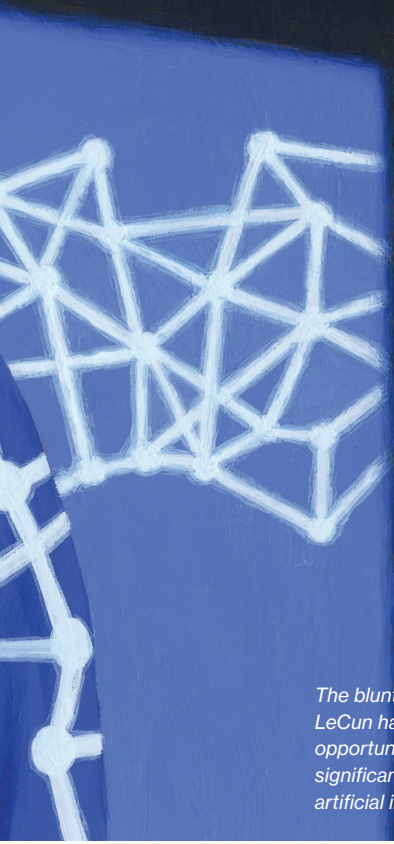# Teaching Machines to Understand Us

A reincarnation of one of the oldest ideas in artificial intelligence could finally make it possible to truly converse with our computers. And Facebook has a chance to make it happen first.

By Tom Simonite

The blunt and puckish LeCun has a second opportunity to significantly advance artificial intelligence.

THOMAS EHRETSMANN

The first time Yann LeCun revolutionized artificial intelligence, it was a false dawn. It was 1995, and for almost a decade, the young Frenchman had been dedicated to what many computer scientists considered a bad idea: that crudely mimicking certain features of the brain was the best way to bring about intelligent machines. But LeCun had shown that this approach could produce something strikingly smart—and useful. Working at Bell Labs, he made software that roughly simulated neurons and learned to read handwritten text by looking at many different examples. Bell Labs' corporate parent, AT&T, used it to sell the first machines capable of reading the handwriting on checks and written forms. To LeCun and a few fellow believers in artificial neural networks, it seemed to mark the beginning of an era in which machines could learn many other skills previously limited to humans. It wasn't.

"This whole project kind of disappeared on the day of its biggest success," says LeCun. On the same day he celebrated the launch of bank machines that could read thousands of checks per hour, AT&T announced it was splitting into three companies dedicated to different markets in communications and computing. LeCun became head of research at a slimmer AT&T and was directed to work on other things; in 2002 he would leave AT&T, soon to become a professor at New York University. Meanwhile, researchers elsewhere found that they could not apply LeCun's breakthrough to other computing problems. The brain-inspired approach to AI went back to being a fringe interest.

LeCun, now a stocky 55-year-old with a ready smile and a sideways sweep of dark hair touched with gray, never stopped pursuing that fringe interest. And remarkably, the rest of the world has come around. The ideas that he and a few others nurtured in the face of over two decades of apathy and sometimes outright rejection have in the past few years produced striking results in areas like face and speech recognition. Deep learning, as the field is now known, has become a new battleground between Google and other leading technology companies that are racing to use it in consumer services. One such company is Facebook, which hired LeCun from NYU in December 2013 and put him in charge of a new artificial-intelligence research group, FAIR, that today has 50 researchers but will grow to 100. LeCun's lab is Facebook's first significant investment in fundamental research, and it could be crucial to the company's attempts to become more than just a virtual social venue. It might also reshape our expectations of what machines can do.

Facebook and other companies, including Google, IBM, and Microsoft, have moved quickly to get into this area in the past few years because deep learning is far better than previous AI techniques at getting computers to pick up skills that challenge machines, like understanding photos. Those more established techniques require human experts to laboriously program certain abilities, such as how to detect lines and corners in images. Deep-learning software figures out how to make sense of data for itself, without any such programming. Some systems can now recognize images or faces about as accurately as humans.

