha followed on very rapidly afterwards, mid 80s and late 80s. And when I started to follow up on those papers from those eras, the interesting thing is that there's almost no equivalence outside Finland or in the whole world. There are very small pockets of researchers who are working on this direction here and there, from the neurophysiology side, or a bit of a learning side, but not a lot of them. But they just continued on.

They continued on and on until Aapo joined the group in, let's say, mid to late 90s, and then started working on the independent component analysis. And then Aapo brought in this kind of idea of the statistics. Can we be statistically rigorous about these various types of the component analysis. And then they continued on. They went on until 2005, 2006. And then I came here in 2009. Now the whole field had shrunk a bit back then. You know, they were working on these kind of unsupervised neural networks, but they were still working on it. Juha was still working on this kind of ICA and then the nonlinear PCA.

One of the major papers that I read, and we were all very happy about, was the Alexander and Tapani's paper that was published at JMLR back then on how to deal with the missing values when you perform the PCA, for the purpose of the recommendation systems. This kind of persistence of this lineage of the researchers going all the way from Erkki, all the way down, to me, I guess, that is amazing. Thereby all those people in this line I'm a very big fan of.

I need to actually mention Harri Valpola as well along this line, because I got to read his master's thesis on sparse encoding or sparse coding from 1995 or 1996. I mean, I read it very late, only about 2016 or 2017 when I was already in New York. It's brilliant. It's kind of, say, 20 years ahead of time. But yeah, there is this lineage of the Finnish researchers who have done amazing work. The amazing thing is their persistence. Despite the fact that everyone was doing something else, they just continued on.

Eric:
Yeah, we heard some interesting stories. We interviewed both Harri and Erkki previously. Kind of how small the field is and it was hard to get funding.

Kyunghyun:
That's true.

Eric:
You have to be very persistent. Keep believing. But it paid out in the end.

So then after Aalto, you moved to Montreal to work with Yoshua Bengio's team. And I think some of your most influential work you did during your period there in Montreal. One of these works already mentioned was the paper that introduced the concept of attention. Now one of the goals of this podcast is to make the fundamental technology behind today's AI systems understandable to a broader audience. So, since you're an educator as well as a scientist, could you explain what is attention? What does it do?

Kyunghyun:
Yeah. That's a big question. But you're right. I should be able to explain it pretty well. I can actually take one step back and try to talk about why it's difficult to build machine translation or these large scale language models in general.

In order for you to be able to answer any questions or try to follow up on the long conversation that you have been part of, just like we are having now, the important thing is for you to be able to remember all the salient and important information from the previous, let's say, utterances or the previous context that we have about the conversation or the book that you're translating or whatnot.

But then here's the question. Before you say something or before you hear the question that you need to answer, how do you know what you need to remember? One way is that you say that I'm going to remember every single detail and then try to go back and replay the whole thing over and over until the point at which you find the right information, bring it back, and try to answer this new question.

Or the second way is that you try to predict what the question is going to be, or what the question is going to be like. Thereby, you compress all the information, but leaving only the ones that are potentially relevant for the coming, let's say, question. Now, what most of the people were doing back then was to do the second way. As we continue on this kind of conversation, or as we read the books, we're going to figure out what are the things that are going to be more relevant for the future questions or the future sentences that we need to translate. And then we're going to compress the whole context, regardless of how long or how short they are into a single representation or the single, let's say, storage from which we can actually get all those relevant information as soon as we get the new question or the new sentencing.

And in some sense, that's what we actually do. We have just one brain that has a fixed number of neurons at any point. And we're trying to actually save things that we think are going to be relevant in the future. Unfortunately, we haven't actually figured out how we learn ourselves in our brain. And what that means is that many of the existing learning algorithms that we rely on are not as good as human learning, in the sense that we almost always fail to detect what are those relevant bits, or what are those things that are going to be potentially relevant in the future.

So then we had to essentially go back to the first option that is that we're going to store everything and hope that we're going to be able to figure out what are the relevant things in the future by scanning them again. The idea I had back then was a dumb one. It's that

we're literally going to store the original, let's say, context. If it's a conversation, everything that we have talked about, we're going to store them in the original form. And every time there's a new sentence coming in, we're going to just reread the whole thing all together and then figure out, given this sentence, what are the things that I need to know in order to answer or to translate this sentence.

Of course, it's a horrible idea. Computationally doesn't make any sense. As a computer scientist, I should feel really ashamed that I was thinking in that way. But fortunately, I did not actually bring it up because it was one of those very hot days in Montreal in 2013 summer or 2014 summer when I came to the office thinking, this is the right way to solve this problem. Yes, we're going to store everything. We're going to reread it over and over. Every time we receive a new sentence or new question.

