

Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts

Big-data boondoggles and brain-inspired chips are just two of the things we're really getting wrong

By Lee Gomes

Posted 20 Oct 2014 | 19:37 GMT



Photo-Illustration: Randi Klett

The overeager adoption of big data (<http://spectrum.ieee.org/tag/big+data>) is likely to result in catastrophes of analysis comparable to a national epidemic of collapsing bridges. Hardware designers creating chips based on the human brain (<http://spectrum.ieee.org/computing/hardware/how-ibm-got-brainlike-efficiency-from-the-truenorth-chip>) are engaged in a faith-based undertaking likely to prove a fool's errand. Despite recent claims to the contrary, we are no further along with computer vision than we were with physics when Isaac Newton sat under his apple tree.

Those may sound like the Luddite ravings of a crackpot who breached security at an IEEE conference. In fact, the opinions belong to IEEE Fellow Michael I. Jordan (<http://www.cs.berkeley.edu/~jordan/>), Pehong Chen Distinguished Professor at the University of California, Berkeley. Jordan is one of the

world's most respected authorities on machine learning and an astute observer of the field. His CV would require its own massive database, and his standing in the field is such that he was chosen to write the introduction to the 2013 National Research Council report "[Frontiers in Massive Data Analysis \(http://www.nap.edu/catalog.php?record_id=18374\)](http://www.nap.edu/catalog.php?record_id=18374)." San Francisco writer Lee Gomes interviewed him for *IEEE Spectrum* on 3 October 2014.

Michael Jordan on...

[Why We Should Stop Using Brain Metaphors When We Talk About Computing](#)

[Our Foggy Vision About Machine Vision](#)

[Why Big Data Could Be a Big Fail](#)

[What He'd Do With US \\$1 Billion](#)

[How Not to Talk About the Singularity](#)

[What He Cares About More Than Whether \$P = NP\$](#)

[What the Turing Test Really Means](#)

Why We Should Stop Using Brain Metaphors When We Talk About Computing

IEEE Spectrum: I infer from your writing that you believe there's a lot of misinformation out there about deep learning, big data, computer vision, and the like.

Michael Jordan: Well, on all academic topics there is a lot of misinformation. The media is trying to do its best to find topics that people are going to read about. Sometimes those go beyond where the achievements actually are. Specifically on the topic of deep learning, it's largely a rebranding of neural networks, which go back to the 1980s. They actually go back to the 1960s; it seems like every 20 years there is a new wave that involves them. In the current wave, the main success story is the convolutional neural network, but that idea was already present in the previous wave. And one of the problems with both the previous wave, that has unfortunately persisted in the current wave, is that people continue to infer that something involving neuroscience is behind it, and that deep learning is taking advantage of an understanding of how the brain processes information, learns, makes decisions, or copes with large amounts of data. And that is just patently false.

Spectrum: As a member of the media, I take exception to what you just said, because it's very often the case that academics are desperate for people to write stories about them.

Michael Jordan: Yes, it's a partnership.

Spectrum: It's always been my impression that when people in computer science describe how the brain works, they are making horribly reductionist statements that you would never hear from neuroscientists. You called these "cartoon models" of the brain.

Michael Jordan: I wouldn't want to put labels on people and say that all computer scientists work one way, or all neuroscientists work another way. But it's true that with neuroscience, it's going to require decades or even hundreds of years to understand the deep principles. There is progress at the very lowest levels of neuroscience. But for issues of higher cognition—how we perceive, how we remember, how we act—we have no idea how neurons are storing information, how they are computing, what the rules are, what the algorithms are, what the representations are, and the like. So we are not yet in an era in which we can be using an understanding of the brain to guide us in the construction of intelligent

systems.

Spectrum: In addition to criticizing cartoon models of the brain, you actually go further and criticize the whole idea of “neural realism”—the belief that just because a particular hardware or software system shares some putative characteristic of the brain, it’s going to be more intelligent. What do you think of computer scientists who say, for example, “My system is brainlike because it is massively parallel.”

Michael Jordan: Well, these are metaphors, which can be useful. Flows and pipelines are metaphors that come out of circuits of various kinds. I think in the early 1980s, computer science was dominated by sequential architectures, by the von Neumann paradigm of a stored program that was executed sequentially, and as a consequence, there was a need to try to break out of that. And so people looked for metaphors of the highly parallel brain. And that was a useful thing.

