

## Facebook AI Director Yann LeCun on His Quest to Unleash Deep Learning and Make Machines Smarter

By Lee Gomes

Posted 18 Feb 2015 | 14:45 GMT



Photo-illustration: Randi Klett

Deep Learning expert Yann LeCun leads Facebook's AI research lab.

Artificial intelligence has gone through some dismal periods, which those in the field gloomily refer to as “AI winters.” This is not one of those times; in fact, AI is so hot right now that tech giants like Google, Facebook, Apple, Baidu, and Microsoft are battling for the leading minds in the field. The current excitement about AI stems, in great part, from groundbreaking advances involving what are known as “[convolutional neural networks](http://en.wikipedia.org/wiki/Convolutional_neural_network) ([http://en.wikipedia.org/wiki/Convolutional\\_neural\\_network](http://en.wikipedia.org/wiki/Convolutional_neural_network)).” This machine learning technique promises dramatic improvements in things like computer vision, speech recognition, and natural language processing. You probably have heard of it by its more layperson-friendly name: “Deep Learning.”

Few people have been more closely associated with Deep Learning than [Yann LeCun](https://research.facebook.com/researchers/684639631606527/yann-lecun/) (<https://research.facebook.com/researchers/684639631606527/yann-lecun/>), 54. Working as a Bell

Labs researcher during the late 1980s, LeCun developed the convolutional network technique and showed how it could be used to significantly improve handwriting recognition; many of the checks written in the United States are now processed with his approach. Between the mid-1990s and the late 2000s, when neural networks had fallen out of favor, LeCun was one of a handful of scientists who persevered with them. He became a [professor \(http://www.cs.nyu.edu/~yann/\)](http://www.cs.nyu.edu/~yann/) at New York University in 2003, and has since spearheaded many other Deep Learning advances.

More recently, Deep Learning and its related fields grew to become one of the most active areas in computer research. Which is one reason that at the end of 2013, LeCun was appointed head of the newly-created [Artificial Intelligence Research Lab \(https://research.facebook.com/ai\)](https://research.facebook.com/ai) at Facebook, though he continues with his NYU duties.

LeCun was born in France, and retains from his native country a sense of the importance of the role of the “public intellectual.” He writes and speaks frequently in his technical areas, of course, but is also not afraid to opine outside his field, including about current events.

*IEEE Spectrum* contributor Lee Gomes spoke with LeCun at his Facebook office in New York City. The following has been edited and condensed for clarity.

Yann LeCun on...

[Explaining Deep Learning . . . in Eight Words](#)

[A Black Box With 500 Million Knobs](#)

[The Pursuit of Beautiful Ideas \(Some Hacking Required\)](#)

[Hype and Things That Look Like Science but Are Not](#)

[Unsupervised Learning: The Learning That Machines Need](#)

[Facebook Does Deep Learning](#)

[Can Deep Learning Give Machines Common Sense?](#)

[The Inevitable Singularity Questions](#)

[“Sometimes I Need to Build Things With My Hands”](#)

---

## Explaining Deep Learning . . . in Eight Words

***IEEE Spectrum:*** We read about Deep Learning in the news a lot these days. What’s your least favorite definition of the term that you see in these stories?

**Yann LeCun:** My least favorite description is, “It works just like the brain.” I don’t like people saying this because, while Deep Learning gets an inspiration from biology, it’s very, very far from what the brain actually does. And describing it like the brain gives a bit of the aura of magic to it, which is dangerous. It leads to hype; people claim things that are not true. AI has gone through a number of AI winters because people claimed things they couldn’t deliver.

***Spectrum:*** So if you were a reporter covering a Deep Learning announcement, and had just eight words to describe it, which is usually all a newspaper reporter might get, what would you say?

**LeCun:** I need to think about this. [Long pause.] I think it would be “machines that learn to represent the world.” That’s eight words. Perhaps another way to put it would be “end-to-end machine learning.” Wait, it’s only five words and I need to

kind of unpack this. [Pause.] It's the idea that every component, every stage in a learning machine can be trained.

**Spectrum:** Your editor is not going to like that.

**LeCun:** Yeah, the public wouldn't understand what I meant. Oh, okay. Here's another way. You could think of Deep Learning as the building of learning machines, say pattern recognition systems or whatever, by assembling lots of modules or elements that all train the same way. So there is a single principle to train everything. But again, that's a lot more than eight words.