I was about to say that, but then Dzmitry Bahdanau, who in fact was a master's student intern from Germany in Montreal, he said that he has an idea. I was like, okay, let's hear your first. You know, I thought that it wasn't going to be good. So I was going to give my brilliant idea. But then Dzmitry actually gave a similar idea, but in a much more elegant and computationally sensible manner. That is that we don't save the entire context as they are. We turn them into a much more readily accessible form that is a vectorized form. And then what we do is we use the dot product in order to figure out, given the new question, and its vectorized form, what the relevant things are going to be. And then this determination is going to be done automatically through the process of learning.

As soon as I heard the description, I could tell that this is like the idea that's going to work. And usually, you know, you do research yourself, it's very rare to have an idea that you can tell that it is going to work. Almost whenever we think that the idea is going to work, it doesn't. But this one I could see that it was going to work. So I stopped actually talking about my idea at all. I just listened to it. Let's go with this one.

So coming back to your question, the idea of the attention is that instead of trying to guess what is going to be relevant for the future in advance, rather let's try to keep as much information as we can in a growing manner. So if we receive more information, we're going to grow our memory. And then this memory is now going to be populated in the original form, but in a form that allows us to rapidly and very effectively find the relevant information in the future. And then attention is a way for us to figure out which part of the information from the past our machine learning model or the AI system needs to attend to based on the question that needs to be answered now.

So that's the idea behind the attention. It turned out that there are a lot of nice mathematical properties that allow us to scale up these AI systems by using attention as well. That's the reason why we're using it everywhere these days.

Eric:
You mentioned you kind of realized immediately that this is a great idea. Was it still a surprise to some extent, like how actually influential this concept ended up being?

Kyunghyun:

Yeah, absolutely. In a sense that what we were thinking back then was to view this as a way to solve the problem of machine translation. Now, of course, translation is a kind of universal problem because you translate something to another thing. Of course, it can be a question to answer, Korean to Finnish, English to Spanish, audio to, let's say, transcript. That is the case. Except that we weren't even thinking that big back then.

We were thinking very narrowly because there was one problem of translating a sentence in one language to a sentence in another language that we just could not solve at all. So we were, or at least I was, very hyperfocused on what the issues in these AI systems are that prevent us from building these machine translation systems. And then attention turned out to be the perfect way to solve this problem.

And then the reason was, of course, there is an intuition that I just used to explain the whole concept of attention, but there were mathematical reasons, like it's a perfect way to address the issue of the vanishing gradient and whatnot. But somehow it worked really well for machine translation. It took me another, let's say, 2 to 3 years to realize that this is a pretty much universal mechanism that can help us build AI systems for different problems.

Eric:
And actually, many of the, I would say, architectural innovations in the space of language models, they've been developed initially for machine translation. So is there something special about that application, why it has been such a fertile ground for breeding these new ideas?

Kyunghyun:
My perspective is that computer science or information science or machine learning or artificial intelligence, however we call our field, it is all about problem solving. We don't have a kind of elusive intelligence or computation that we're trying to uncover the mystery of. I mean, sometimes we do almost by accident. But it's very different from, let's say, biology, physics, chemistry or whatnot, where they have a very specific object in the nature that they are trying to figure something out of. But in our case, that object, in my view, is the problems.

How do we actually come up with the solutions to all those different problems while being scalable? That is, that we don't want to come up with a new solution for a new problem every single time. We want to come up with a common set of potential solutions, or the components behind these solutions so that we can mix and match in order to solve the future problems as efficiently as possible.

So if you think about it from this angle, what we actually need is the motivations each time, motivating problems that we want to solve. And thereby when those problems are solved, almost as a side effect, we get a bunch of these principles that we can use to solve the future problems.

And I believe that what is really important here is that how to determine what are the problems that are motivating, that inspire people to work on solving those problems.

Machine translation or just translation in general, has this amazing quality that it allows people to communicate better with each other, and it allows people to learn about what others have already learned and then be able to communicate what we have learned first to the others who haven't learned it. And it goes all the way to this kind of biblical time. They talk about the Tower of Babel and whatnot, right? So in a sense, it feels like the translation is something that really motivates and inspires a lot of people, almost inherently by our kind of desire to communicate with others.

And then there is one interesting motivation I had in my mind, the fact that there is a gigantic digital divide on the internet. In fact, I learned that when I moved to Finland. It's a great country. I immediately fell in love with the country. I immediately fell in love with how people actually behave in the society, how the society is structured and whatnot. But also, I was very surprised to learn that Finland is a tiny country. In terms of the landmass it's a pretty large country. However, when you look at the population, it's a very small country. Back when I was here I think it was about 5,4 million people, probably slightly higher now I guess. More or less. So it's a very small country.

