

Harvard Data Science Review • 1.1

Dr. AI or: How I Learned to Stop Worrying and Love Economics

Michael I. Jordan

Published on: Jun 23, 2019

Updated on: Dec 11, 2019

DOI: [10.1162/99608f92.b9006d09](https://doi.org/10.1162/99608f92.b9006d09)

This is a rejoinder to the [invited commentaries](#) on the article, "[Artificial Intelligence—The Revolution Hasn't Happened Yet.](#)"

It is a pleasure to be participating in this wide-ranging discussion of issues of such importance to modern life. Given that much of the recent public discussion around AI has had the flavor of far-out science fiction, I hope that thinkers of all kinds, particularly those in the media, will find food for thought in the near-term, sober perspectives that appear here.

Parenthetically, I do not dismiss the importance of far-out science fiction in allowing humans to muse upon potential long-term consequences of current technology. There is, however, a need to return to the current century, where information technology is having a profound, exciting, and sometimes pathological effect on society, culture, and individual human lives. What do we want out of the technology of our era?

With my phrase "human-imitative AI," I had hoped to circumscribe some of the themes that have dominated public discussion, pulling them aside momentarily and allowing a focus on other kinds of intellectual, technological, and social issues. Many of the discussants seem to be in strong agreement with this framing. Two discussants were not—Brendan McCord and Andrew Lo—for which I am grateful, as it permits me to more clearly delineate where I think that the focus might usefully be directed.

McCord asks the following: "The question arises, what should be the researcher's guiding light—the brain, or something else? Practically speaking, it is exceedingly difficult to imagine what other intelligence there could be." I wish to argue that it is not so difficult.

Lest there be any doubt, I view the scientific study of the brain as one of the grandest challenges that science has ever undertaken, and the accompanying engineering discipline of 'human-imitative AI' as equally grand and worthy. The two enterprises offer the hope of one day understanding—in a concrete, mathematical way—what is this mysterious thing called 'thought.' But we are currently very far from such understanding, and the pursuit of such understanding is not our sole touchstone. Indeed, as I argued in my article, it is unrealistic and distracting to view the major goal for information technology in our era as being that of putting 'thought' into the computer, and expecting that 'thinking computers' will be able to solve our problems and make our lives better.

Let me indulge in a bit of my own science fiction. Let us suppose that there is a fledgling Martian computer science industry, and suppose that the Martians look down at Earth to get inspiration for making their current clunky computers more 'intelligent.' What do they see that is intelligent, and worth imitating, as they look down at Earth?

They will surely take note of human brains and minds, and perhaps also animal brains and minds, as intelligent and worth emulating. But they will also find it rather difficult to uncover the underlying principles or algorithms that give rise to that kind of intelligence — the ability to form abstractions, to give semantic interpretation to thoughts and percepts, and to reason. They will see that it arises from neurons, and that each neuron is an exceedingly complex structure — a cell with huge numbers of proteins, membranes, and ions interacting in complex ways to yield complex three-dimensional electrical and chemical activity. Moreover, they will likely see that these cells are connected in complex ways (via highly arborized dendritic trees; please type "dendritic tree and spines" into your favorite image browser to get some sense of a real neuron). A human brain contains on the order of a hundred billion neurons connected via these trees, and it is the network that gives rise to intelligence, not the individual neuron.

Daunted, the Martians may step away from considering the imitation of human brains as the principal path forward for Martian AI. Moreover, they may reassure themselves with the argument that humans evolved to do certain things well, and certain things poorly, and human intelligence may be not necessarily be well suited to solve Martian problems.

What else is intelligent on Earth? Perhaps the Martians will notice that in any given city on Earth, most every restaurant has at hand every ingredient it needs for every dish that it offers, day in and day out. They may also realize that, as in the case of neurons and brains, the essential ingredients underlying this capability are local decisions being made by small entities that each possess only a small sliver of the information being processed by the overall system. But, in contrast to brains, the underlying principles or algorithms may be seen to be not quite as mysterious as in the case of neuroscience. And they may also determine that this system is intelligent by any reasonable definition—it is adaptive (it works rain or shine), it is robust, it works at small scale and large scale, and it has been working for thousands of years (with no software updates needed). Moreover, not being anthropocentric creatures, the Martians may be happy to conceive of this system as an ‘entity’—just as much as a collection of neurons is an ‘entity.’

Am I arguing that we should simply bring in microeconomics in place of computer science? And praise markets as the way forward for AI? No, I am instead arguing that we should bring microeconomics in as a first-class citizen into the blend of computer science and statistics that is currently being called ‘AI.’ This blend was hinted at in my discussion piece; let me now elaborate.

