ARTIFICIAL INTELLIGENCE

Yann LeCun has a bold new vision for the future of AI

One of the godfathers of deep learning pulls together old ideas to sketch out a fresh path for AI, but raises as many questions as he answers.

By

- **Melissa Heikkiläarchive page**
- **Will Douglas Heavenarchive page**

June 24, 2022

BRIAN ACH/GETTY IMAGES

Around a year and a half ago, Yann LeCun realized he had it wrong.

LeCun, who is chief scientist at Meta's AI lab and a professor at New York University, is one of the most influential AI researchers in the world. He had been trying to give machines a basic grasp of how the world works—a kind of common sense—by training neural networks to predict what was going to happen next in video clips of everyday events. But guessing future frames of a video pixel by pixel was just too complex. He hit a wall.

Now, after months figuring out what was missing, he has a bold new vision for the next generation of AI. In a draft document shared with MIT Technology Review, LeCun sketches out an approach that he thinks will one day give machines the common sense they need to navigate the world. (Update: LeCun has since posted the document online.)

For LeCun, the proposals could be the first steps on a path to building machines with the ability to reason and plan like humans—what many call artificial general intelligence, or AGI. He also steps away from today's hottest trends in machine learning, resurrecting some old ideas that have gone out of fashion.

But his vision is far from comprehensive; indeed, it may raise more questions than it answers. The biggest question mark, as LeCun points out himself, is that he does not know how to build what he describes.

Related Story

Artificial general intelligence: Are we close, and does it even make sense to try?

A machine that could think like a person has been the guiding vision of AI research since the earliest days—and remains its most divisive idea.

The centerpiece of the new approach is a neural network that can learn to view the world at different levels of detail. Ditching the need for pixel-perfect predictions, this network would focus only on those features in a scene that are relevant for the task at hand. LeCun proposes pairing this core network with another, called the configurator, which determines what level of detail is required and tweaks the overall system accordingly.

For LeCun, AGI is going to be a part of how we interact with future tech. His vision is colored by that of his employer, Meta, which is pushing a virtual-reality metaverse. He says that in 10 or 15 years people won't be carrying smartphones in their pockets, but augmented-reality glasses fitted with virtual assistants that will guide humans through their day. "For those to be most useful to us, they basically have to have more or less human-level intelligence," he says.

"Yann has been talking about many of these ideas for some time," says Yoshua Bengio, an AI researcher at the University of Montreal and scientific director at the Mila-Quebec Institute. "But it is good to see it all together, in one big picture." Bengio thinks that LeCun asks the right questions. He also thinks it's great that LeCun is willing to put out a document that has so few answers. It's a research proposal rather than a set of clean results, he says.

"People talk about these things in private, but they're not usually shared publicly," says Bengio. "It's risky."

A matter of common sense

LeCun has been thinking about AI for nearly 40 years. In 2018 he was joint winner of computing's top prize, the Turing Award, with Bengio and Geoffrey Hinton, for his pioneering work on deep learning. "Getting machines to behave like humans and animals has been the quest of my life," he says.

LeCun thinks that animal brains run a kind of simulation of the world, which he calls a world model. Learned in infancy, it's the way animals (including humans) make good guesses about what's going on around them. Infants pick up the basics in the first few months of life by observing the world, says LeCun. Seeing a dropped ball fall a handful of times is enough to give a child a sense of how gravity works.

"Common sense" is the catch-all term for this kind of intuitive reasoning. It includes a grasp of simple physics: for example, knowing that the world is three-dimensional and that objects don't actually disappear when they go out of view. It lets us predict where a bouncing ball or a speeding bike will be in a few seconds' time. And it helps us join the dots between incomplete pieces of information: if we hear a metallic crash from the kitchen, we can make an educated guess that someone has dropped a pan, because we know what kinds of objects make that noise and when they make it.

In short, common sense tells us what events are possible and impossible, and which events are more likely than others. It lets us foresee the consequences of our actions and make plans—and ignore irrelevant details.

But teaching common sense to machines is hard. Today's neural networks need to be shown thousands of examples before they start to spot such patterns.

In many ways common sense amounts to the ability to predict what's going to happen next. "This is the essence of intelligence," says LeCun. That's why he—and a few other researchers—have been using video clips to train their models. But existing machine-learning techniques required the models to predict exactly what is going to happen in the next frame and generate it pixel by pixel. Imagine you hold up a pen and let it go, LeCun says. Common sense tells you that the pen will fall, but not the exact position it will end up in. Predicting that would require crunching some tough physics equations.

That's why LeCun is now trying to train a neural network that can focus only on the relevant aspects of the world: predicting that the pen will fall but not exactly how. He sees this trained network as the equivalent of the world model that animals rely on.