But as the topic evolved, it was not neural realism that led to most of the progress. The algorithm that has proved the most successful for deep learning is based on a technique called back propagation. You have these layers of processing units, and you get an output from the end of the layers, and you propagate a signal backwards through the layers to change all the parameters. It’s pretty clear the brain doesn’t do something like that. This was definitely a step away from neural realism, but it led to significant progress. But people tend to lump that particular success story together with all the other attempts to build brainlike systems that haven’t been nearly as successful.

Spectrum: Another point you’ve made regarding the failure of neural realism is that there is nothing very neural about neural networks.

Michael Jordan: There are no spikes in deep-learning systems. There are no dendrites. And they have bidirectional signals that the brain doesn’t have.

We don’t know how neurons learn. Is it actually just a small change in the synaptic weight that’s responsible for learning? That’s what these artificial neural networks are doing. In the brain, we have precious little idea how learning is actually taking place.

Spectrum: I read all the time about engineers describing their new chip designs in what seems to me to be an incredible abuse of language. They talk about the “neurons” or the “synapses” on their chips. But that can’t possibly be the case; a neuron is a living, breathing cell of unbelievable complexity. Aren’t engineers appropriating the language of biology to describe structures that have nothing remotely close to the complexity of biological systems?

Michael Jordan: Well, I want to be a little careful here. I think it’s important to distinguish two areas where the word *neural* is currently being used.

One of them is in deep learning. And there, each “neuron” is really a cartoon. It’s a linear-weighted sum that’s passed through a nonlinearity. Anyone in electrical engineering would recognize those kinds of nonlinear systems. Calling that a neuron is clearly, at best, a shorthand. It’s really a cartoon. There is a procedure called logistic regression in statistics that dates from the 1950s, which had nothing to do with neurons but which is exactly the same little piece of architecture.

A second area involves what you were describing and is aiming to get closer to a simulation of an actual brain, or at least to a simplified model of actual neural circuitry, if I understand correctly. But the problem I see is that the research is not coupled with any understanding of what algorithmically this system might do. It’s not coupled with a learning system that takes in data and solves problems, like in vision. It’s really just a piece of architecture with the hope that someday people will discover algorithms that are useful for it. And there’s no clear reason that hope should be borne out. It is based, I believe, on faith, that if you build something like the brain, that it will become clear what it can do.

Spectrum: If you could, would you declare a ban on using the biology of the brain as a model in computation?

Michael Jordan: No. You should get inspiration from wherever you can get it. As I alluded to before, back in the 1980s, it was actually helpful to say, “Let’s move out of the sequential, von Neumann paradigm and think more about highly parallel systems.” But in this current era, where it’s clear that the detailed processing the brain is doing is not informing algorithmic process, I think it’s inappropriate to use the brain to make claims about what we’ve achieved. We don’t know how the brain processes visual information.

[Back to top](#)

Our Foggy Vision About Machine Vision

Spectrum: You’ve used the word *hype* in talking about vision system research. Lately there seems to be an epidemic of stories about how computers have tackled the vision problem, and that computers have become just as good as people at vision. Do you think that’s even close to being true?

Michael Jordan: Well, humans are able to deal with cluttered scenes. They are able to deal with huge numbers of categories. They can deal with inferences about the scene: “What if I sit down on that?” “What if I put something on top of something?” These are far beyond the capability of today’s machines. Deep learning is good at certain kinds of image classification. “What object is in this scene?”

But the computational vision problem is vast. It’s like saying when that apple fell out of the tree, we understood all of physics. Yeah, we understood something more about forces and acceleration. That was important. In vision, we now have a tool that solves a certain class of problems. But to say it solves all problems is foolish.

Spectrum: How big of a class of problems in vision are we able to solve now, compared with the totality of what humans can do?

Michael Jordan: With [face recognition \(http://spectrum.ieee.org/tag/face+recognition\)](http://spectrum.ieee.org/tag/face+recognition), it’s been clear for a while now that it can be solved. Beyond faces, you can also talk about other categories of objects: “There’s a cup in the scene.” “There’s a dog in the scene.” But it’s still a hard problem to talk about many kinds of different objects in the same scene and how they relate to each other, or how a person or a robot would interact with that scene. There are many, many hard problems that are far from solved.

Spectrum: Even in facial recognition, my impression is that it still only works if you’ve got pretty clean images to begin with.

Michael Jordan: Again, it’s an engineering problem to make it better. As you will see over time, it will get better. But this business about “revolutionary” is overwrought.