Now LeCun is aiming for something much more powerful. He wants to deliver software with the language skills and common sense needed for basic conversation. Instead of having to communicate with machines by clicking buttons or entering carefully chosen search terms, we could just tell them what we want as if we were talking to another person. "Our relationship with the digital world will completely change due to intelligent agents you can interact with," he predicts. He thinks deep learning can produce software that understands our sentences and can respond with appropriate answers, clarifying questions, or suggestions of its own.

Agents that answer factual questions or book restaurants for us are one obvious—if not exactly world-changing—application. It's also easy to see how such software might lead to more stimulating video-game characters or improve online learning. More provocatively, LeCun says systems that grasp ordinary language could get to know us well enough to understand what's good for us. "Systems like this should be able to understand not just what people would be entertained by but

Deep Learning's Leaders

Geoff Hinton
Google and the University of Toronto
Did his PhD on artificial neural networks in the 1970s. Showed how to train larger, "deep" neural networks on large data sets in the 2000s, and proved their power for speech and image recognition.

Yann LeCun
Facebook
Got interested in neural networks as an undergraduate, and later pioneered the use of deep learning for image recognition. Now leads a group at Facebook trying to create software that understands text and can hold conversations.

Yoshua Bengio
University of Montreal
Started working on artificial neural networks after meeting LeCun at Bell Labs in the 1980s. Was one of the first to apply the technique to understanding words and language. Now working with IBM to improve its Watson software.

Andrew Ng
Baidu
Led a project at Google that worked out how neural networks could be trained on millions of pieces of data, allowing greater accuracy. Now oversees research at Baidu, which is working on improved speech recognition.

Demis Hassabis
DeepMind
Worked on AI in the games industry, then researched neuroscience to get ideas about building intelligence. He founded DeepMind, which Google bought last year and runs as a quasi-independent unit.

what they need to see regardless of whether they will enjoy it," he says. Such feats aren't possible using the techniques behind the search engines, spam filters, and virtual assistants that try to understand us today. They often ignore the order of words and get by with statistical tricks like matching and counting keywords. Apple's Siri, for example, tries to fit what you say into a small number of categories that trigger scripted responses. "They don't really understand the text," says LeCun. "It's amazing that it works at all." Meanwhile, systems that seem to have mastered complex language tasks, such as IBM's *Jeopardy!* winner Watson, do it by being super-specialized to a particular format. "It's cute as a demonstration, but not work that would really translate to any other situation," he says.

In contrast, deep-learning software may be able to make sense of language more the way humans do. Researchers at Facebook, Google, and elsewhere are developing software that has shown progress toward understanding what words mean. LeCun's team has a system capable of reading simple stories and answering questions about them, drawing on faculties like logical deduction and a rudimentary understanding of time.

However, as LeCun knows firsthand, artificial intelligence is notorious for blips of progress that stoke predictions of big leaps forward but ultimately change very little. Creating software that can handle the dazzling complexities of language is a bigger challenge than training it to recognize objects in pictures. Deep learning's usefulness for speech recognition and image detection is beyond doubt, but it's still just a guess that it will master language and transform our lives more radically. We don't yet know for sure whether deep learning is a blip that will turn out to be something much bigger.

Deep history

The roots of deep learning reach back further than LeCun's time at Bell Labs. He and a few others who pioneered the technique were actually resuscitating a long-dead idea in artificial intelligence.

When the field got started, in the 1950s, biologists were just beginning to develop simple mathematical theories of how intelligence and learning emerge from signals passing between neurons in the brain. The core idea—still current today—was that the links between neurons are strengthened if those cells communicate frequently. The fusillade of neural activity triggered by a new experience adjusts the brain's connections so it can understand it better the second time around.

In 1956, the psychologist Frank Rosenblatt used those theories to invent a way of making simple simulations of neurons in software and hardware. The *New York Times* announced his work with the headline "Electronic 'Brain' Teaches Itself."

