

The AI technique that could imbue machines with the ability to reason

Karen Hao

At six months old, a baby won't bat an eye if a toy truck drives off a platform and [seems to hover in the air](#). But perform the same experiment a mere two to three months later, and she will instantly recognize that something is wrong. She has already learned the concept of gravity.

"Nobody tells the baby that objects are supposed to fall," said Yann LeCun, the chief AI scientist at Facebook and a professor at NYU, [during a webinar](#) on Thursday organized by the Association for Computing Machinery, an industry body. And because babies don't have very sophisticated motor control, he hypothesizes, "a lot of what they learn about the world is through observation."

That theory could have important implications for researchers hoping to advance the boundaries of artificial intelligence.

[Deep learning](#), the category of AI algorithms that kick-started the field's most recent revolution, has made immense strides in giving machines perceptual abilities like vision. But it has fallen short in imbuing them with sophisticated reasoning, grounded in a conceptual model of reality. In other words, machines don't truly understand the world around them, which makes them fall short in their ability to engage with it. New techniques are helping to overcome this limitation—for example, by giving machines a kind of working memory so that as they learn and derive basic facts and principles, they can accumulate them to draw on in future interactions.

But LeCun believes that is only a piece of the puzzle. "Obviously we're missing something," he said. A baby can develop an understanding of an elephant after seeing two photos, while [deep-learning algorithms](#) need to see thousands, if not millions. A teen can learn to drive safely by practicing for 20 hours and manage to avoid crashes without first experiencing one, while [reinforcement-learning algorithms](#) (a subcategory of deep learning) must go through tens of millions of trials, including many egregious failures.

The answer, he thinks, is in the underrated deep-learning subcategory known as unsupervised learning. While algorithms based on supervised and reinforcement learning are taught to achieve an objective through human input, unsupervised ones extract patterns in data entirely on their own. (LeCun prefers the term "self-supervised learning" because it essentially uses part of the training data to predict the rest of the training data.)

In recent years, such algorithms have [gained significant traction](#) in natural-language processing because of their ability to find the relationships between billions of words. This proves useful for building text prediction systems like autocomplete or for generating [convincing prose](#). But the vast majority of AI research in other domains has focused on

supervised or reinforcement learning.

LeCun believes the emphasis should be flipped. “Everything we learn as humans—almost everything—is learned through self-supervised learning. There’s a thin layer we learn through supervised learning, and a tiny amount we learn through reinforcement learning,” he said. “If machine learning, or AI, is a cake, the vast majority of the cake is self-supervised learning.”

What does this look like in practice? Researchers should begin by focusing on temporal prediction. In other words, train large neural networks to predict the second half of a video when given the first. While not everything in our world can be predicted, this is the foundational skill behind a baby’s ability to realize that a toy truck should fall. “This is kind of a simulation of what’s going on in your head, if you want,” LeCun said.

Once the field develops techniques that refine those abilities, they will have important practical uses as well. “It’s a good idea to do video prediction in the context of self-driving cars because you might want to know in advance what other cars on the streets are gonna do,” he said.

Ultimately, unsupervised learning will help machines develop a model of the world that can then predict future states of the world, he said. It’s a lofty ambition that has eluded AI research but would open up an entirely new host of capabilities. LeCun is confident: “The next revolution of AI will not be supervised.”

*To have more stories like this delivered directly to your inbox, [sign up](#) for our Webby-nominated AI newsletter *The Algorithm*. It's free.*