Recommendation systems are a wonderful example of a blend of computer science and statistics. The basic idea is simple. If someone goes to a website and buys some books, and subsequently another customer buys many of the same books, then the website will recommend to the second customer some of the other books that the first customer bought. Real recommendation systems are far more complex than this—they use advanced computer science and advanced statistics to work at huge scale

and to handle the ‘long tail’ of unusual humans and unusual products. They are ‘prediction machines,’ and, as such, they are limited kinds of ‘intelligence.’

Indeed, classical recommendation systems can and do cause serious problems if they are rolled out in real-world domains where there is scarcity. Consider building an app that recommends routes to the airport. If few people in a city are using the app, then it is benign, and perhaps useful. When many people start to use the app, however, it will likely recommend the same route to large numbers of people and create congestion. The best way to mitigate such congestion is not to simply assign people to routes willy-nilly, but to take into account human preferences—on a given day some people may be in a hurry to get to the airport and others are not in such a hurry. An effective system would respect such preferences, letting those in a hurry opt to pay more for their faster route and allowing others to save for another day. But how can the app know the preferences of its users? It is here that major IT companies stumble, in my humble opinion. They assume that, as in the advertising domain, it is the computer's job to figure out human users' preferences, by gathering as much information as possible about their users, and by using AI. But this is absurd; in most real-world domains—where our preferences and decisions are fine-grained, contextual, and in-the-moment—there is no way that companies can collect enough data to know what we really want. Nor would we want them to collect such data—doing so would require getting uncomfortably close to prying into the private thoughts of individuals. A more appealing approach is to empower individuals by creating a two-way market where (say) street segments bid on drivers, and drivers can make in-the-moment decisions about how much of a hurry they are in, and how much they're willing to spend (in some currency) for a faster route.

Similarly, a restaurant recommendation system could send large numbers of people to the same restaurant. Again, fixing this should not be left to a platform or an omniscient AI system that purportedly knows everything about the users of the platform; rather, a two-way market should be created where the two sides of the market see each other via recommendation systems.

It is this last point that takes us beyond classical microeconomics and brings in machine learning. In the same way as modern recommendation systems allowed us to move beyond classical catalogs of goods, we need to use computer science and statistics to build new kinds of two-way markets. For example, we can bring relevant data about a diner's food preferences, budget, physical location, etc., to bear in deciding which entities on the other side of the market (the restaurants) are best to connect to, out of the tens of thousands of possibilities. That is, we need two-way markets where each side sees the other side via an appropriate form of recommendation system.

From this perspective, business models for modern information technology should be less about providing ‘AI avatars’ or ‘AI services’ for us to be dazzled by (and put out of work by)—on platforms that are monetized via advertising because they do not provide sufficient economic value directly to

the consumer—and more about providing new connections between (new kinds of) producers and consumers.

Consider the fact that precious few of us are directly connected to the humans who make the music we listen to (or listen to the music that we make), to the humans who write the text that we read (or read the text that we write), and to the humans who create the clothes that we wear. Making those connections in the context of a new engineering discipline that builds market mechanisms on top of data flows would create new ‘intelligent markets’ that currently do not exist. Such markets would create jobs and unleash creativity.

Implementing such platforms is a task worthy of a new branch of engineering. It would require serious attention to data flow and data analysis, it would require blending such analysis with ideas from market design and game theory, and it would require integrating all of the above with innovative thinking in the social, legal, and public policy spheres. The scale and scope is surely at least as grand as that envisaged when chemical engineering was emerging as a way to combine ideas from chemistry, fluid mechanics, and control theory at large scale.

Certainly market forces are not a panacea. But market forces are an important source of algorithmic ideas for constructing intelligent systems, and we ignore them at our peril. We are already seeing AI systems that create problems regarding fairness, congestion, and bias. We need to reconceptualize the problems in such a way that market mechanisms can be taken into account at the algorithmic level, as part and parcel of attempting to make the overall system be ‘intelligent.’ Ignoring market mechanisms in developing modern societal-scale information-technology systems is like trying to develop a field of civil engineering while ignoring gravity.

Markets need to be regulated, of course, and it takes time and experience to discover the appropriate regulatory mechanisms. But this is not a problem unique to markets. The same is true of gravity, when we construe it as a tool in civil engineering. Just as markets are imperfect, gravity is imperfect. It sometimes causes humans, bridges, and buildings to fall down. Thus it should be respected, understood, and tamed. We will require new kinds of markets, which will require research into new market designs and research into appropriate regulation. Again, the scope is vast.