But what you notice is that on the internet there are so much materials that are written in Finnish. Finnish is a language that is spoken by less than 10 million people in the world. Like entire world. Right? And still, the amount of the content you see is much greater than the amount of content you see on the internet that is written in, say, some of the languages that are spoken in Indonesia.

Eric:
Finnish Wikipedia, for instance.

Kyunghyun:
Exactly. It's massive. So there's this discrepancy in the amount of information that is available on the internet versus the number of speakers of those languages. And then translation, machine translation, automated translation is one way for us to directly address this issue of the digital divide. So I thought it's an interesting problem and it could have a very positive impact as well. So I think there are a lot of people actually bought into this mission very naturally. It could have been another problem, I think. But machine translation back then was the problem that we can solve. And then thereby we actually ended up with a lot of artifacts that turned out to be very useful down the road.

Eric:
Yeah. And other methodological breakthrough that happened while you worked on machine translation is the gated recurrent unit used in recurrent neural networks. Nowadays transformer-based models have largely overtaken these RNNs. But still, I feel every few years there is some new paper that makes a lot of noise. For example, there was the Mamba architecture a few years ago that had some recurrence in it. So what do you feel as you've been in this? Like, do you think they're going to make a comeback and take over transformers?

Kyunghyun:

Yeah, I almost feel like we are actually coming back to recurrent now. Mamba is one of the examples. A couple of days ago, I was just taking a quick look at the Kimi K3 model, let's say, a technical report. And one of the ways by which these large scale models are handling a very long context is to ensure that we don't have to do a strict quadratic attention, that is, that we look at every single representation of every single token from the past, one at a time. But rather than that, we're going to actually compress some of the representation or some of the chunks of the tokens into smaller bits or a smaller number of representations. So that we ensure that the number of the things that we need to attend over from the past is somewhat manageable.

And in doing so, what they use is a recurrent neural network. Now, of course, recurrent neural net comes in many different flavors. Like there could be infinitely many variations, but the one that is being used these days is so-called linear recurrent neural network, which actually connects very well to a lot of the innovations that happened in Finland a long time ago, starting from the Oja's rule, so Erkki Oja's rule, and then on and on, like the ICA and whatnot.

Doing this actually gives me a thought that in fact, the right thing to do was not to replace recurrent neural nets completely with the attention, or vice versa, or trying to replace the attention completely with the recurrent neural net. It's actually a bit of a combination. And this makes sense. We don't know what needs to be stored perfectly to solve the problems in the future. However, often we know that there are local regularities that allow us to compress the local or the small chunk dramatically. So then we can use a recurrent net in order to compress those chunks. But then we use the attention to ensure that the global regularities that cannot be compressed down are preserved.

So I think it's going to be a little bit of a mix. But then eventually, ultimately, of course, everything has to be fully recurrent neural network in a very high level view. Because that's how brains are and how the world is. Because we want to build on agents that are going to be here forever and ever. There's no way you can save everything. That just hits the practical limitation that there is no such thing as infinity in real life.

Eric:
But when you say that's how brain is, should we try to follow the brain design as much as possible?

Kyunghyun:
Not the design of the brain, but the problem that our brain is solving is somewhat different from many of the problems that we're solving with the AI, in the sense that it's about the continual never ending kind of episode that we are in. And we don't have the multiple episodes, we just go one direction only. So if we really want our AI systems to be like us in a sense that they are going to be here over and over, over time, and it's going to learn new things without having to reset, then it'll inevitably have to be fully recurrent neural network.

Eric:

So this problem of continual learning that a lot of people are talking about, you're saying recurrence is going to be important for.

Kyunghyun:
Yes. I think the major difference between AI systems and biological systems at the moment is this reset ability. Can you reset the system and then continue on? Biological systems, there's no such thing as a reset for us. AI systems, we can always reset them. But then being able to reset also implies the redesigned things to be reset. It's almost like there's a kind of weird chicken or egg problem or this cycle that happens here.

But if we start thinking about the problems we want to solve, that cannot be reset, then we'll have to come up with algorithms that cannot be reset. Thereby, it just learns or adapts continuously, forever and ever. For that one, eventually it'll have to be somewhat recurrent because you won't be able to just save everything.

Eric:
Okay. So, after Montreal, if we look forward in your career, you moved to NYU, you started there as a professor. And at some point you also joined Facebook as a research scientist part-time. I remember kind of looking up to you back then a lot, and I still do. I had started working on NLP and you had clearly established yourself as one of the leading NLP researchers in the world. Then I remember that somewhere in late 2020, maybe early 2021, you first announced you had left Facebook and then you had founded or co-founded this company called Prescient Design that specialized in protein design. And I remember being quite surprised. Maybe you had history with working on this kind of biological applications. But being in the position you were, wasn't it kind of a bit of a reset starting all over in a new domain?