**Spectrum:** What can a Deep Learning system do that other machine learning systems can't do?

**LeCun:** That may be a better question. Previous systems, which I guess we could call "shallow learning systems," were limited in the complexity of the functions they could compute. So if you want a shallow learning algorithm like a "linear classifier" to recognize images, you will need to feed it with a suitable "vector of features" extracted from the image. But designing a feature extractor "by hand" is very difficult and time consuming.

An alternative is to use a more flexible classifier, such as a "support vector machine" or a two-layer neural network fed directly with the pixels of the image. The problem is that it's not going to be able to recognize objects to any degree of accuracy, unless you make it so gigantically big that it becomes impractical.

**Spectrum:** It doesn't sound like a very easy explanation. And that's why reporters trying to describe Deep Learning end up saying...

**LeCun:** ...that it's like the brain.

[Back to top](#)


---

## A Black Box With 500 Million Knobs

**Spectrum:** Part of the problem is that machine learning is a surprisingly inaccessible area to people not working in the field. Plenty of educated lay people understand semi-technical computing topics, like, say, the [PageRank algorithm](http://en.wikipedia.org/wiki/PageRank) (<http://en.wikipedia.org/wiki/PageRank>) that Google uses. But I'd bet that only professionals know anything detailed about linear classifiers or vector machines. Is that because the field is inherently complicated?

**LeCun:** Actually, I think the basics of machine learning are quite simple to understand. I've explained this to high-school students and school teachers without putting too many of them to sleep.

A pattern recognition system is like a black box with a camera at one end, a green light and a red light on top, and a whole bunch of knobs on the front. The learning algorithm tries to adjust the knobs so that when, say, a dog is in front of the camera, the red light turns on, and when a car is put in front of the camera, the green light turns on. You show a dog to the machine. If the red light is bright, don't do anything. If it's dim, tweak the knobs so that the light gets brighter. If the green light turns on, tweak the knobs so that it gets dimmer. Then show a car, and tweak the knobs so that the red light get dimmer and the green light gets brighter. If you show many examples of the cars and dogs, and



**“Imagine a box with 500 million knobs, 1,000 light bulbs, and 10 million images to train it with. That’s what a typical Deep Learning**

you keep adjusting the knobs just a little bit each time, eventually the machine will get the right answer every time.

**system is.”**

The interesting thing is that it may also correctly classify cars and dogs it has never seen before. The trick is to figure out in which direction to tweak each knob and by how much without actually fiddling with them. This involves computing a “gradient,” which for each knob indicates how the light changes when the knob is tweaked.

Now, imagine a box with 500 million knobs, 1,000 light bulbs, and 10 million images to train it with. That’s what a typical Deep Learning system is.

**Spectrum:** I assume that you use the term “shallow learning” somewhat tongue-in-cheek; I doubt people who work with linear classifiers consider their work “shallow.” Doesn’t the expression “Deep Learning” have an element of PR to it, since it implies that what is “deep” is what is being learned, when in fact the “deep” part is just the number of steps in the system?

**LeCun:** Yes, it is a bit facetious, but it reflects something real: shallow learning systems have one or two layers, while deep learning systems typically have five to 20 layers. It is not the learning that is shallow or deep, but the architecture that is being trained.

[Back to top](#)

---

## The Pursuit of Beautiful Ideas (Some Hacking Required)

**Spectrum:** The standard Yann LeCun biography says that you were exploring new approaches to neural networks at a time when they had fallen out of favor. What made you ignore the conventional wisdom and keep at it?

**LeCun:** I have always been enamored of the idea of being able to train an entire system from end to end. You hit the system with essentially raw input, and because the system has multiple layers, each layer will eventually figure out how to transform the representations produced by the previous layer so that the last layer produces the answer. This idea—that you should integrate learning from end to end so that the machine learns good representations of the data—is what I have been obsessed with for over 30 years.

**Spectrum:** Is the work you do “hacking,” or is it science? Do you just try things until they work, or do you start with a theoretical insight?

**LeCun:** It’s very much an interplay between intuitive insights, theoretical modeling, practical implementations, empirical studies, and scientific analyses. The insight is creative thinking, the modeling is mathematics, the implementation is engineering and sheer hacking, the empirical study and the analysis are actual science. What I am most fond of are beautiful and simple theoretical ideas that can be translated into something that works.

I have very little patience for people who do theory about a particular thing simply because it’s easy—particularly if they dismiss other methods that actually work empirically, just because the theory is too difficult. There is a bit of that in the machine learning community. In fact, to some extent, the “Neural Net Winter” during the late 1990s and early 2000s was a consequence of that philosophy; that you had to have ironclad theory, and the empirical results didn’t count. It’s a very bad way to approach an engineering problem.