Rosenblatt's perceptron, as he called his design, could learn how to sort simple images into categories—for instance, triangles and squares. Rosenblatt usually implemented his ideas on giant machines thickly tangled with wires, but they established the basic principles at work in artificial neural networks today.

One computer he built had eight simulated neurons, made from motors and dials connected to 400 light detectors. Each of the neurons received a share of the signals from the light detectors, combined them, and, depending on what they added up to, spit out either a *1* or a *0*. Together those digits amounted to the perceptron's "description" of what it saw. Initially the results were garbage. But Rosenblatt used a method called supervised learning to train a perceptron to generate results that correctly distinguished different shapes. He would show

Deep learning is good at taking dictation and recognizing images. But can it master human language?

the perceptron an image along with the correct answer. Then the machine would tweak how much attention each neuron paid to its incoming signals, shifting those "weights" toward settings that would produce the right answer. After many examples, those tweaks endowed the computer with enough smarts to correctly categorize images it had never seen before. Today's deep-learning networks use sophisticated algorithms and have millions of simulated neurons, with billions of connections between them. But they are trained in the same way.

Rosenblatt predicted that perceptrons would soon be capable of feats like greeting people by name, and his idea became a linchpin of the nascent field of artificial intelligence. Work focused on making perceptrons with more complex networks, arranged into a hierarchy of multiple learning layers. Passing images or other data successively through the layers would allow a perceptron to tackle more complex problems. Unfortunately, Rosenblatt's learning algorithm didn't work on multiple layers. In 1969 the AI pioneer Marvin Minsky, who had gone

to high school with Rosenblatt, published a book-length critique of perceptrons that killed interest in neural networks at a stroke. Minsky claimed that getting more layers working wouldn't make perceptrons powerful enough to be useful. Artificial-intelligence researchers abandoned the idea of making software that learned. Instead, they turned to using logic to craft working facets of intelligence—such as an aptitude for chess. Neural networks were shoved to the margins of computer science.

Nonetheless, LeCun was mesmerized when he read about perceptrons as an engineering student in Paris in the early 1980s. "I was amazed that this was working and wondering why people abandoned it," he says. He spent days at a research library near Versailles, hunting for papers published before perceptrons went extinct. Then he discovered that a small group of researchers in the United States were covertly working on neural networks again. "This was a very underground movement," he says. In papers carefully purged of words like "neural" and "learning" to avoid rejection by reviewers, they were working on something very much like Rosenblatt's old problem of how to train neural networks with multiple layers.

LeCun joined the underground after he met its central figures in 1985, including a wry Brit named Geoff Hinton, who now works at Google and the University of Toronto. They immediately became friends, mutual admirers—and the nucleus of a small community that revived the idea of neural networking. They were sustained by a belief that using a core mechanism seen in natural intelligence was the only way to build artificial intelligence. "The only method that we knew worked was a brain, so in the long run it had to be that systems something like that could be made to work," says Hinton.

LeCun's success at Bell Labs came about after he, Hinton, and others perfected a learning algorithm for neural networks with multiple layers. It was known as backpropagation, and it sparked a rush of interest from psychologists and computer scientists. But after LeCun's check-reading project ended, backpropagation proved tricky to adapt to other problems, and a new way to train software to sort data was invented by a Bell Labs researcher down the hall from LeCun. It didn't involve simulated neurons and was seen as mathematically more elegant. Very quickly it became a cornerstone of Internet companies such as Google, Amazon, and LinkedIn, which use it to train systems that block spam or suggest things for you to buy.

After LeCun got to NYU in 2003, he, Hinton, and a third collaborator, University of Montreal professor Yoshua Bengio, formed what LeCun calls "the deep-learning conspiracy." To prove that neural networks would be useful, they quietly developed ways to make them bigger, train them with larger data sets, and run them on more powerful computers. LeCun's handwriting recognition system had had five layers of neurons, but now they could have 10 or many more. Around 2010, what was now dubbed deep learning started to beat established techniques on real-world tasks like sorting images. Microsoft, Google, and IBM added it to speech recognition systems. But neural networks were still alien to most researchers and not considered widely useful. In early 2012 LeCun wrote a fiery letter—initially published anonymously—after a paper claiming to have set a new record on a standard vision task was rejected by a leading conference. He accused the reviewers of being "clueless" and "negatively biased."

