Comment on Michael I. Jordan’s “Artificial Intelligence—The Revolution Hasn’t Happened Yet”

Andrew W. Lo

Published on: Jun 23, 2019

Updated on: Dec 10, 2019

DOI: 10.1162/99608f92.9f99661b

This piece is a commentary on the article: [“Artificial Intelligence—The Revolution Hasn't Happened Yet.”](#)

Michael I. Jordan has written a thought-provoking article on the paradox of progress in Artificial Intelligence (AI). We are surrounded by previously unthinkable advances in this field—facial recognition, natural language translation, voice recognition and generation, autonomous vehicles, and the casual mastery of complex games such as chess and Go—and new breakthroughs seem to emerge almost daily. On the other hand, few of these advances have felt like a true revolution, and when they have, such as the ability to exploit social networks for targeted advertising and behavioral modification, it has often felt like a step backward from a societal perspective. Jordan makes a strong case for the necessity of a new field of human-centered engineering for systems within the several fields now comprising AI research.

However, I would like to focus on the broader arc of research in these fields. There is a lot of utility in Jordan's reorganization of AI into machine learning and related statistical methods, intelligence augmentation, intelligent infrastructure, and human-imitative artificial intelligence technology. The seeming fragmentation that is currently happening in AI, in my opinion, is a natural phenomenon that occurs in any rapidly advancing field of study.

AI is accumulating results as quickly as Moore's Law will permit and quite naturally, and predictably, it is beginning to splinter under its own weight into different subdisciplines. This is a natural consequence of the finiteness of human cognitive capacity compared to the growth of scientific knowledge. Even with intelligence augmentation technologies such as the search engine, it is impossible to stay current about everything in one's areas of interest. Eventually, however, as these new subdisciplines mature and develop truths of their own, they will come together to form a more robust and compelling narrative of the original field.

Consider the field of biology. We can date the beginnings of a truly modern biology to a dramatic reorganization, Linnaeus's 18th-century classification of the world's species into a consistent taxonomy, one that is still used in modified form today. Biology was then primarily a descriptive field, its subdisciplines based on the form of life being studied; for example, ornithology, the study of birds, or botany, the study of plants. Linnaeus's scheme provided a universal framework for all forms of life.

A century after Linnaeus, Charles Darwin provided the field with an even stronger theoretical underpinning after his discovery of evolution by natural selection. His powerful idea provided biology with testable narratives for the mysteries documented by the descriptive biologists, such as the geographical distribution of species, the unusual forms of newly discovered species, and perhaps most importantly to ourselves as *Homo sapiens*, the origins of humanity itself. Darwin's idea was so powerful that evolutionary biology became a vigorous foundational discipline, a rarity in the sciences.

However, Watson and Crick's famous discovery of the double helix of DNA overturned the previous order. The obvious connection between the structure of DNA and the mechanism of heredity created a new discipline out of older pieces of genetics and biochemistry: molecular biology, based on the study of information-containing molecules in living systems. The cracking of the genetic code and the deciphering of the genome made biology a much more informationally intensive field than it had been before, an early example of the potential of Big Data.

The rise of molecular biology was also a battle over scientific turf. Just as data science in its machine-learning guise appears to have taken over much of AI's old territory, molecular biology seemed poised to overthrow evolutionary biology as the leading paradigm in the biological sciences. The great evolutionary biologist E.O. Wilson called it "the Molecular Wars" in his autobiography, *Naturalist*. Unlike many academic battles, the stakes were very high: the direction of the biological sciences for generations. Who would tell the story of life, the scientists who studied the change in the forms of life over time, or the scientists who could decipher the genes responsible for the change? For a time, the evolutionary biologists were deposed from their position, and the molecular biologists took over, deriding their rivals as mere 'stamp collectors' (Wilson, 1994, pp. 218–237).

But something unusual happened after this takeover. Molecular biology was very good at explaining the 'how' of living systems at the molecular level. Evolutionary biology, on the other hand, was still required to explain the 'why.' Eventually, the two branches experienced a rapprochement, the 'Big Data' of molecular biology creating more powerful evolutionary hypotheses, and the relationships of evolutionary biology informing the genetic sequencers. Two separate subdisciplines—each with its own methodology, perspective, and forms of evidence—converged, and started reinforcing each other, rather than competing with each other.

This process impressed Wilson, who was an early participant in the molecular wars. Wilson used the 19th-century word 'consilience' to describe this process of scientific unification across disciplines (Wilson, 1998). Over the decades, it has happened multiple times among the sciences, in various branches of physics, biology, and the social sciences. I believe something similar will occur in the field of AI. What appears to be a split is, in fact, a sowing of new crops in the field, to produce an even more bountiful harvest in the future.

