Yann LeCun - Evolving AI Q&A

Q: If you could just introduce yourself, that would be great.

A: My name is Yann LeCun. I'm the Chief AI Scientist at Meta and I'm also a professor at New York University.

Q: How long have you worked at Meta?

A: A little over 10 years.

Q: And what's your favourite thing about working at Meta?

A: Openness.

Q: We'll dive into the questions. The first one is, given the rapid advancements in AI, how do you envision its role in shaping the future of our society and the global community?

A: Short term, it's easier to predict but long term the impact is going to be very, very broad and to some extent difficult to predict, but my picture of it is, every single one of our interactions with the digital world and largely with each other are going to be mediated by AI assistants. And so we're going to be talking to them through our smart glasses and smartphones, but they're going to be with us at all times, helping us, so to some extent they're going to amplify our own intelligence. It's as if everyone will have a staff of really smart virtual people working for them. So, we will all be managers, to some extent. And an AI assistant will take care of the details, execute for us. But we'll set their goals, they're not going to work off their own backs. What that means globally, at the level of humanity, we will be smarter because we're going to be augmented by those AI systems. And the

transformation of society might be similar to what happened after the invention of the printing press. In the 15th century, people got smarter, they started to learn to read; we have the enlightenment; we also had the feudal system which was then displaced by democracy – there could be a similar trend here.

Q: So, that kind of links to one of the questions that we've got, which is that do you see AI integration in daily tasks enhancing or hindering human creativity? I guess taking away the admin from our lives, which would enhance creativity. But what's your view on that?

A: It will certainly enhance creativity and allow people to be more creative. For example, people can now produce music but they couldn't do before given they couldn't play an instrument, but now they can use computers and digital audio workstations. Similarly with digital art, AI can produce art in ways that gets around the technicality of it.

Q: What are some of the projects that you're working on that you're most excited about at the company?

A: The projects I work on personally are fairly long term. They are about the next generation of AI systems. Over the next few years, we're going to see progress in systems that can answer questions, can help you in your daily life like AI assistants, help you write texts, create imagery, audio, music, videos etc. This is going to be relatively continuous, but there's probably going to be a major change with the underlying architecture of AI systems that will allow them to do things that they can't do currently; understanding the physical world, having persistent memory, being able to plan and reason. So that's what I'm working on: how to get systems to learn how the world works, like, you know, human babies or animals.

Q: How do language models work exactly?

A: They're primarily language models. The way language models are trained is, you take a piece of text and you remove some of the words, so you train some gigantic neural net system to predict the words that are missing. In some cases, you can constrain it so that it can only predict a word from the words that are to the left of it. So if you have a system like this that's been trained, it basically has learned a lot about language and grammar and semantics- and then you can fine tune it for a particular application, answering questions in a particular domain. In Meta's case, detecting hate speech or bullying. You can also use them to generate text because they've been trained to produce a word from the previous words You produce next word, and then you shift that word into the input, and then produce the second next word, and shift that into the input and so on. So they can spit out one word after the other without really thinking, which is why they can hallucinate..

Q: And what the challenges involved with that?

A: The challenge is that it's relatively easy to predict the next word in a text - you can't really predict exactly which word is missing, but you can produce what amounts to a probability distribution of all possible words in the dictionary. For example, if the cat chases the 'blank' in the kitchen, you know the 'blank' *could* be a 'mouse'; of course it could be a laser pointer, or a toy, but there's only a small number of things and likely to be a noun. This is where you can have a computer produce a probability for every word in the dictionary. So we know how to handle uncertainty in prediction. Now, if you do the same thing with video, you have a cat running in a kitchen, and you start the video and you ask what's going to happen next. And you ask the system to predict every pixel in the video, there's absolutely no way it can do this.This prediction task is impossible. There's just too many things that could plausibly happen and we don't know how to represent them on the computer. The solution to this is

not to have the system predict every pixel in the video, but have it basically learn an abstract representation of what goes on in the video and make the prediction in this abstract level. And that's much easier, but here's the funny thing: that's not a generative model. So the future of AI is non-generative, which goes a little against the current chatter.

Q. As AI continues to advance, how do you propose we manage its risks—such as those highlighted by concerns over AI safety and control—while also staying critically engaged with its development as it begins to match and possibly surpass human intelligence?

A: Currently our systems are very far from matching human intelligence. At the moment, they can't really invent new things at least current LLMs. And the reason they can't do this is because they're trained purely on language, and language, it turns out, is only a tiny portion of human knowledge. Most of what you learn in the first few months of life has nothing to do with language, and similarly, everything that animals learn has nothing to do with language.

To put it into perspective, consider this simple calculation: LLMs are trained on approximately 10 trillion tokens, which can be equated to words or sub-word units. Each token is about two bytes, so that amounts to roughly 10 trillion multiplied by 13 bytes. If you were to read through all of that, it would take you 100,000 years. Essentially, it encompasses the entirety of publicly available texts on the internet. Now, consider the information intake of a young child. By age four, they've been awake for 16,000 hours. The optical nerve processes about 20 megabytes per second, translating to 10 to 15 bytes through the visual cortex in four years—50 times more than the largest LLM. In just 300 hours of sight, babies absorb more than these models.