[Back to top](#)

Why Big Data Could Be a Big Fail

Spectrum: If we could turn now to the subject of [big data \(http://spectrum.ieee.org/tag/big+data\)](http://spectrum.ieee.org/tag/big+data), a theme that runs through your remarks is that there is a certain fool’s gold element to our current obsession with it. For example, you’ve predicted that society is about to experience an epidemic of false positives coming out of big-data projects.

Michael Jordan: When you have large amounts of data, your appetite for hypotheses tends to get even larger. And if it's growing faster than the statistical strength of the data, then many of your inferences are likely to be false. They are likely to be white noise.

Spectrum: How so?

Michael Jordan: In a classical database, you have maybe a few thousand people in them. You can think of those as the rows of the database. And the columns would be the features of those people: their age, height, weight, income, et cetera.

Now, the number of combinations of these columns grows exponentially with the number of columns. So if you have many, many columns—and we do in modern databases—you'll get up into millions and millions of attributes for each person.

Now, if I start allowing myself to look at all of the combinations of these features—if you live in Beijing, and you ride bike to work, and you work in a certain job, and are a certain age—what's the probability you will have a certain disease or you will like my advertisement? Now I'm getting combinations of millions of attributes, and the number of such combinations is exponential; it gets to be the size of the number of atoms in the universe.

Those are the hypotheses that I'm willing to consider. And for any particular database, I will find some combination of columns that will predict perfectly any outcome, just by chance alone. If I just look at all the people who have a heart attack and compare them to all the people that don't have a heart attack, and I'm looking for combinations of the columns that predict heart attacks, I will find all kinds of spurious combinations of columns, because there are huge numbers of them.

So it's like having billions of monkeys typing. One of them will write Shakespeare.

Spectrum: Do you think this aspect of big data is currently underappreciated?

Michael Jordan: Definitely.

Spectrum: What are some of the things that people are promising for big data that you don't think they will be able to deliver?

Michael Jordan: I think data analysis can deliver inferences at certain levels of quality. But we have to be clear about *what* levels of quality. We have to have error bars around all our predictions. That is something that's missing in much of the current machine learning literature.

Spectrum: What will happen if people working with data don't heed your advice?

Michael Jordan: I like to use the analogy of building bridges. If I have no principles, and I build thousands of bridges without any actual science, lots of them will fall down, and great disasters will occur.

Similarly here, if people use data and inferences they can make with the data without any concern about error bars, about heterogeneity, about noisy data, about the sampling pattern, about all the kinds of things that you have to be serious about if you're an engineer and a statistician—then you will make lots of predictions, and there's a good chance that you will occasionally solve some real interesting problems. But you will occasionally have some disastrously bad decisions. And you won't know the difference a priori. You will just produce these outputs and hope for the best.

And so that's where we are currently. A lot of people are building things hoping that they work, and sometimes they will. And in some sense, there's nothing wrong with that; it's exploratory. But society as a whole can't tolerate that; we can't just hope that these things work. Eventually, we have to give real guarantees. Civil engineers eventually learned to build bridges that were guaranteed to stand up. So with big data, it will take decades, I suspect, to get a real engineering approach, so

that you can say with some assurance that you are giving out reasonable answers and are quantifying the likelihood of errors.

Spectrum: Do we currently have the tools to provide those error bars?

Michael Jordan: We are just getting this engineering science assembled. We have many ideas that come from hundreds of years of statistics and computer science. And we're working on putting them together, making them scalable. A lot of the ideas for controlling what are called familywise errors, where I have many hypotheses and want to know my error rate, have emerged over the last 30 years. But many of them haven't been studied computationally. It's hard mathematics and engineering to work all this out, and it will take time.

It's not a year or two. It will take decades to get right. We are still learning how to do big data well.

Spectrum: When you read about big data and health care, every third story seems to be about all the amazing clinical insights we'll get almost automatically, merely by collecting data from everyone, especially in the cloud.

Michael Jordan: You can't be completely a skeptic or completely an optimist about this. It is somewhere in the middle. But if you list all the hypotheses that come out of some analysis of data, some fraction of them will be useful. You just won't know which fraction. So if you just grab a few of them—say, if you eat oat bran you won't have stomach cancer or something, because the data seem to suggest that—there's some chance you will get lucky. The data will provide some support.

But unless you're actually doing the full-scale engineering statistical analysis to provide some error bars and quantify the errors, it's gambling. It's better than just gambling without data. That's pure roulette. This is kind of partial roulette.