But there are dangers in the purely empirical approach too. For example, the speech recognition community has

traditionally been very empirical, in the sense that the only stuff that's being paid attention to is how well you are doing on certain benchmarks. And that stifles creativity, because to get to the level where if you want to beat other teams that have been at it for years, you need to go underground for four or five years, building your own infrastructure. That's very difficult and very risky, and so nobody does it. And so to some extent with the speech recognition community, the progress has been continuous but very incremental, at least until the emergence of Deep Learning in the last few years.

**Spectrum:** You seem to take pains to distance your work from neuroscience and biology. For example, you talk about “convolutional nets,” and not “convolutional neural nets.” And you talk about “units” in your algorithms, and not “neurons.”

**LeCun:** That's true. Some aspects of our models are inspired by neuroscience, but many components are not at all inspired by neuroscience, and instead come from theory, intuition, or empirical exploration. Our models do not aspire to be models of the brain, and we don't make claims of neural relevance. But at the same time, I'm not afraid to say that the architecture of convolutional nets is inspired by some basic knowledge of the visual cortex. There are people who indirectly get inspiration from neuroscience, but who will not admit it. I admit it. It's very helpful. But I'm very careful not to use words that could lead to hype. Because there is a huge amount of hype in this area. Which is very dangerous.

[Back to top](#)

---

## Hype and Things That Look Like Science but Are Not

**Spectrum:** Hype is bad, sure, but why do you say it's “dangerous”?

**LeCun:** It sets expectations for funding agencies, the public, potential customers, start-ups and investors, such that they believe that we are on the cusp of building systems that are as powerful as the brain, when in fact we are very far from that. This could easily lead to another “winter cycle.”

And then there is a little bit of “[cargo cult science \(http://en.wikipedia.org/wiki/Cargo\\_cult\\_science\)](http://en.wikipedia.org/wiki/Cargo_cult_science)” in this. This is a Richard Feynman expression. He talked about cargo cult science to describe things that look like science, but basically are not.

**Spectrum:** Give me some examples.

**LeCun:** In a cargo cult, you reproduce the appearance of the machine without understanding the principles behind the machine. You build radio stations out of straw. The cargo cult approach to aeronautics—for actually building airplanes—would be to copy birds very, very closely; feathers, flapping wings, and all the rest. And people did this back in the 19th century, but with very limited success.

The equivalent in AI is to try to copy every detail that we know of about how neurons and synapses work, and then turn on a gigantic simulation of a large neural network inside a supercomputer, and hope that AI will emerge. That's cargo cult AI. There are very serious people who get a huge amount of money who basically—and of course I'm sort of simplifying here—are pretty close to believing this.

**“What I am most fond of are beautiful and simple theoretical ideas that can be translated into something that works.”**



***Spectrum:*** Do you think the [IBM True North project \(http://spectrum.ieee.org/computing/hardware/how-ibm-got-brainlike-efficiency-from-the-truenorth-chip\)](http://spectrum.ieee.org/computing/hardware/how-ibm-got-brainlike-efficiency-from-the-truenorth-chip) is cargo cult science?

**LeCun:** That would be a little harsh! But I do believe that some of the claims by the IBM group have gone a bit too far and were easily misinterpreted. Some of their announcements look impressive on the surface, but aren't actually implementing anything useful. Before the [True North project \(http://www.research.ibm.com/articles/brain-chip.shtml\)](http://www.research.ibm.com/articles/brain-chip.shtml), the group used an IBM supercomputer to "simulate a rat-scale brain." But it was just a random network of neurons that did nothing useful except burn cycles.

The sad thing about the True North chip is that it could have been useful if it had not tried to stick too close to biology and not implemented "spiking integrate-and-fire neurons." Building a chip is very expensive. So in my opinion—and I used to be a chip designer—you should build a chip only when you're pretty damn sure it can do something useful. If you build a convolutional net chip—and it's pretty clear how to do it—it can go into a lot of devices right away. IBM built the wrong thing. They built something that we can't do anything useful with.

***Spectrum:*** Any other examples?

**LeCun:** I'm going to get a lot of heat for this, but basically a big chunk of the [Human Brain Project in Europe \(https://www.humanbrainproject.eu/\)](https://www.humanbrainproject.eu/) is based on the idea that

we should build chips that reproduce the functioning of neurons as closely as possible, and then use them to build a gigantic computer, and somehow when we turn it on with some learning rule, AI will emerge. I think it's nuts.