Kyunghyun:
Yeah, you're absolutely correct. And by the way, a big fan of you as well. I've been following your work. So yes, thank you very much.

Yeah, an interesting thing happened in 2017 or so. It's that, you know, once we started to see that these attention based machine translation models work, there were so many problems I wanted to solve. So instead of building a machine translation system that translates one language to another, I wanted a machine translation system to translate any sentence in any of the many languages to any of the many languages, so many-to-many translation.

I wanted to build a machine translation system that knows how to do simultaneous translation. So translating the things in stream. And then we would use reinforcement learning to make a decision when to actually start emitting the translated, let's say, tokens and whatnot. And I wanted to mix it with the multimodality. So I wanted to add in images, audio and video to facilitate translation.

So I had all these ideas. I admitted a lot of PhD students. I received a lot of visiting students. It was really fun time. But then 2017 or so, one thing I realized is that there is a recipe to solve any of these problems. The recipe is that you first collect as much data

as you can. You put them into the format that is easily accessible by these sequence models. So put them into sequences. And then you start training a model.

Let's say the model is not working as well as I wanted or how I want it to work. Then what I would do is that I'm going to first try to see if I can collect more data. If I collected more data, it got better. Let's say it didn't get as well as I wanted to again. Then what I'm going to do is I'm going to make the model larger. It will work better. Then if it did not work as well still, then what I would do, is I would just wait longer. I'll just let it train longer. And then it just got better and better.

And then, in other words, what happened was that the things started to become predictable. That's why I started to look for different applications or the use cases that may share some of these regularities, or the structures of the problem of machine translation. So that I can easily transfer some of my existing knowledge and know-how, and then try to learn something new and try to solve new problems.

Of course, in the case of the natural language processing, I was still doing so by going beyond this kind of autoregressive left-to-right generation, by looking at, let's say, sequence generation using, I guess now we should call it diffusion models, but the denoising type of models. But still I wanted to see what are the next problems that I need to be inspired by and motivated by where we do not have this kind of predictability. Because once the technology becomes predictable or the progress becomes predictable, and then if the technology is in fact useful for some actual problems, and then especially if there are business opportunities that are tied to it, then of course, that's already what the companies should do, not the universities should do or the university researchers should do.

So that's the reason why I started to explore quite a lot. And then, I actually even dabbled a bit on the text processing for political science. You know, like, let's look at some chemistry. Let's look at some music technology, which you are very familiar with yourself. And then let's look at the particle physics. I looked at all those different things.

But eventually I was talking with one of my colleagues, one of the co-founders of Prescient Design, Richard Bonneau, who was a professor of biology and computer science at NYU. We realized that there is a stark similarity between the language of biology or the language of life and the language of humans. That's how we started the collaboration. This is what I thought was the very beginning. But then I realized one thing when I was actually walking on the campus this morning. That of course, biology is fascinating.

Life is fascinating. When I came to Aalto, I had an option to choose one of those kind of elective courses. And I couldn't tell what I wanted to take. So I sat in at the course on bioinformatics. Now, back then, I was...

Eric:
So the one that you took in Korea hadn't discouraged you too much.

Kyunghyun:
Not at all. Yeah. That one did not teach me anything. So I sat in and I was like, what is this? Let's learn about this. I remember that they were talking a lot about the graphs and trying to learn their interactions and all those different techniques. I got completely lost. I was so interested in the whole concept, but I couldn't follow anything, so I dropped off of the course.

So maybe I always wanted to do that. I just didn't have a right kind of skillset or the interest back then. But anyhow, so we found a stark similarity between language of life and language of humans. Then we started a little bit of a collaboration, 2018, 2019. And we realized that a lot of techniques that we were able to build for the machine translation, in particular the non-autoregressive machine translation, were applicable to the protein generation side. So we worked on it and we saw that there was a good potential here.

And then we had to make a decision. Do we actually do it at university or do we actually start a company and do it as a company? One major downside that we identified back then doing the same thing at the university was that because we were all computer scientists, despite the fact that my colleague Richard Bonneau was one of the original authors of the Rosetta tool, that is one of the most widely used protein modeling tools. We're all computer scientists. We didn't have overlap.