Mystery ingredients

LeCun says he has built an early version of this world model that can do basic object recognition. He is now working on training it to make predictions. But how the configurator should work remains a mystery, he says. LeCun imagines that neural network as the controller for the whole system. It would decide what kind of predictions the world model should be making at any given time and what level of detail it should focus on to make those predictions possible, adjusting the world model as required.

LeCun is convinced that something like a configurator is needed, but he doesn't know how to go about training a neural network to do the job. "We need to figure out a good recipe to make this work, and we don't have that recipe yet," he says.

In LeCun's vision, the world model and the configurator are two key pieces in a larger system, known as a cognitive architecture, that includes other neural networks—such as a perception model that senses the world and a model that uses rewards to motivate the AI to explore or curb its behavior.

Each neural network is roughly analogous to parts of the brain, says LeCun. For example, the configurator and world model are meant to replicate functions of the prefrontal cortex. The motivation model corresponds to certain functions of the amygdala, and so on.

The idea of cognitive architectures, especially ones inspired by the brain, has been around for decades. So have many of LeCun's ideas about prediction using models with different levels of detail. But when deep learning became the dominant approach in AI, many of these older ideas went out of fashion. "People in AI research have kind of forgotten about this a little bit," he says.

What he has done is taken these older ideas and rehabilitated them, suggesting ways that they can be combined with deep learning. For LeCun, revisiting these out-of-fashion ideas is essential, because he believes the two dominant approaches in modern AI are dead ends.

When it comes to building general-purpose AI, there are two main camps. In one, many researchers think the remarkable success of very large language or image-making models like OpenAI's GPT-3 and DALL-E show that all we need to do is just build bigger and bigger models.

In the other camp are champions of reinforcement learning, the AI technique that rewards specific behaviors to make neural networks learn by trial and error. This is the approach DeepMind used to train its game-playing AIs like AlphaZero. Get the rewards right, the argument goes, and reinforcement learning will eventually produce more general intelligence.

Related Story

Why GPT-3 is the best and worst of AI right now

Open AI's language AI wowed the public with its apparent mastery of English – but is it all an illusion?

LeCun is having none of it: “This idea that we're going to just scale up the current large language models and eventually human-level AI will emerge—I don’t believe this at all, not for one second.” These large models just manipulate words and images, he says. They have no direct experience of the world.

He is equally skeptical about reinforcement learning, because it requires vast amounts of data to train models to do even simple tasks. “I think that has no chance of working at all,” says LeCun.

David Silver at DeepMind, who led the work on AlphaZero and is a big advocate of reinforcement learning, disagrees with this assessment but welcomes LeCun’s overall vision. “It’s an exciting new proposal for how a world model could be represented and learned,” he says.

Melanie Mitchell, an AI researcher at the Santa Fe Institute, is also excited to see a whole new approach. “We really haven’t seen this coming out of the deep-learning community so much,” she says. She also agrees with LeCun that large language models cannot be the whole story. “They lack memory and internal models of the world that are actually really important,” she says.

Natasha Jaques, a researcher at Google Brain, thinks that language models should still play a role, however. It’s odd for language to be entirely missing from LeCun’s proposals, she says: “We know that large language models are super effective and bake in a bunch of human knowledge.”

Jaques, who works on ways to get AIs to share information and abilities with each other, points out that humans don't have to have direct experience of something to learn about it. We can change our behavior simply by being told something, such as not to touch a hot pan. "How do I update this world model that Yann is proposing if I don't have language?" she asks.

There's another issue, too. If they were to work, LeCun's ideas would create a powerful technology that could be as transformative as the internet. And yet his proposal doesn't discuss how his model's behavior and motivations would be controlled, or who would control them. This is a weird omission, says Abhishek Gupta, the founder of the Montreal AI Ethics Institute and a responsible-AI expert at Boston Consulting Group.

"We should think more about what it takes for AI to function well in a society, and that requires thinking about ethical behavior, amongst other things," says Gupta.

Yet Jaques notes that LeCun's proposals are still very much ideas rather than practical applications. Mitchell says the same: "There's certainly little risk of this becoming a human-level intelligence anytime soon."

LeCun would agree. His aim is to sow the seeds of a new approach in the hope that others build on it. "This is something that is going to take a lot of effort from a lot of people," he says. "I'm putting this out there because I think ultimately this is the way to go." If nothing else, he wants to convince people that large language models and reinforcement learning are not the only ways forward.

"I hate to see people wasting their time," he says.

hide

by **Melissa Heikkilä** **&** **Will Douglas Heaven**