In short, to those who argue that we should focus on emulating human intelligence because it's the only kind that we know, I have given you a second kind. I suspect that there are others. We need to liberate such forms of intelligence from an overly anthropocentric perspective. Indeed, while markets have historically been built on top of human decisions, it is clear that many of those decisions do not require very much of the human. Mimicking those simple decision-making processes, while focusing on the overall system, will liberate our algorithmic creativity. Market forces provide a kind of

algorithmic intelligence that is complementary to that sought by classical AI researchers focused on ‘thought.’

I won't go through my other disagreements with McCord, but I do wish to note that his assertion that I'm being contradictory regarding ‘revolution’ doesn't hold up. I certainly do believe that we're in an era in which there are revolutionary new forces at work—these include the ease of collecting data, computing as a utility, and effective data analysis methods, all at massive scale. But that's not the same thing as saying that we are in a revolution defined by *solutions* to the problems that these forces have brought to the fore. And it is certainly not the same thing as saying that we are in an ‘AI Revolution.’ If the point is still not coming through, I refer the reader to David Donoho's lucid deconstruction of these issues. See also Candès, Duchi, and Sabatti, who are content to call the underlying phenomenon and challenges, "Data Science."

Andrew Lo is a financial economist, and I had hoped that he would have picked up on my nod in the direction of microeconomics. He instead focuses on a theme that will be familiar to researchers in human-computer interaction (HCI)—that of modeling humans computationally, including their frailties and lapses, and aiming to design interfaces and algorithms that bring human thinking and decision-making up a notch. This is perfectly reasonable, but it is just intelligence augmentation, described differently (one might call it ‘stupidity mitigation’). He also offers a curious historical perspective in which AI was once a unified field that is currently splintering under its success, and hopefully will come back together. That's not the history that I know and have lived. Here's my history: work in machine learning, supported by allied work in statistics and optimization, began to yield impressive results, in particular on human-imitative tasks, and someone (my best bet is the public relations wing of some IT company in California) started to refer to this work as ‘AI’ because it sounded better. AI researchers in academia, including classical AI researchers who historically had engaged very little with machine learning, suddenly found their discipline back in the news. Some were delighted; some were uncomfortable.

Lo also has an unusual perspective on statistics and data science, saying, "Statistics has, for years, been the field of choice with which to analyze data, but while under the mantle of AI, data science has emerged as a very worthy competitor." I know of no data scientist who sees things this way. Data science is an umbrella that brings statisticians together with database and distributed systems researchers. It brings together inference, data, modeling, and scalability.

Where I do agree with Lo is in his nice turn of phrase "modern AI is simply too powerful a set of tools to entrust to the likes of Homo sapiens," including his argument that "the current environment in which we operate differs significantly from that of the Neolithic Ice Age." Indeed, human intelligence is susceptible to poor statistical performance when confronted with modern datasets that are unrelated to our evolutionary past. We are quite susceptible to ‘false positives’ and susceptible to

confusing correlation for causation. Modeling such susceptibility, while a reasonable scientific endeavor, does not seem like the best path forward if the goal is to avoid that susceptibility.

Barbara Grosz's commentary revisits the AI-versus-HCI wars of many years ago. I am not sure which of these two camps I belong to, particularly given that neither field was very much interested in my home fields of statistics, optimization, pattern recognition, or numerical methods—the fields that laid the ground for most of the recent progress. Grosz seems to have put me in the "AI camp" (although not the "classical AI camp"). Away with these labelings! If anything, by training, disposition, and a personal intellectual journey, I'm more interested in linking computation with the external world than I am in using it to explore a putative internal world. So, call me "HCI" if you must, but in a broad sense, and acknowledging that classical HCI researchers would probably not agree to have their brand affixed to me. If this is all confusing to you, that is part of my point.

I mentioned false positives; here is an example: researchers who work on the West Coast share Jordan's perspectives on AI and researchers on the East Coast do not. Silly? Well, it accounts for all of the data I've reviewed thus far, and much of the data I'm soon to present. It's accordingly not hard to imagine a current-generation AI system picking up on such a correlation, not to mention a modern journalist. Indeed, sadly, most current AI systems don't do a very good job with the error bars that are needed to prevent leaping to this conclusion when considering millions of such hypotheses, nor do current AI systems do a very good job with the pragmatics needed to rule out considering such a hypothesis in the first place.

Let me turn more briefly to some of the other discussants. I'm generally in agreement with them and I mainly wish to cheerlead with respect to their deepening of some of the perspectives that I raised in my article.