So then what we need. We need to write a grant proposal to NIH hoping that it's going to be funded eventually. Once it's funded, we're going to talk to our biology colleagues who have the wet labs, and then try to convince them to work with us, teaching some of their own efforts and working with us, and then trying to hire a postdoc who's going to work on this one. And perhaps within 3 to 4 years, we would have some proof of concept that it actually works. And I was like, no, that's too much. Let's start a company instead. So that's how I got into the biology and in particular the molecular biology there.

Eric:
That's interesting. So you're saying that methodologically, it was not just a big jump, maybe.

Kyunghyun:
Yeah. There was something that was shared.

Eric:
Is there any advice you would give to students who are thinking if they should sort specialize on a specific domain versus like branch out to many different areas?

Kyunghyun:
Yes. For students, I think it's always a good idea to focus on one problem or one domain at a time. Now I always encourage my own students, as well as the others to be ready to always switch the domains without having to worry too much about the consequences, because the uncertainty is anyway very, very high. When the uncertainty is high, then it doesn't make any sense to plan out based on how the society is going to change or how

the field is going to change. But the only thing that can guide them is their own interest, right? Because even if the field is going amazingly well, if you're not interested in this, it's like very horrifying. It's frustrating.

So I think it's a good idea to be able to switch the domains. But being focused on one domain at a time, I think is really important. But then over time, as you get more and more senior, you have more bandwidth and then you have know-how, knowledge and experience that allows you to minimize the disruption that is inevitable when you switch the domain. So thereby you can actually almost switch so quickly and fluidly to the point that it looks from the outside that people are working on multiple things in parallel. Despite the fact that it is actually faster context switching rather than doing everything in parallel. So for students, go deep, one at a time. But over time it's going to get easier to move into the new domains.

Eric:
Yeah. How do you generally think about selecting the problems to work on? I mean, you mentioned that machine translation was great, people find it motivating. But I guess nowadays it's somewhat solved. So what do you say to your students today? How should they pick?

Kyunghyun:
Yeah. Machine translation is definitely close to the state where we can declare that we may have actually solved the problem. Not perfectly yet. There are some corner cases that need some work. But generally for the broader population, absolutely, yes. And I can tell you why. Because my mom speaks Korean only. My dad speaks Korean only. They know some English, but not really much. My wife speaks only English and they actually talk to each other on WhatsApp using machine translation. And I'm actually on the same group chat. I try not to intervene. I try to see how they talk to each other. It kind of works, not perfectly, but it kind of works. I'm very happy about the whole state of the machine translation relative to, let's say 10, 15 years ago.

But I think that this is an example of how at least I choose problem. It's that it is very tempting to work on problems that I see have a good potential to be solved with my own contribution in a reasonable timeframe, so that I also feel the joy of the contribution. And then I move on to the next problem and so on. But I think that's actually a bad idea to go about. Because this requires one to predict the future. We need to forecast what is going to be more promising, what is going to be more interesting, or what is easier or faster to solve. And I don't think that kind of forecasting works that easily in general.

Rather, I think it's important to think about the consequences of those problems being solved or solved to a certain degree. What I had in mind was that if the machine translation was to be solved even to a degree that is slightly better than the current state back in 2014-2015, then that would actually make so many more people communicate with each other much more freely and be able to consume this amazing amount of digital materials that are online, regardless of the language barrier. So that was a consequence that I was looking at. I couldn't tell how long it was going to take. My guess

was anywhere between 5 to 50 years. I mean, it turned out it was on the lower end. But I had a massive, let's say, confidence interval there.

But thinking about the consequence seems to be a better way for us to be motivated better, and also to be able to choose the problems where I can actually contribute to have a better or the more positive consequence. Because making a progress doesn't mean that the consequences will be positive necessarily. So for students, and for others who are looking for a new problem, my suggestion is to think about the consequences of the solved problem or what's going to happen when the problem is solved.

Eric:
If we go from students to academia as a whole, how do you see the role of academia now that you know much of the so-called frontier model development happens on the industry side? So, how do you kind of decide what to do?

Kyunghyun:
That's true. I think there are two roles that we need to serve in academia. So the first role is that despite the fact that there are amazing progresses that are happening in these frontier labs, including the DeepMind, OpenAI, Anthropic, and Meta nowadays, and so on, that they are making progress at building products, in my view, does not necessarily translate to we as a humanity, having a good understanding of how those progresses are being made, and thereby being able to make even further progress in the future.

So moving fast forward really quickly does not necessarily mean that you will win in marathon kind of thing. We're in a sprint mode at the moment. We're actually forgetting the fact that there may be some other problems that we need to solve down the road. There are aspects that we need to understand in order to go from the 100 meter sprint to the marathon, but we're probably missing that part. And the academia's job could be to, in fact, kind of reverse engineer the current progress and trying to figure out what are the kind of fundamentals that we need to know in order to not get to the 100 meter line, but to get to the 42 kilometer line eventually down the road.