Everything changed six months later. Hinton and two grad students used a network like the one LeCun made for reading checks to rout the field in the leading contest for image recognition. Known as the ImageNet Large Scale Visual Recognition Challenge, it asks software to identify 1,000 types of objects as diverse as mosquito nets and mosques. The Toronto entry correctly identified the object in an image within five guesses about 85 percent of the time, more than 10 percentage points better than the second-best system (see Innovator Under 35 Ilya Sutskever, page 47). The deep-learning soft-

Journey to Acceptance

1956: Psychologist Frank Rosenblatt uses theories about how brain cells work to design the perceptron, an artificial neural network that can be trained to categorize simple shapes.

1969: AI pioneers Marvin Minsky and Seymour Papert write a book critical of perceptrons that quashes interest in neural networks for decades.

1986: Yann LeCun and Geoff Hinton perfect backpropagation to train neural networks that pass data through successive layers of artificial neurons, allowing them to learn more complex skills.

1987: Terry Sejnowski at Johns Hopkins University creates a system called NETtalk that can be trained to pronounce text, going from random babbling to recognizable speech.

1990: At Bell Labs, LeCun uses backpropagation to train a network that can read handwritten text. AT&T later uses it in machines that can read checks.

1995: Bell Labs mathematician Vladimir Vapnik publishes an alternative method for training software to categorize data such as images. This sidelines neural networks again.

LeCun at Bell Labs in 1993, with a computer that could read the handwriting on checks.
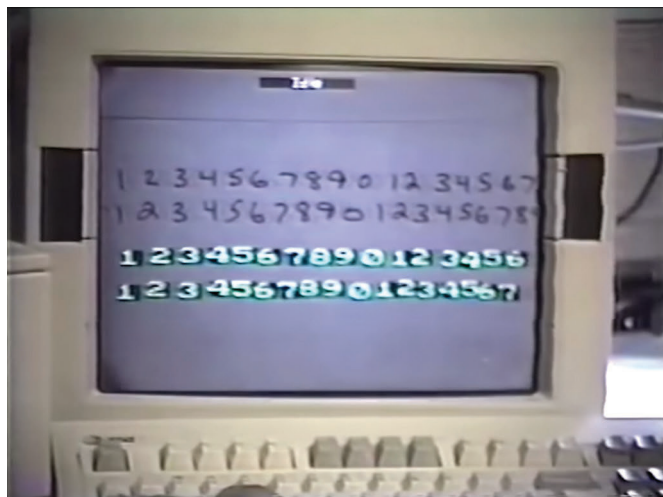
ware's initial layers of neurons optimized themselves for finding simple things like edges and corners, with the layers after that looking for successively more complex features like basic shapes and, eventually, dogs or people.

LeCun recalls seeing the community that had mostly ignored neural networks pack into the room where the winners presented a paper on their results. "You could see right there a lot of senior people in the community just flipped," he says. "They said, 'Okay, now we buy it. That's it, now—you won.'"

Academics working on computer vision quickly abandoned their old methods, and deep learning suddenly became one of the main strands in artificial intelligence. Google bought a company founded by Hinton and the two others behind the 2012 result, and Hinton started working there part time on

a research team known as Google Brain. Microsoft and other companies created new projects to investigate deep learning. In December 2013, Facebook CEO Mark Zuckerberg stunned academics by showing up at the largest neural-network research conference, hosting a party where he announced that LeCun was starting FAIR (though he still works at NYU one day a week).