Now, what I just said is a caricature of the Human Brain Project, to be sure. And I don't want to include in my criticism everyone who is involved in the project. A lot of participants are involved simply because it's a very good source of funding that they can't afford to pass up.

[Back to top](#)

**“If you build a convolutional net chip . . . it can go into a lot of devices right away. IBM built the wrong thing. They built something that we can't do anything useful with.”**

---

## Unsupervised Learning: The Learning That Machines Need

***Spectrum:*** How much more about machine learning in general remains to be discovered?

**LeCun:** A lot. The type of learning that we use in actual Deep Learning systems is very restricted. What works in practice in Deep Learning is "supervised" learning. You show a picture to the system, and you tell it it's a car, and it adjusts its parameters to say "car" next time around. Then you show it a chair. Then a person. And after a few million examples, and after several days or weeks of computing time, depending on the size of the system, it figures it out.

Now, humans and animals don't learn this way. You're not told the name of every object you look at when you're a baby. And yet the notion of objects, the notion that the world is three-dimensional, the notion that when I put an object behind another one, the object is still there—you actually learn those. You're not born with these concepts; you learn them. We call that type of learning "unsupervised" learning.

A lot of us involved in the resurgence of Deep Learning in the mid-2000s, including [Geoff Hinton](#)

(<http://www.cs.toronto.edu/~hinton/>), Yoshua Bengio ([http://www.iro.umontreal.ca/~bengioy/yoshua\\_en/index.html](http://www.iro.umontreal.ca/~bengioy/yoshua_en/index.html)), and myself—the so-called “Deep Learning conspiracy”—as well as Andrew Ng (<http://cs.stanford.edu/people/ang/>), started with the idea of using unsupervised learning more than supervised learning. Unsupervised learning could help “pre-train” very deep networks. We had quite a bit of success with this, but in the end, what ended up actually working in practice was good old supervised learning, but combined with convolutional nets, which we had over 20 years ago.

But from a research point of view, what we’ve been interested in is how to do unsupervised learning properly. We now have unsupervised techniques that actually work. The problem is that you can beat them by just collecting more data, and then using supervised learning. This is why in industry, the applications of Deep Learning are currently all supervised. But it won’t be that way in the future.

The bottom line is that the brain is much better than our model at doing unsupervised learning. That means that our artificial learning systems are missing some very basic principles of biological learning.

[Back to top](#)

---

## Facebook Does Deep Learning

**Spectrum:** What are some of the reasons Facebook was interested in setting up an AI lab?

**LeCun:** Facebook’s motto is to connect people. Increasingly, that also means connecting people to the digital world. At the end of 2013, when Mark Zuckerberg decided to create Facebook AI Research, the organization I direct, Facebook was about to turn 10 years old. The company thought about what “connecting people” would entail 10 years in the future, and realized that AI would play a key role.

Facebook can potentially show each person on Facebook about 2,000 items per day: posts, pictures, videos, etc. But no one has time for this. Hence Facebook has to automatically select 100 to 150 items that users want to see—or need to see. Doing a good job at this requires understanding people, their tastes, interests, relationships, aspirations and even goals in life. It also requires understanding content: understanding what a post or a comment talks about, what an image or a video contains, etc. Only then can the most relevant content be selected and shown to the person. In a way, doing a perfect job at this is an “AI-complete” problem: it requires understanding people, emotions, culture, art. Much of our work at Facebook AI focuses on devising new theories, principles, methods, and systems to make machines understand images, video, speech, and language—and then to reason about them.

**“Much of our work at Facebook AI focuses on devising new theories, principles, methods, and systems to make machines understand images, video, speech, and language—and then to reason about them.”**

**Spectrum:** We were talking earlier about hype, and I have a hype complaint of my own. Facebook recently announced a face-verification algorithm called “DeepFace,” with results that were widely reported to involve (<http://money.cnn.com/2014/04/04/technology/innovation/facebook-facial-recognition/>) near-human accuracy in facial recognition. But weren’t those results only true with carefully curated data sets? Would the system have the same success looking at whatever pictures it happened to come across on the Internet?

**LeCun:** The system is more sensitive to image quality than humans would be, that's for sure. Humans can recognize faces in a lot of different configurations, with different facial hair and things like that, which computer systems are slightly more sensitive to. But those systems can recognize humans among very large collections of people, much larger collections than humans could handle.