So that's the one thing that I think we need to do. For this one, I hope these frontier AI companies realize the importance of this activity as well, and then are able to contribute back to the society by funding the researchers, as well as funding the academic institutions as well.

But the other side is that, at the end of the day, the reason I think we as a society decided collectively to fund research activities and also have the universities host these research activities is because we have agreed on collectively that there are some surprising aspects or surprises that we want to see down the road. Because those surprises can actually contribute back to society in a positive way. But then it's impossible to plan out the surprises. It doesn't make any sense. By definition, it's a surprise. So the only thing we can do is to encourage people to work on things that are not predictable, so that there will be some manufactured surprises. And some of them, a very small number of those surprises will be actually impactful down the road.

So what that means is that anything that is predictable is probably not what we need to actually work on as a major direction. I want us, including myself, including you and then everyone here, to think about the things that are not predictable. When I talk to my students, they always tell me about these scaling laws and whatnot. I was like, great, that's awesome. But if you think that there is indeed that kind of law that tells us where we are going to get at after a certain amount of time is spent or after a certain amount of effort is spent, then that's exactly the thing that we should avoid working on. Because if that is so, why would we do it? There has to be an actual business or actual business-like activities that are done out there so that we as a society get to the point that is predicted.

What we need to do is to prepare for the future by working on the fundamentals. That includes education as well as interpreting what is happening, kind of how the progress was made. And second, to continue to go for this kind of unpredictable thing or things that we don't know how to predict yet. So I think that we have good roles to serve still. Yes.

Eric:
Okay. So talking a bit more about advice for students, you have an interesting blog post called I sensed anxiety and frustration at NeurIPS'24. I was also there that year. I think you talk about kind of the anxiety about the job market that students are facing. And of course, now with the AI, there's a lot of discussion of how junior roles potentially are going away. What's the advice you give today? Or how has this evolved since the two years?

Kyunghyun:
I always try to avoid giving any career advices because I feel like there's always something that's going to backfire a lot. I think that kind of the anxiety that I sensed in 2024 December, because it was NeurIPS, still continues. And there are a number of factors behind it.

One is, of course, the fact that there are so many more people working in this area. So despite the fact that the whole field of AI, both the research as well as the commercialization, has grown so much, when you are a part of this whole thing and what you see is not the field growing only, but also the participating, let's say you see how much more crowded the whole field is, thereby the anxiety seems to be continuing. Despite the fact that if you just divide the number of participants with the actual, let's say, size of the field, I'm pretty sure it is okay. But of course, individuals' anxiety has nothing to do with the objective nature. It's how one feels about it. So it seems like the anxiety continues.

And there are another factors that I actually blame, not the students nor the researchers, but a lot of these so-called thought leaders or business leaders. That they are sending out this kind of signal that things are going to get worse. And things are going to get worse especially for the junior developers or the new students who are graduating from the schools. This is kind of a false narrative, in my opinion, that exaggerates the

whole level of anxiety that the students, as well as the junior people, are feeling unnecessarily.

Why do I feel it's unnecessary? It's that a lot of these thought leaders or the leaders are saying that how AI is going to remove all those jobs, how no one will need to work or anything like that. But as if they are the ones who are making decisions for the people who are coming into the job markets or coming into the society. When in fact it is their choices, whether they want to work on this or that, or whether they want to work at all, or whether they want to be part of the society and what kind of roles they want to take. It's their roles. Not these so-called leaders' roles. But they are actually sending out false narratives that make people feel more anxious and frustrated.

So the first one, I don't think we can actually fix that people are feeling anxious because they feel that there is increasingly more competition in their job market, job search, as well as the everyday life. So that one continues. All I can tell people is that, well, everyone has always been frustrated and anxious about the job search as well as the study and whatnot. Unfortunately, that's kind of human nature. We compete with each other, so that will continue. But I hope they don't feel that this is something that unique that they are going through. So they don't have to feel too bad. That's what every generation has felt so far. And every generation will continue to feel. So you know, embrace it a bit.

Second thing, I think that we can fix it. And the fix is not going to be by the students or the new entrants to the society. But for people like us or the even more senior people to realize that our job is not to make decisions for them. Our job is to provide an environment where they can make their best decisions, and their best choices, and trying not to impose all that kind of doomerism on them. When they feel doomed, they try to impose it on the others. No, we should not do that. We should just try our best to provide the best environment for them to make their own choices. So in that sense, I don't have any career advice to the young people or anything like that. I think I have advice for myself and our generation, that we shouldn't try to impose anything on the next generation.