If AI has undergone a metamorphosis in the public mind away from the imitation of human intelligence, that is a sign of its conceptual strength. Take the new discipline of data science, for example. While data science is not in the classical lineage of human-imitative AI research, it developed out of the scientific need to efficiently interpret very large datasets, in much the way that statistics originally developed out of a need to interpret merely large datasets in astronomy, biology, and the social sciences. Because its methods, such as machine learning, are typically algorithmic, this new discipline has kept a strong flavor of computer science. The resulting hybrid field has produced results

that have impressed even economists, and as a group, we pride ourselves on the rigor of our econometric tests and do not impress easily. Statistics has so far been the field of choice with which to analyze data, but under the mantle of AI, data science has emerged as a very worthy competitor.

This is not meant to minimize the advances in the other areas of AI, especially under people like Jordan. From the overall perspective of the increase of knowledge, a tug of war between the statisticians and the computer scientists is very healthy! At some point, though, a rapprochement like the one that took place between the different branches of biology will need to be reached—a consilience of data scientists, computer scientists, statisticians, probabilists, and even topologists such as Gunnar Carlsson at Stanford (Carlsson, 2009)—to understand deeply the myriad properties of data. We might think of this rapprochement as part of intelligence augmentation, or as a new subcategory of nonhuman AI (since many of its algorithms are beyond strictly human capacity), or simply as part of a unified and reinvigorated field of AI.

There is another aspect to Jordan's taxonomy that I would like to see expanded, under his category of human-imitative AI. There is an unspoken assumption that human-imitative AI will capture the behavior of humans on their best day—the best chess player using human reasoning skills, or the best translator between languages. However, as a financial economist, I work in a field where the standard hypothesis, the Efficient Markets Hypothesis, already assumes such perfection, i.e., that the market fully reflects all available information in determining the price of a financial asset. (Given the behavior of the market in recent years, this hypothesis has obviously been questioned.) In contrast, I would prefer to see the development of artificial intelligence that accurately models the imperfections of human behavior, warts and all.

In order to do so, we will need to think systematically about how to capture the mistakes of human behavior at the algorithmic level. This is still a form of human-imitative artificial intelligence, but it might not seem very intelligent to an observer! Call it 'artificial stupidity.' An artificially stupid system might be as complex as a more capable artificial intelligence, or even more so. After all, many professions have the expression, "It takes a very smart person to make so large of a mistake."

Artificial stupidity—or, less pejoratively, 'artificial humanity'—is already part of the classical inheritance of AI (although not under either rubric). For example, in the early 1960s, the psychologist and political scientist Robert Abelson simulated the directed reasoning of ideologues of the time, basing it on his model of the cognition of belief (Abelson & Carroll, 1965). In the early 1970s, the psychiatrist Kenneth Colby at the Stanford Artificial Intelligence Lab developed a program named PARRY, a simulation of an individual with paranoid schizophrenia (Colby, 1975). In fact, a comparison of these kinds of artificial intelligence with the artificial intelligences of today is particularly instructive in highlighting the challenges that Jordan outlines.

In the early days of AI, so-called 'expert systems' used incredibly complicated algorithms but virtually no data to analyze real-world situations. These programs were not simple chatterbots based on statistical relationships between pieces of text, as might develop from a data science approach. Instead, they were designed to imitate the hypothesized thought processes of a supremely rational individual, even a specific real-world expert in some cases. In practice, unexpected inputs would cause improbable outputs, a classic case of "garbage in, garbage out," but it was a valiant effort, especially for the time.

Now fast forward to the artificial intelligences of today, which are the exact opposite of classical expert systems: much simpler algorithms applied to massive datasets. When I add a book to my shopping cart on Amazon, the site engages a fiendishly effective recommender system (Smith & Linden, 2017), showing me five other books purchased by users who bought the book I just ordered; I ordered two of the five other books. Simple algorithm, massive database.

This is neither accidental nor trivial. I believe that the overwhelming success of modern artificial intelligence versus previous generations is the direct result of this new data-intensive and algorithmically simple approach, and it is because it more closely matches human intelligence. The tech-entrepreneur-turned-neuroscientist Jeff Hawkins (2004) got it right when he argued that the core of what is considered to be intelligence is encapsulated by his 'memory-prediction' model. We search our memory banks for patterns that are closest to the current situation in which we must act, and based on what those patterns tell us about "if X, then Y," we choose the most favorable action. This memory-prediction model happens to lie at the core of virtually all industrial machine-learning applications, whether they be recommender systems, chatbots, or autonomous vehicles. Before there was machine learning, there was human learning.