**Spectrum:** So could DeepFace do a better job of looking through pictures on the Internet and seeing if, say, Obama is in the picture than I could?

**LeCun:** It will do it faster, that's for sure.

**Spectrum:** Would it be more accurate?

**LeCun:** Probably not. No. But it can potentially recognize people among hundreds of millions. That's more than I can recognize!

**Spectrum:** Would it have 97.25 percent accuracy, like it did in the study?

**LeCun:** It's hard to quote a number without actually having a data set to test it on. It completely depends on the nature of the data. With hundreds of millions of faces in the gallery, the accuracy is nowhere near 97.25 percent.

**Spectrum:** One of the problems here seems to be that computer researchers use certain phrases differently than lay people. So when researchers talk about "accuracy rates," they might be talking about what they get with curated data sets. Whereas lay people might think the computers are looking at the same sorts of random pictures that people look at every day. But the upshot is that claims made for computer systems usually need to be much more qualified than they typically are in news stories.

**LeCun:** Yes. We work with a number of well-known benchmarks, like Labeled Faces in the Wild (<http://vis-www.cs.umass.edu/lfw/>) that other groups use as well, so as to compare our methods with others. Naturally, we also have internal datasets.

**Spectrum:** So in general, how close to humans would a computer be at facial recognition, against real pictures like you find on the Internet?

**LeCun:** It would be pretty close.

**Spectrum:** Can you attach a number to that?

**LeCun:** No, I can't, because there are different scenarios.

**Spectrum:** How well will Deep Learning do in areas beyond image recognition, especially with issues associated with generalized intelligence, like natural language?

**LeCun:** A lot of what we are working on at Facebook is in this domain. How do we combine the advantages of Deep Learning, with its ability to represent the world through learning, with things like accumulating knowledge from a temporal signal, which happens with language, with being able to do reasoning, with being able to store knowledge in a different way than current Deep Learning systems store it. Currently with Deep Learning systems, it's like learning a motor skill. The way we train them is similar to the way you train yourself to ride a bike. You learn a skill, but there's not a huge amount of factual memory or knowledge involved.

But there are other types of things that you learn where you have to remember facts, where you have to remember things and store them. There's a lot of work at Facebook, at Google, and at various other places where we're trying to have a



neural net on one side, and then a separate module on the other side that is used as a memory. And that could be used for things like natural language understanding.

We are starting to see impressive results in natural language processing with Deep Learning augmented with a memory module. These systems are based on the idea of representing words and sentences with continuous vectors, transforming these vectors through layers of a deep architecture, and storing them in a kind of associative memory. This works very well for question-answering and for language translation. A particular model of this type called “Memory Network” was recently proposed by Facebook scientists [Jason Weston](https://research.facebook.com/researchers/773431699369432/jason-weston/) (<https://research.facebook.com/researchers/773431699369432/jason-weston/>), [Sumit Chopra](https://research.facebook.com/researchers/944923105534452/sumit-chopra/) (<https://research.facebook.com/researchers/944923105534452/sumit-chopra/>), and [Antoine Bordes](https://research.facebook.com/researchers/547539832014269/antoine-bordes/) (<https://research.facebook.com/researchers/547539832014269/antoine-bordes/>). A somewhat related idea called the “Neural Turing Machine (<http://arxiv.org/abs/1410.5401>)” was also proposed by scientists at Google/Deep Mind.

**Spectrum:** So you don’t think that Deep Learning will be the one tool that will unlock generalized intelligence?

**LeCun:** It will be part of the solution. And, at some level, the solution will look like a very large and complicated neural net. But it will be very different from what people have seen so far in the literature. You’re starting to see papers on what I am talking about. A lot of people are working on what’s called “recurrent neural nets.” These are networks where the output is fed back to the input, so you can have a chain of reasoning. You can use this to process sequential signals, like speech, audio, video, and language. There are preliminary results that are pretty good. The next frontier for Deep Learning is natural language understanding.

**Spectrum:** If all goes well, what can we expect machines to soon be able to do that they can’t do now?

**LeCun:** You might perhaps see better speech recognition systems. But they will be kind of hidden. Your “digital companion” will get better. You’ll see better question-answering and dialog systems, so you can converse with your computer; you can ask questions and it will give you answers that come from some knowledge base. You will see better machine translation. Oh, and you will see [self-driving cars](http://spectrum.ieee.org/blog/cars-that-think) (<http://spectrum.ieee.org/blog/cars-that-think>) and [smarter robots](http://spectrum.ieee.org/blog/automaton) (<http://spectrum.ieee.org/blog/automaton>). Self-driving cars will use convolutional nets.