LeCun still harbors mixed feelings about the 2012 research that brought the world around to his point of view. "To some extent this should have come out of my lab," he says. Hinton shares that assessment. "It was a bit unfortunate for Yann that he wasn't the one who actually made the breakthrough system," he says. LeCun's group had done more work than anyone else to prove out the techniques used to win the ImageNet challenge. The victory could have been his had student gradua-

2006: Hinton's research group at the University of Toronto develops ways to train much larger networks with tens of layers of artificial neurons.

June 2012: Google uses deep learning to cut the error rate of its speech recognition software by 25 percent.

October 2012: Hinton and two colleagues from the University of Toronto win the largest challenge for software that recognizes objects in photos, almost halving the previous error rate.

March 2013: Google buys DNN Research, the company founded by the Toronto team to develop their ideas. Hinton starts working at Google.

March 2014: Facebook starts using deep learning to power its facial recognition feature, which identifies people in uploaded photos.

May 2015: Google Photos launches. The service uses deep learning to group photos of the same people and let you search your snapshots using terms like "beach" or "dog."

tion schedules and other commitments not prevented his own group from taking on ImageNet, he says. LeCun's hunt for deep learning's next breakthrough is now a chance to even the score.

Language learning

Facebook's New York office is a three-minute stroll up Broadway from LeCun's office at NYU, on two floors of a building constructed as a department store in the early 20th century. Workers are packed more densely into the open plan than they are at Facebook's headquarters in Menlo Park, California, but they can still be seen gliding on articulated skateboards past notices for weekly beer pong. Almost half of LeCun's team of leading AI researchers works here, with the rest at Facebook's California campus or an office in Paris. Many of them are trying to make neural networks better at understanding language. "I've hired all the people working on this that I could," says LeCun.

A neural network can "learn" words by spooling through text and calculating how each word it encounters could have been predicted from the words before or after it. By doing this, the software learns to represent every word as a vector that indicates its relationship to other words—a process that uncannily captures concepts in language. The difference between the vectors for "king" and "queen" is the same as for "husband" and "wife," for example. The vectors for "paper" and "cardboard" are close together, and those for "large" and "big" are even closer.

The same approach works for whole sentences (Hinton says it generates "thought vectors"), and Google is looking at using it to bolster its automatic translation service. A recent paper from researchers at a Chinese university and Microsoft's Beijing lab used a version of the vector technique to make software that beats some humans on IQ-test questions requiring an understanding of synonyms, antonyms, and analogies.

Several companies have opened deep-learning labs. "I've hired all the people working on this that I could," says LeCun.

LeCun's group is working on going further. "Language in itself is not that complicated," he says. "What's complicated is having a deep understanding of language and the world that gives you common sense. That's what we're really interested in building into machines." LeCun means common sense as Aristotle used the term: the ability to understand basic physical reality. He wants a computer to grasp that the sentence "Yann picked up the bottle and walked out of the room" means the bottle left with him. Facebook's researchers have invented a deep-learning system called a memory network that displays what may be the early stirrings of common sense.

A memory network is a neural network with a memory bank bolted on to store facts it has learned so they don't get washed away every time it takes in fresh data. The Facebook AI lab has created versions that can answer simple common-sense questions about text they have never seen before. For example, when researchers gave a memory network a very simplified summary of the plot of *Lord of the Rings*, it could answer questions such as "Where is the ring?" and "Where was Frodo before Mount Doom?" It could interpret the simple world described in the text despite having never previously encountered many of the names or objects, such as "Frodo" or "ring."

The software learned its rudimentary common sense by being shown how to answer questions about a simple text in which characters do things in a series of rooms, such as "Fred moved to the bedroom and Joe went to the kitchen." But LeCun wants to expose the software to texts that are far better at capturing the complexity of life and the things a virtual assistant might need to do. A virtual concierge called Moneypenny that Facebook is expected to release could be one source of that data. The assistant is said to be powered by a team of human operators who will help people do things like make restaurant reservations. LeCun's team could have a memory network watch over Moneypenny's shoulder before eventually letting it learn by interacting with humans for itself.