Spectrum: What adverse consequences might await the big-data field if we remain on the trajectory you're describing?

Michael Jordan: The main one will be a "big-data winter." After a bubble, when people invested and a lot of companies overpromised without providing serious analysis, it will bust. And soon, in a two- to five-year span, people will say, "The whole big-data thing came and went. It died. It was wrong." I am predicting that. It's what happens in these cycles when there is too much hype, i.e., assertions not based on an understanding of what the real problems are or on an understanding that solving the problems will take decades, that we will make steady progress but that we haven't had a major leap in technical progress. And then there will be a period during which it will be very hard to get resources to do data analysis. The field will continue to go forward, because it's real, and it's needed. But the backlash will hurt a large number of important projects.

[Back to top](#)

What He'd Do With \$1 Billion

Spectrum: Considering the amount of money that is spent on it, the science behind serving up ads still seems incredibly primitive. I have a hobby of searching for information about silly Kickstarter projects, mostly to see how preposterous they are, and I end up getting served ads from the same companies for many months.

Michael Jordan: Well, again, it's a spectrum. It depends on how a system has been engineered and what domain we're talking about. In certain narrow domains, it can be very good, and in very broad domains, where the semantics are much murkier, it can be very poor. I personally find Amazon's [recommendation system \(http://spectrum.ieee.org/computing/software/deconstructing-recommender-systems\)](http://spectrum.ieee.org/computing/software/deconstructing-recommender-systems) for books and music to be very, very good. That's because they have

large amounts of data, and the domain is rather circumscribed. With domains like shirts or shoes, it's murkier semantically, and they have less data, and so it's much poorer.

There are still many problems, but the people who build these systems are hard at work on them. What we're getting into at this point is semantics and human preferences. If I buy a refrigerator, that doesn't show that I am interested in refrigerators in general. I've already bought my refrigerator, and I'm probably not likely to still be interested in them. Whereas if I buy a song by Taylor Swift, I'm more likely to buy more songs by her. That has to do with the specific semantics of singers and products and items. To get that right across the wide spectrum of human interests requires a large amount of data and a large amount of engineering.

Spectrum: You've said that if you had an unrestricted \$1 billion grant, you would work on natural language processing. What would you do that Google isn't doing with Google Translate?

Michael Jordan: I am sure that Google is doing everything I would do. But I don't think Google Translate, which involves machine translation, is the only language problem. Another example of a good language problem is question answering, like "What's the second-biggest city in California that is not near a river?" If I typed that sentence into Google currently, I'm not likely to get a useful response.

Spectrum: So are you saying that for a billion dollars, you could, at least as far as natural language is concerned, solve the problem of generalized knowledge and end up with the big enchilada of AI: machines that think like people?

Michael Jordan: So you'd want to carve off a smaller problem that is not about everything, but which nonetheless allows you to make progress. That's what we do in research. I might take a specific domain. In fact, we worked on question-answering in geography. That would allow me to focus on certain kinds of relationships and certain kinds of data, but not everything in the world.

Spectrum: So to make advances in question answering, will you need to constrain them to a specific domain?

Michael Jordan: It's an empirical question about how much progress you could make. It has to do with how much data is available in these domains. How much you could pay people to actually start to write down some of those things they knew about these domains. How many labels you have.

Spectrum: It seems disappointing that even with a billion dollars, we still might end up with a system that isn't generalized, but that only works in just one domain.

Michael Jordan: That's typically how each of these technologies has evolved. We talked about vision earlier. The earliest vision systems were face-recognition systems. That's domain bound. But that's where we started to see some early progress and had a sense that things might work. Similarly with speech, the earliest progress was on single detached words. And then slowly, it started to get to be where you could do whole sentences. It's always that kind of progression, from something circumscribed to something less and less so.

Spectrum: Why do we even need better question-answering? Doesn't Google work well enough as it is?

Michael Jordan: Google has a very strong natural language group working on exactly this, because they recognize that they are very poor at certain kinds of queries. For example, using the word *not*. Humans want to use the word *not*. For example, "Give me a city that is not near a river." In the current Google search engine, that's not treated very well.

[Back to top](#)

How Not to Talk About the Singularity

Spectrum: Turning now to some other topics, if you were talking to someone in Silicon Valley, and they said to you, “You know, Professor Jordan, I’m a really big believer in [the singularity \(http://spectrum.ieee.org/static/singularity\)](http://spectrum.ieee.org/static/singularity),” would your opinion of them go up or down?