[Back to top](#)

**“A lot of people are working on what’s called ‘recurrent neural nets.’ . . . You can use this to process speech, audio, video, and language. There are preliminary results that are pretty good. The next frontier for Deep Learning is natural language understanding.”**

---

## Can Deep Learning Give Machines Common Sense?

**Spectrum:** In preparing for this interview, I asked some people in computing what they’d like to ask you. [Oren Etzioni](http://oren-etzioni.people.allenai.org/) (<http://oren-etzioni.people.allenai.org/>), head of the [Allen Institute for Artificial Intelligence](http://www.allenai.org/), (<http://www.allenai.org/>) was specifically curious about [Winograd Schemas](http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/winograd-schemas-replace-turing-test-for-defining-humanlevel-artificial-intelligence) (<http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/winograd-schemas-replace-turing-test-for-defining-humanlevel-artificial-intelligence>), which involve not only natural language and common sense, but also even an understanding of how contemporary society works. What approaches

might a computer take with them?

**LeCun:** The question here is how to represent knowledge. In “traditional” AI, factual knowledge is entered manually, often in the form of a graph, that is, a set of symbols or entities and relationships. But we all know that AI systems need to be able to acquire knowledge automatically through learning. The question becomes, “How can machines learn to represent relational and factual knowledge?” Deep Learning is certainly part of the solution, but it’s not the whole answer. The problem with symbols is that a symbol is a meaningless string of bits. In Deep Learning systems, entities are represented by large vectors of numbers that are learned from data and represent their properties. Learning to reason comes down to learning functions that operate on these vectors. A number of Facebook researchers, such as [Jason Weston](https://research.facebook.com/researchers/773431699369432/jason-weston/) (<https://research.facebook.com/researchers/773431699369432/jason-weston/>), [Ronan Collobert](https://research.facebook.com/researchers/1476710302612726/ronan-collobert/) (<https://research.facebook.com/researchers/1476710302612726/ronan-collobert/>), [Antoine Bordes](https://research.facebook.com/researchers/547539832014269/antoine-bordes/) (<https://research.facebook.com/researchers/547539832014269/antoine-bordes/>), and [Tomas Mikolov](https://research.facebook.com/researchers/643234929129233/tomas-mikolov/) (<https://research.facebook.com/researchers/643234929129233/tomas-mikolov/>) have pioneered the use of vectors to represent words and language.

**Spectrum:** One of the classic problems in AI is giving machines common sense. What ideas does the Deep Learning community have about this?

**LeCun:** I think a form of common sense could be acquired through the use of predictive unsupervised learning. For example, I might get the machine to watch lots of videos where objects are being thrown or dropped. The way I would train it would be to show it a piece of video, and then ask it, “What will happen next? What will the scene look like a second from now?” By training the system to predict what the world is going to be like a second, a minute, an hour, or a day from now, you can train it to acquire good representations of the world. This will allow the machine to know about the constraints of the physical world, such as “Objects thrown in the air tend to fall down after a while,” or “A single object cannot be in two places at the same time,” or “An object is still present while it is occluded by another one.” Knowing the constraints of the world would enable a machine to “fill in the blanks” and predict the state of the world when being told a story containing a series of events. Jason Weston, Sumit Chopra, and Antoine Bordes are working on such systems here at Facebook using the “Memory Network” I mentioned previously.

**Spectrum:** When discussing human intelligence and consciousness, many scientists often say that we don’t even know what we don’t know. Do you think that’s also true of the effort to build artificial intelligence?

**LeCun:** It’s hard to tell. I’ve said before that working on AI is like driving in the fog. You see a road and you follow the road, but then suddenly you see a brick wall in front of you. That story has happened over and over again in AI; with the [Perceptrons](http://en.wikipedia.org/wiki/Perceptron) (<http://en.wikipedia.org/wiki/Perceptron>) in the ’50s and ’60s, then the syntactic-symbolic approach in the ’70s, and then the expert systems in the ’80s, and then neural nets in the early ’90s, and then graphical models, kernel machines, and things like that. Every time, there is some progress and some new understanding. But there are also limits that need to be overcome.