Building something that can hold even a basic, narrowly focused conversation still requires significant work. For example, neural networks have shown only very simple reasoning, and researchers haven't figured out how they might be taught to make plans, says LeCun. But results from the work that has been done with the technology so far leave him confident about where things are going. "The revolution is on the way," he says.

Some people are less sure. Deep-learning software so far has displayed only the simplest capabilities required for what we would recognize as conversation, says Oren Etzioni, CEO of the Allen Institute for Artificial Intelligence in Seattle. The logic and planning capabilities still needed, he says, are very

It's not clear how much we'd benefit from smarter virtual assistants, but we may not have to wait long to find out.

different from the things neural networks have been doing best: digesting sequences of pixels or acoustic waveforms to decide which image category or word they represent. "The problems of understanding natural language are not reducible in the same way," he says.

Gary Marcus, a professor of psychology and neural science at NYU who has studied how humans learn language and recently started an artificial-intelligence company called Geometric Intelligence, thinks LeCun underestimates how hard it would be for existing software to pick up language and common sense. Training the software with large volumes of carefully annotated data is fine for getting it to sort images. But Marcus doubts it can acquire the trickier skills needed for language, where the meanings of words and complex sentences can flip depending on context. "People will look back on deep learning and say this is a really powerful technique—it's the first time that AI became practical," he says. "They'll also say those things required a lot of data, and there were domains where people just never had enough." Marcus thinks language may be one of those domains. For software to master conversation, it would need to learn more like a toddler who picks it up without explicit instruction, he suggests.

Deep belief

At Facebook's headquarters in California, the West Coast members of LeCun's team sit close to Mark Zuckerberg and Mike Schroepfer, the company's CTO. Facebook's leaders know that LeCun's group is still some way from building something you can talk to, but Schroepfer is already thinking about how to use it. The future Facebook he describes retrieves and coördinates information, like a butler you communicate with by typing or talking as you might with a human one.

"You can engage with a system that can really understand concepts and language at a much higher level," says Schroepfer.

He imagines being able to ask that you see a friend's baby snapshots but not his jokes, for example. "I think in the near term a version of that is very realizable," he says. As LeCun's systems achieve better reasoning and planning abilities, he expects the conversation to get less one-sided. Facebook might offer up information that it thinks you'd like and ask what you thought of it. "Eventually it is like this super-intelligent helper that's plugged in to all the information streams in the world," says Schroepfer.

The algorithms needed to power such interactions would also improve the systems Facebook uses to filter the posts and ads we see. And they could be vital to Facebook's ambitions to become much more than just a place to socialize. As Facebook begins to host articles and video on behalf of media and entertainment companies, for example, it will need better ways for people to manage information. Virtual assistants and other spinouts from LeCun's work could also help Facebook's more ambitious departures from its original business, such as the Oculus group working to make virtual reality into a mass-market technology.

None of this will happen if the recent impressive results meet the fate of previous big ideas in artificial intelligence. Blooms of excitement around neural networks have withered twice already. But while complaining that other companies or researchers are over-hyping their work is one of LeCun's favorite pastimes, he says there's enough circumstantial evidence to stand firm behind his own predictions that deep learning will deliver impressive payoffs. The technology is still providing more accuracy and power in every area of AI where it has been applied, he says. New ideas are needed about how to apply it to language processing, but the still-small field is expanding fast as companies and universities dedicate more people to it. "That will accelerate progress," says LeCun.

It's still not clear that deep learning can deliver anything like the information butler Facebook envisions. And even if it can, it's hard to say how much the world really would benefit from it. But we may not have to wait long to find out. LeCun guesses that virtual helpers with a mastery of language unprecedented for software will be available in just two to five years. He expects that anyone who doubts deep learning's ability to master language will be proved wrong even sooner. "There is the same phenomenon that we were observing just before 2012," he says. "Things are starting to work, but the people doing more classical techniques are not convinced. Within a year or two it will be the end." ✠

Tom Simonite is MIT Technology Review's *San Francisco bureau chief.*