I have long been a fan of Maja Mataric's work on "socially assistive robotics," and I enjoyed her further elaboration of this perspective. In particular, she reminds us to take care in formulating the overall goals for any system that interacts with humans, because "human wellbeing and longevity, our health and wellness, fundamentally hinge on physical activity, social connectedness, and a sense of purpose."

I also enjoyed Greg Crane's comments, notably the thought-provoking association that he draws between the humanities and intelligence augmentation. In his words, "For the humanist, intelligence augmentation must now and forever be our goal. Machine learning only matters insofar as it makes us fundamentally more intelligent and deepens our understanding." I also appreciate Rebecca Willett's somewhat different take on machine learning, coming from an engineering point of view. She offers a compelling general definition of engineering that includes the human: "Engineering integrates fundamental scientific and mathematical disciplines, adding to and synergizing them, resulting in useful technology for human benefit." I also want to highlight Maria Fasli's comments on some of the

key remaining challenges in human-imitative AI (in part to help ensure that no one will misinterpret my own article as arguing against ongoing work on the open problems in human-imitative AI). She puts it very clearly: "But despite our flaws, we are pretty good decision makers, we deal with uncertainty pretty well, and we are able, astonishingly, to draw generalizations and apply learnings from one problem to a completely different domain."

I quite enjoyed the commentary from Emmanuel Candès, John Duchi, and Chiara Sabatti, and I was pleased to see their raising the bar with a quote from Heidegger. I see that quote and I raise them one from Wittgenstein: "The limits of my language means the limits of my world" (1922). As they note, imitation of humans means inheriting those limits, and in many applications those limits will not be reasonable. It was useful, for example, for technology to take us beyond the visible light spectrum. I also want to call attention to their discussion of fairness. They are careful to talk about equality of opportunity as a way to conceive of fairness, which is rather different from naive approaches that aim to provide equality of outcome, i.e., treating everyone the same. I might add that again the economic perspective seems essential; in economic language, we might ask that the system assays and respects everyone's utilities.

These issues—of how to define and operationalize fairness, using concepts such as opportunity and utility—have long been discussed in the social sciences, and they will be a particularly productive arena in which to bring social science thinking together with computational and inferential thinking.

As for Max Welling's comments, I agree that industry's involvement is an essential, and even revolutionary, aspect of the current landscape. But I am also exercised by industry irresponsibility in repeated claims that 'AI will solve the problem,' when the 'problem' is one that AI is not remotely prepared to solve. This is a feint, one that avoids rethinking the entire business model and rethinking whether the business model is based on a relatively easy, but ultimately unhealthy, way to use the current internet. Yes, I have *that* company in mind.

Finally, let me return to David Donoho's commentary, which I found particularly helpful. His phrase "recycled intelligence" is spot-on as a description of supervised learning, the workhorse of modern machine learning. That said, not all of the recent wave of AI hype is about supervised learning. In particular, AlphaGo exemplifies the kind of system ("reinforcement learning") that learns by trying out vast numbers of actions (using a simulator to try out these actions inside the computer) and storing away those actions that succeed. We might refer to this as 'trial-and-error-with-simulator intelligence.' It is interesting to ponder whether recycled intelligence and trial-and-error-with-simulator intelligence are solving McCarthy's AI problem. Inexplicably to me (and to Rod Brooks, who also raises this question, and provides some commentary on what's missing from those two kinds of intelligence), many of the current purveyors of AI technology seem to think that McCarthy's, and Turing's, goal is within reach. Unfortunately, I suspect that we will therefore soon be witness to another public

relations spectacle—after yet more data is collected and yet more trial-and-error searches have been performed—in which some of these purveyors will trumpet that they have solved the Turing test. To forestall this spectacle, can we perhaps agree to retire the Turing test? It was useful philosophically and historically, but it incentivizes fakery and it is a poor way to evaluate progress in a real-world engineering discipline.

Discussion

Read invited 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)
- [Andrew Lo](#) (MIT)
- [Maja Mataric](#) (USC)
- [Brendan McCord](#) (Tulco Labs)
- [Max Welling](#) (University of Amsterdam)
- [Rebecca Willett](#) (University of Chicago)

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

Acknowledgements

I wish to acknowledge helpful comments from Pieter Abbeel, Mingli Chen, Trevor Darrell, Jelena Diakonakolis, Ken Goldberg, Xiao-Li Meng, Michael Mühlebach, Robert Nishihara, and Chelsea Zhang.

References

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. New York: Harcourt, Brace & Co.

This article is © 2019 by Michael I. Jordan. 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.