Eric:
That makes a lot of sense. That's kind of a fresh perspective. It's good to hear that from you.

Another risk that is brought up often with AI is the risk of de-skilling. If we kind of outsource our thinking more and more to these tools, what is it going to do to our cognitive capabilities? Maybe I'll ask it from the point of view of students again. So what do you tell your students? Like how much should they use the agentic tools in their research? And how should they not use these?

Kyunghyun:
Yeah. I tell my students, both my PhD students as well as the students in my courses, to use all these tools as much as they want. In fact, I tell them, you better use them while you're at school. Because you know, when people start using a new set of tools, they

inevitably make mistakes here and there. I want them to make mistakes when they're under the roof of the university or the schools, rather than when they're on the job. That can be devastating.

So I tell them to use it as much as they want. In fact, in my view, I'm not actually worried about the de-skilling. The reason is a lot of the professors, in particular, my professor colleagues at NYU, as well as the other universities, they all complain or they are all concerned how they should teach the students when all these AI tools are available to them.

I'm actually not concerned at all. I understand why they are concerned. This reason is actually completely incorrect. So they are concerned because they are trying to teach the students in this new era, the same amount of knowledge that they were taught when they were students without these kinds of tools. Thereby, they are worried that if we teach the same amount of knowledge, then the students will just simply ask ChatGPT or whatnot and then get the answer immediately. So that's why they are worried.

But what I tell them, and what I tell my students as well, is it's not about the amount of knowledge that we need to preserve over time. That's crazy. If you think about the people 1000 years ago, what they were learning was nothing. The amount of knowledge that we teach our students and we learn ourselves has dramatically improved. Then one needs to be preserved.

In my view, we need to strive to preserve the amount of struggles that students go through, amount of suffering that they go through. Then they'll be able to learn 10 x 100 x the amount of knowledge that we were able to learn, because now they have the amazing tools. So we just need to let them struggle as much as we struggled when we grew up. Then it's not going to be the problem of de-skilling. We'll have to worry about how the next generation knows so much more than we do, so that we get actually completely obsoleted.

Eric:
I got the same advice actually, as a junior, that it doesn't matter too much what you study as long as it's hard enough and forces you to push yourself.

Okay, switching gears, on a more personal side. You've been actually writing very openly in your blog about deeply personal topics such as facing health issues, battling with thyroid cancer, and the loss of your friend. What would you say, has been faced with our mortality affected the way that you approach science in your work? And has being a scientist changed the way you approach our mortality?

Kyunghyun:
Yeah. So one thing that I have realized in few occasions that I'm going to talk about is the fact that the whole human society is really diverse in terms of the situations people are in and the consequences people are actually experiencing. One of them was, for instance, that I had some health scares here and there. Before the thyroid cancer, I had Ramsay Hunt syndrome. That is the shingles that happen inside the face. It hurts a lot.

Then because I couldn't tell why I was having that kind of headache nonstop for a week and couldn't sleep at all I went to the emergency room a couple of times.

So when you go to the ER in New York City, in the night in particular, you see everything. Like the consequences of the whole spectrum of what kind of things people do and what kind of things people suffer from. And that's when you see that how I actually experience New York City as well as the workplaces in New York City and whatnot, is extremely narrow. It's extremely limited. And the kind of lifestyles or the lives that everyone actually experiences are so different.

And this actually gave me the sense that when I think about these problems and the consequences, I shouldn't really think of things only from my own perspective, but I need to actually broaden my perspective. I had a similar experience when I moved to Finland from Korea as well. As I said at the very beginning, South Korea is extremely homogeneous country. We're almost like an island because of the North Korea. There's almost no exchange between South and North Korea. And then anything I saw about the foreign countries or the abroad was largely based on the US.

So when I came to Finland and I started to learn about how the people behave here or how people live here, I was so shocked. It was so different from what I kind of was implicitly assuming. That also gave me a much broader perspective in the world. And one of the most interesting things I noticed back then was that in South Korea, it's all about either the relationship with North Korea, relationship with China, or relationship with the US or Japan. That's the only thing that you hear on the international section or the global section of the newspapers. And then I came here and I started to read some of the newspapers and whatnot, as I was learning some Finnish. And what I realized is that most of the news articles in the international or the global politics section was about European countries and Russia. And I was like, what? This is a completely different world, right?

And then along this line, when I went to Montreal, to Quebec, so it's kind of, say the isolated region of Canada as well. So everything that was mentioned there was again, very different from how it was in Finland and South Korea as well. So the reason why I tried to be open and transparent about the experiences I have is that I learned so much about how different everyone is, and every society is. And that gave me a broader and broader perspective. I feel like I can still get an even broader perspective. So I try to always be where I haven't been to, trying to see how people are working. And the reason why I'm being transparent is I want that the small number of people who read my blogs or who talk to me to get motivated or to feel that they can and they should also try to experience more and then see the broader context.