**“I’ve said before that working on AI is like driving in the fog. You see a road and you follow the road, but then suddenly you see a brick wall in front of you.”**

**Spectrum:** Here’s another question, this time from [Stuart](http://www.ieor.berkeley.edu/People/Faculty/dreyfus.htm) (<http://www.ieor.berkeley.edu/People/Faculty/dreyfus.htm>) and [Hubert Dreyfus](http://socrates.berkeley.edu/~hdreyfus/index.html) (<http://socrates.berkeley.edu/~hdreyfus/index.html>), brothers and well-known professors at the University of California, Berkeley: “What do you think of press reports that computers are now robust enough to be able to identify and attack targets on their own, and what do you

think about the morality of that?"

**LeCun:** I don't think moral questions should be left to scientists alone! There are ethical questions surrounding AI that must be discussed and debated. Eventually, we should establish ethical guidelines as to how AI can and cannot be used. This is not a new problem. Societies have had to deal with ethical questions attached to many powerful technologies, such as nuclear and chemical weapons, nuclear energy, biotechnology, genetic manipulation and cloning, information access. I personally don't think machines should be able to attack targets without a human making the decision. But again, moral questions such as these should be examined collectively through the democratic/political process.

**Spectrum:** You often make quite caustic comments about political topics. Do your Facebook handlers worry about that?

**LeCun:** There are a few things that will push my buttons. One is political decisions that are not based on reality and evidence. I will react any time some important decision is made that is not based on rational decision-making. Smart people can disagree on the best way to solve a problem, but when people disagree on facts that are well established, I think it is very dangerous. That's what I call people on. It just so happens that in this country, the people who are on side of irrational decisions and religious-based decisions are mostly on the right. But I also call out people on the left, such as those who think GMOs are all evil—only some GMOs are!—or who are against vaccinations or nuclear energy for irrational reasons. I'm a rationalist. I'm also an atheist and a humanist; I'm not afraid of saying that. My idea of morality is to maximize overall human happiness and minimize human suffering over the long term. These are personal opinions that do not engage my employer. I try to have a clear separation between my personal opinions—which I post on my personal Facebook timeline—and my professional writing, which I post on my public Facebook page.

[Back to top](#)

## The Inevitable Singularity Questions

**Spectrum:** You've already [expressed your disagreement \(http://www.kdnuggets.com/2014/02/exclusive-yann-lecun-deep-learning-facebook-ai-lab-part2.html\)](http://www.kdnuggets.com/2014/02/exclusive-yann-lecun-deep-learning-facebook-ai-lab-part2.html) with many of the ideas associated with the [Singularity movement \(http://spectrum.ieee.org/static/singularity\)](http://spectrum.ieee.org/static/singularity). I'm interested in your thoughts about its sociology. How do you account for its popularity in Silicon Valley?

**LeCun:** It's difficult to say. I'm kind of puzzled by that phenomenon. As [Neil Gershenfeld \(http://ng.cba.mit.edu/\)](http://ng.cba.mit.edu/) has noted, the first part of a [sigmoid \(http://en.wikipedia.org/wiki/Sigmoid\\_function\)](http://en.wikipedia.org/wiki/Sigmoid_function) looks a lot like an exponential. It's another way of saying that what currently looks like exponential progress is very likely to hit some limit—physical, economical, societal—then go through an inflection point, and then saturate. I'm an optimist, but I'm also a realist.

There are people that you'd expect to hype the Singularity, like [Ray Kurzweil \(http://en.wikipedia.org/wiki/Ray\\_Kurzweil\)](http://en.wikipedia.org/wiki/Ray_Kurzweil). He's a futurist. He likes to have this positivist view of the future. He [sells a lot of books \(http://www.amazon.com/Ray-Kurzweil/e/B001ILHHDS\)](http://www.amazon.com/Ray-Kurzweil/e/B001ILHHDS) this way. But he has not contributed anything to the science of AI, as far as I can tell. He's sold products based on technology, some of which were somewhat innovative, but nothing conceptually new. And certainly he has never written papers that taught the world anything on how to make progress in AI.

**“There are people that you'd expect to hype the Singularity, like Ray Kurzweil. He's a futurist. . . . But he has not contributed**

**Spectrum:** What do you think he is going to accomplish in his job at Google? (<http://www.popsoci.com/technology/article/2012-12/google-hires-ray-kurzweil-head-its-engineering-lab>)

**anything to the science of AI, as far as I can tell.”**

**LeCun:** Not much has come out so far.

**Spectrum:** I often notice when I talk to researchers about the Singularity that while privately they are extremely dismissive of it, in public, they're much more temperate in their remarks. Is that because so many powerful people in Silicon Valley believe it?