However, there is a key difference between humans and machines, and it is the size of our database. An example I use in class to illustrate this point is how humans are able to navigate social situations effortlessly by piecing together an impression of other people they might encounter at a cocktail party. Imagine making small talk with a number of people over the course of an evening's random conversations through which you learn a number of facts about their backgrounds, e.g., their gender, sexual preference, marital status, educational background, etc. Table 1 summarizes the kind of features likely to be gathered about each individual, the number of unique categories for each feature (e.g., two major genders, male and female, two major sexual orientations, gay and heterosexual, which yields four possible types), and the features of two specific individuals, José and Susan.

Feature	# Possibilities	José	Susan

Gender and sexual orientation	4	Gay Male	Hetero Female
Marital status	3	Single	Married
Race/ethnicity	4	Latino	White
Age group	4	Young	Middle-Age
Current home state	50	CA	TX
Religious affiliation	4	None	Christian
Political party	3	Democrat	Republican
Economic status	3	Middle Class	Upper Class
Education	3	M.B.A.	B.A.

Table 1. Illustration of human cognitive ability to render rapid judgment of individuals based on sparse data.

After verbally describing each of these individuals to my class (“Jose is a young single gay Latino male from California who’s middle class, no religious affiliation, has an M.B.A. and is a Democrat”; “Susan is a middle-aged married female from Texas who’s Christian, Republican, and has a B.A.”), I ask my students to make three judgments about them:

1. If you were launching a tech startup and needed to hire someone to help you get the company set up, who would you rather hire, José or Susan?
2. If you needed help organizing a charity event to raise money for breast cancer research, who would you ask first, José or Susan?
3. If you were an auditor in the Tax Fraud Division at the Internal Revenue Service and could only audit the tax returns of one of these two individuals, who would you audit, José or Susan?

My students answer each of these questions decisively and with no hesitation: the vast majority choose José for the startup, Susan for the fundraiser, and Susan for the IRS audit. The students are surprised and somewhat taken aback when my reaction of feigned amazement is not due to the accuracy of their judgments—which they all take for granted—but rather because of how judgmental they are! After all, without ever having met these two individuals, the audience had no hesitation in making judgments about who to hire, who to delegate organizational tasks to, and who to investigate for tax fraud.

Hawkins's memory-prediction model implies that, using our database of memories of various people in a multitude of contexts, we can extrapolate the performance of José and Susan in the roles we are considering them for and predict accordingly. Accordingly, José's youth and M.B.A. seem ideal for a tech startup, and Susan's gender and wealth seem ideal for a breast cancer fundraiser. However, a closer look at this example reveals the fact that the features listed in Table 1 yield a remarkable 1,036,800 unique categories of individuals, which is greater resolution than the number of pixels in a 600×800 photograph.

The problem, of course, is that unlike Amazon's recommender system—based on a gargantuan database of every online transaction Amazon has conducted since its founding in 1994—none of us humans have access to such vast amounts of data, not to mention the means for accessing and processing it. Returning to the example in Table 1, because none of us have met more than a billion people in our lifetimes, our database of memories is extremely sparse. Yet from an evolutionary perspective, our memory-prediction model is a finely honed adaptation that makes split-second judgments, irrespective of the sparseness of our database. Adages like "don't judge a book by its cover" are passed down from one generation to the next precisely because we do judge books by their covers all the time. Moreover, such snap judgments likely offered survival benefits to our ancestors, hence their presence in our current behavioral repertoire.

The current environment in which we operate differs significantly from that of the Neolithic Ice Age which shaped that repertoire. We now have the ability to manipulate, with increasing accuracy, the specific cells of the database of the 1,036,800 categories of people we've met. And because of the sparseness of that database—most of those cells will in fact remain empty for most of our lives—it does not take much effort to change our behavior by changing just a few judiciously chosen cells. For example, if we falsely report that married well-to-do Christian women from Texas tend to exploit donors at fundraising events to enrich themselves, we may be able to prevent Susan from being hired to help with our breast cancer event if we so choose. This is why 'fake news' is so dangerous. When coupled with an understanding of our cognitive biases and sparse memory banks, modern artificial intelligence is simply too powerful a set of tools to entrust to the likes of *Homo sapiens*.