And then this really helps me choose the problems that I want to solve. And that's the reason why I try not to leave or relocate to Bay area. You know, you tend to actually have an even narrower view after living there, it seems like.

Eric:

Yeah. And by the way, I recommend people to read your blog. There's a lot of gems there that people can learn from. I mean, when preparing for this interview, I was thinking your kind of, if it's fair to call pivot, to this biology and nowadays you're a professor of health statistics. Was there any connection between your own sort of health issues you were tackling with?

Kyunghyun:
Yeah. More recently, one of the areas that I'm actually quite interested in, since about 2021 or so, is not the AI for the biology or the treatment or the diagnosis, but more like AI for the whole health care system itself. This is not about the thyroid cancer. But a couple of months ago, I got a letter with a check inside that has about $300 for me. I was like, this is interesting. And who sent it to me was the dentist that I used to go to that has closed down already. And then it has a very concise, very short description saying that apparently they owed me $300 because I had an Invisalign a few years back, and then apparently I paid more. But it took them about three years to resolve it and then send the money back to me.

Now because I'm, you know, financially pretty well off now so I'm okay with it and I didn't even notice that I was missing $300. But when you look at the health care, it's like no one opts out of the health care. No one decides that they are not going to be sick. People get sick, or people get into an accident without actually planning. So what that means is that there are people who are going to be very sensitive to this $300 missing for three years. But this is a really weird thing, that yes, everyone needs to go through this kind of process at one point or another. Sometimes they have cancer. They need to get some kind of, let's say, surgery like I did six years ago. So all those things happen.

But somehow to support that kind of system, the back end is extremely inefficient to the point that we don't know whether any treatment that is given to the patient by the doctors is even going to be paid for, or is going to have a better outcome. We often don't know, and then we just follow the standard procedures. That is called the standard of care or the guidelines from the various medical associations. But it's not really individualized.

What that means is that we need to do much better. And I think the AI is really giving us an opportunity. But unfortunately, the back end is too messy for us to think about how to use these kind of data-driven approaches. And then my experiences of being at the hospitals here and there, at the ER here and there, and then being at the hospitals in different countries, gives me a perspective. It's that one of the major problems that we can solve is to use AI in order to make that kind of system better. It's going to be a very messy job that's not going to be appreciated, nor sexy enough for a lot of people to jump in.

But these are kind of hidden problems that you don't get to see until you are like part of the hospital. Then you see all those random bills here and there, like, what is going on here? Do I owe them like $500 000? But then a week later, you get a bill saying that no, you owe us, say $50. What happened in between? No one actually knows. So these experiences that I had have been kind of pushing me into the direction of thinking

initially about the diagnosis, the treatments or the therapeutics, but now more like thinking about the whole system. Because I was able to benefit from the whole healthcare system. But as we know, not a lot of people can benefit from a healthcare system at the same level. And often we miss out on these kind of rare diseases as well as the conditions that are just not part of the standardized care. Then what we see is that they often are the ones that suffer in exchange of a kind of average case and care. So how do you actually make it much more personalized, much more individualized? That's the thing that I'm thinking a lot about these days.

Eric:
Okay. It's been a fantastic conversation. We usually finish this by asking what your P(doom) is, but maybe I'll just rephrase it a bit and ask, do you generally feel optimistic or more pessimistic about the potential of AI and the kind of developments that you see?

Kyunghyun:
I'm super optimistic. I mean, I'm generally an optimist wherever I am. I'm on like the far extreme of optimism. I'm optimistic for two reasons. One is that, yes, I am an optimist by nature. But the second reason is that, there is definitely, or of course, I do acknowledge that there can be horrible scenarios. Absolutely. But what I see is that all those scenarios, good to bad, are so uncertain to the point that we are actually proactively making things happen or not happen. Because we are not passive observers. In fact, everyone, not only the AI researchers and engineers, but everyone is an active participant in this era.

Because if the AI is going to be used everywhere and then used for everything, then what that means is we are all active participants. And I do believe that the humanity has been always moving into slightly more optimistic, better directions. That's the reason why we are here, compared to, I don't know, hundreds or thousands of years. So yes, this kind of, let's say, track record actually gives me the hope. And that makes me much more optimistic. We will be able to, in fact, intervene directly ourselves to make the better future happen than the worst one.

Eric:
Well, that's a great place to end. Thank you so much for coming.

Kyunghyun:
Thanks for the invitation. Kiitos.