Michael Jordan: I luckily never run into such people.

Spectrum: Oh, come on.

Michael Jordan: I really don’t. I live in an intellectual shell of engineers and mathematicians.

Spectrum: But if you did encounter someone like that, what would you do?

Michael Jordan: I would take off my academic hat, and I would just act like a human being thinking about what’s going to happen in a few decades, and I would be entertained just like when I read science fiction. It doesn’t inform anything I do academically.

Spectrum: Okay, but knowing what you do academically, what do you think about it?

Michael Jordan: My understanding is that it’s not an academic discipline. Rather, it’s partly philosophy about how society changes, how individuals change, and it’s partly literature, like science fiction, thinking through the consequences of a technology change. But they don’t produce algorithmic ideas as far as I can tell, because I don’t ever see them, that inform us about how to make technological progress.

[Back to top](#)

What He Cares About More Than Whether $P = NP$

Spectrum: Do you have a guess about whether $P = NP$? Do you care?

Michael Jordan: I tend to be not so worried about the difference between polynomial and exponential. I’m more interested in low-degree polynomial—linear time, linear space. P versus NP has to do with categorization of algorithms as being polynomial, which means they are tractable and exponential, which means they’re not.

I think most people would agree that probably P is not equal to NP . As a piece of mathematics, it’s very interesting to know. But it’s not a hard and sharp distinction. There are many exponential time algorithms that, partly because of the growth of modern computers, are still viable in certain circumscribed domains. And moreover, for the largest problems, polynomial is not enough. Polynomial just means that it grows at a certain superlinear rate, like quadric or cubic. But it really needs to grow linearly. So if you get five more data points, you need five more amounts of processing. Or even sublinearly, like logarithmic. As I get 100 new data points, it grows by two; if I get 1,000, it grows by three.

That’s the ideal. Those are the kinds of algorithms we have to focus on. And that is very far away from the P versus NP issue. It’s a very important and interesting intellectual question, but it doesn’t inform that much about what we work on.

Spectrum: Same question about [quantum computing \(http://spectrum.ieee.org/tag/quantum+computing\)](http://spectrum.ieee.org/tag/quantum+computing).

Michael Jordan: I am curious about all these things academically. It’s real. It’s interesting. It doesn’t really have an impact

on my area of research.

[Back to top](#)

What the Turing Test Really Means

Spectrum: Will a machine pass the [Turing test \(http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/winograd-schemas-replace-turing-test-for-defining-humanlevel-artificial-intelligence\)](http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/winograd-schemas-replace-turing-test-for-defining-humanlevel-artificial-intelligence) in your lifetime?

Michael Jordan: I think you will get a slow accumulation of capabilities, including in domains like speech and vision and natural language. There will probably not ever be a single moment in which we would want to say, “There is now a new intelligent entity in the universe.” I think that systems like Google already provide a certain level of artificial intelligence.

Spectrum: They are definitely useful, but they would never be confused with being a human being.

Michael Jordan: No, they wouldn’t be. I don’t think most of us think the Turing test is a very clear demarcation. Rather, we all know intelligence when we see it, and it emerges slowly in all the devices around us. It doesn’t have to be embodied in a single entity. I can just notice that the infrastructure around me got more intelligent. All of us are noticing that all of the time.

Spectrum: When you say “intelligent,” are you just using it as a synonym for “useful”?

Michael Jordan: Yes. What our generation finds surprising—that a computer recognizes our needs and wants and desires, in some ways—our children find less surprising, and our children’s children will find even less surprising. It will just be assumed that the environment around us is adaptive; it’s predictive; it’s robust. That will include the ability to interact with your environment in natural language. At some point, you’ll be surprised by being able to have a natural conversation with your environment. Right now we can sort of do that, within very limited domains. We can access our bank accounts, for example. They are very, very primitive. But as time goes on, we will see those things get more subtle, more robust, more broad. As some point, we’ll say, “Wow, that’s very different when I was a kid.” The Turing test has helped get the field started, but in the end, it will be sort of like Groundhog Day—a media event, but something that’s not really important.

[Back to top](#)

About the Author

[Lee Gomes \(mailto:lee@leegomes.com\)](mailto:lee@leegomes.com), a former *Wall Street Journal* reporter, has been covering Silicon Valley for more than two decades.