**LeCun:** AI researchers, down in the trenches, have to strike a delicate balance: be optimistic about what you can achieve, but don't oversell what you can do. Point out how difficult your job is, but don't make it sound hopeless. You need to be honest with your funders, sponsors, and employers, with your peers and colleagues, with the public, and with yourself. It is difficult when there is a lot of uncertainty about future progress, and when less honest or more self-deluded people make wild claims of future success. That's why we don't like hype: it is made by people who are either dishonest or self-deluded, and makes the life of serious and honest scientists considerably more difficult.

When you are in the kind of position as Larry Page and Sergey Brin and Elon Musk and Mark Zuckerberg, you have to prepare for where technology is going in the long run. And you have a huge amount of resources to make the future happen in a way that you think will be good. So inevitably you have to ask yourself those questions: what will technology be like 10, 20, 30 years from now. It leads you to think about questions like the progress of AI, the Singularity, and questions of ethics.

**Spectrum:** Right. But you yourself have a very clear notion of where computers are going to go, and I don't think you believe we will be downloading our consciousness into them in 30 years.

**LeCun:** Not anytime soon.

**Spectrum:** Or ever.

**LeCun:** No, you can't say never; technology is advancing very quickly, at an accelerating pace. But there are things that are worth worrying about today, and there are things that are so far out that we can write science fiction about it, but there's no reason to worry about it just now.

[Back to top](#)

---

## “Sometimes I Need to Build Things With My Hands”

**Spectrum:** Another question from a researcher. C++ creator Bjarne Stroustrup asks, “You used to have some really cool toys—many of them flying. Do you still have time for hobbies or has your work crowded out the fun?”

**LeCun:** There is so much fun I can have with the work. But sometimes I need to build things with my hands. This was transmitted to me by my father, an aeronautical engineer. My father and my brother are into building airplanes as well. So when I go on vacation in France, we geek out and build airplanes for three weeks.

**Spectrum:** What is the plane that is on your [Google+ page](https://plus.google.com/+YannLeCunPhD/posts?) (<https://plus.google.com/+YannLeCunPhD/posts?>)

**LeCun:** It's a [Leduc](http://en.wikipedia.org/wiki/Leduc_0.10) ([http://en.wikipedia.org/wiki/Leduc\\_0.10](http://en.wikipedia.org/wiki/Leduc_0.10)), and it's in the Musée de l'Air near Paris. I love that plane. It was the first airplane powered by a ramjet, which is a particular kind of jet engine capable of very high speed. The SR-71

Blackbird, perhaps the fastest plane in the world, uses hybrid ramjet-turbojets. The first Leduc was a prototype that was built in France before World War II, and had to be destroyed before the Germans invaded. Several planes were built after the war. It was a very innovative way of doing things; it was never practical, but it was cool. And it looks great. It's got this incredible shape, where everything is designed for speed, but at the expense of the convenience of the pilot. The noise from the ramjet must have been unbearable for the pilot.

**Spectrum:** You tell a funny story in [a Web post \(http://yann.lecun.com/ex/fun/\)](http://yann.lecun.com/ex/fun/) about running into [Murray Gell-Mann \(http://en.wikipedia.org/wiki/Murray\\_Gell-Mann\)](http://en.wikipedia.org/wiki/Murray_Gell-Mann) years ago, and having him correct you on the pronunciation of your last name. You seemed to be poking gentle fun at the idea of the distinguished-but-pompous senior scientist. Now that you're becoming quite distinguished yourself, do you worry about turning out the same way?

**LeCun:** I try not to pull rank. It's very important when you lead a lab like I do to let young people exercise their creativity. The creativity of old people is based on stuff they know, whereas the creativity of young people is based on stuff they don't know. Which allows for a little wider exploration. You don't want to stunt enthusiasm. Interacting with PhD students and young researchers is a very good remedy against hubris. I'm not pompous, I think, and Facebook is a very non-pompous company. So it's a good fit.

[Back to top](#)

## About the Author

[Lee Gomes \(mailto:lee@leegomes.com\)](mailto:lee@leegomes.com) lives in San Francisco, and has covered technology for more than two decades, including at the *Wall Street Journal*. He [previously interviewed machine learning expert Michael Jordan \(http://spectrum.ieee.org/robotics/artificial-intelligence/machinelearning-maestro-michael-jordan-on-the-delusions-of-big-data-and-other-huge-engineering-efforts\)](http://spectrum.ieee.org/robotics/artificial-intelligence/machinelearning-maestro-michael-jordan-on-the-delusions-of-big-data-and-other-huge-engineering-efforts) for *IEEE Spectrum*.