The study of artificial humanity may be a useful halfway house in developing an understanding of the hazards of AI and the necessary safeguards that will allow us to live in harmony with these tools. In this age of Big Data, large datasets of individual behavioral information are increasingly available to researchers for analysis—and one of the largest sources is the world of finance. In a very practical sense, financial markets are a collection of innumerable behavioral psychology experiments running in parallel, each with an estimable risk and a calculable reward to the experimental subject. This gives researchers the opportunity to simultaneously examine the macroscopic and the microfoundational scales of behavior.

For example, consider how news propagates. A newsworthy event occurs, and word of it begins to spread by personal communication, social media, the broadcast news networks, and the press, and people make decisions adapted to this new information. In financial economics, this sort of 'event study' is used to examine the speed at which new information is incorporated into market prices, most famously in Maloney and Mulherin's (2003) article about the market effects of the Space Shuttle *Challenger* disaster. Participants in the stock market that day had a financial stake in getting information about the explosion as quickly and as accurately as possible. With that tragedy, however, the scope of the disaster was clear, and there was little mediation between the event and the viewer to dilute its import. Other news may be much less reliable, incorrectly evaluated, or deliberately intended to deceive. The appropriate model of artificial stupidity, however, constructed through collaboration between AI researchers, economists, and financial professionals, could simulate the spread of misinformation across the social networks and into the marketplace, in a form of epidemiology to anticipate and prevent the effects of fraud, financial contagion, or even geopolitical panic.

Jordan concludes his manifesto (if that is not too strong of a word) with a call for AI to develop a human-centered form of engineering. This is the niche where I believe artificial humanity would find its most important use. When we think of the great engineering disasters—the Tacoma Narrows Bridge, the New Orleans levees, or the Space Shuttle *Challenger*—the errors that made them possible ultimately come down to human choice. Artificial humanity could not only be used to test artificially intelligent systems and intelligent infrastructure, but also for predictive error analysis in other forms of engineering and design.

To do this properly, however, we will need to understand not only the perfect optimizer—e.g., the best chess player—but also the newcomers, the people who make habitual errors in judgment, the people with imperfect information, and so on. We need that sort of artificial humanity, not least to deal with the worst consequences of those behaviors, but as part of the broader program of AI, which from its beginning has tried to create a mirror of the human mind *in silico*.

In short, we will need to understand, algorithmically, the frailties and foibles that make us human and imperfect, before we can achieve the kind of true AI revolution that Jordan seeks. As the father of AI, Marvin Minsky, apocryphally said, "No computer has ever been designed that is ever aware of what it's doing; but most of the time, we aren't either."

Discussion

Further commentary by:

- [Rodney Brooks](#) (MIT)
- [Emmanuel Candes, John Duchi, and Chiara Sabatti](#) (Stanford University)
- [Greg Crane](#) (Tufts University)
- [David Donoho](#) (Stanford University)
- [Maria Fasli](#) (UNESCO)
- [Barbara Grosz](#) (Harvard University)
- [Maja Mataric](#) (USC)
- [Brendan McCord](#) (Tulco Labs)
- [Max Welling](#) (University of Amsterdam)
- [Rebecca Willett](#) (University of Chicago)

Rejoinder by: [Michael I. Jordan](#) (UC Berkeley)

References

Abelson, R. P., & Carroll, J. D. (1965). Computer simulation of individual belief systems. *The American Behavioral Scientist*, 8(9), 24-30.

Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, 46(2), 255-308.

Colby, K. M. (1975). *Artificial paranoia: a computer simulation of paranoid processes*. New York: Pergamon Press.

Hawkins, J. (2004). *On Intelligence*. New York: Times Books.

Jordan, M. I. (2018). Artificial intelligence—the revolution hasn't happened yet. Retrieved from <https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7>.

Maloney, M. T., & Mulherin, J. H. (2003). The complexity of price discovery in an efficient market: the stock market reaction to the Challenger crash. *Journal of Corporate Finance*, 9(4), 453-479.

Smith, B., & Linden, G. (2017). Two decades of recommender systems at Amazon.com. *IEEE Internet Computing* 21(3), 12-18.

Wilson, E. O. (1994). *Naturalist*. Washington, D.C.: Island Press.

Wilson, E. O. (1998). *Consilience: The Unity of Knowledge*. New York: Knopf.

This article is © 2019 by Andrew W. Lo. The article is licensed under a Creative Commons Attribution (CC BY 4.0) International license (<https://creativecommons.org/licenses/by/4.0/legalcode>), except where otherwise indicated with respect to particular material included in the article. The article should be attributed to the author identified above.