Our intelligence mostly arises from our knowledge of the physical world and interactions within it, and language is kind of an epiphenomenon even though it's a very important one for humans. Attaining human-level AI solely through text-based training is improbable. The real challenge is translating intelligence into actions derived from real-world interactions. Progress toward human-level AI will be gradual, starting with AI levels akin to rats or cats, and advancing with added guardrails. It's unrealistic to think that one day someone will discover the secret of AGI, turn on a computer, and suddenly humanity is annihilated.

Q. As we move closer to achieving AGI, what key scientific breakthroughs or milestones do you believe are still required to achieve it?

A: No, it won't happen suddenly. Basic concepts may emerge over the next five to ten years, with ongoing progress but no sudden event. Its akin to the development of turbo-jet engines, which took decades to advance from frequent failures to reliable, long-distance travel. Similarly, AI will likely take years, if not decades, to evolve. Future AI systems will differ significantly from current ones, capable of planning and imagining consequences. These objective-driven AI systems can be guided by guardrails to ensure safety, unlike humans who can validate laws.

Q. How do you see AI contributing to our understanding of the universe and solving longstanding scientific mysteries?

AI offers hope and initial successes in understanding various phenomena, not just in physics or chemistry but also in biology and society Wherever complex interactions lead to emergent phenomena, traditional models struggle to predict ecological properties.

For example, the mysterious behaviour of materials like graphene when twisted at a certain angle to become superconductors defies reductionist theories.

Here, AI can help predict interactions, such as how water molecules interact with substrates, aiding in processes like hydrogen separation, which could be vital for combating climate change through sustainable energy production.

Q. What role do you envision for AI in addressing fundamental global challenges like climate change and wars?

A: Researchers are exploring AI's potential in predicting catalyst properties and discovering new materials, such as for batteries, where traditional methods struggle due to the need for rare-elements like lithium. AI could revolutionize battery efficiency on a large scale.

AI might also aid in controlling plasma for fusion reactors, a challenge persisting for decades despite understanding its principles.

Could it help with carbon capture or things like that? So a lot of hope for the use of AI in material science. AI also shows promise in carbon capture and in medicine for understanding mechanisms of biology, of interactions between proteins, for example. So there are a lot of people working on this, and it's very fascinating.

Q: And what are the AI products that people can go and use at the moment on Meta platforms that you think are the best ways to engage with AI?

In the US there is something called Meta AI. I don't believe it's been deployed in Europe or the UK yet for regulatory complexities.

It's essentially dialogue system, accessible through various smart devices, like the glasses I'm wearing right now, the Ray Ban Meta. They have a camera and microphones that allow voice interaction and can provide information or translations in real time.

In the near future, we'll have smart letters with displays in them so they all have simultaneous translation capabilities which will become widespread, akin to the movie "Her" from 2013, where AI becomes integral to daily interactions,

Q: Meta has taken a very strong position on open source and its large language models, so, could you tell us a bit more about why you've done that and perhaps answer some of the critics that have criticised this decision?

A: Yes. Firstly, our company's DNA is rooted in open-sourcing, spanning over fifteen years, encompassing platform infrastructure, software, and even hardware design– if not more - including AI platforms that almost everybody uses to develop AI systems.The aim is to foster community contributions to enhance and refine our platforms for various purposes, benefiting from collective insights and feedback.

The second thing is,it creates an entire ecosystem on top of which a complete industry can be built.

For instance, the release of Lama2, our open source LLM, has catalysed the emergence of numerous startups specializing in fine-tuning LLMs for diverse applications, such as adapting them to speak different languages or addressing specific industry needs,

So, for example, there's a project in India to fine tune llama 2 (our LLM) so that it speaks all 22 official languages in India, which of course is great.

Q: Can you expand on translation tools for

A: We have some other projects facilitating translation for languages without written form, preserving local dialects, and enabling voice-based access for those unable to read.

Millions have downloaded our models, adapting them to their language, culture and specific needs, So, this is a huge, enormous impact.

And there is another reason. Envision a future where all digital interactions involve AI assistance.

We cannot afford those AI systems to come from two or three companies on the West Coast of the US, because you know, they've been trained with mostly English there, i mean there's also French or German or Spanish. But what about all the dialects in India for example?

So you can't expect this number of companies to have the kind of wide variety of assistance necessary. Also if all of our interactions are mediated by the same systems, they will carry particular value systems, particular political opinions, particular things like that.

So we need a very wide diversity of AI assistants – for the same reason that we need a free and diverse press. The protection of democracies and cultural diversity. So, for that reason, open source AI platforms, I think are not just desirable, but are necessary.

Contrary to calls for AI regulation to prevent misuse, restricting access could exacerbate risks by limiting diversity of opinion among AI assistants.

The proponents of regulations aiming to restict open-source AI often cite concerns about its potential misuse by terrorist groups or or unsavoury characters. But that's true of every technology ever invented. So there's nothing new in that respect.

I think the danger of closing it off are bigger. The danger would be larger than the slight increase of safety that it would provide.