

**Comments on Michael Jordan’s Essay  
“Artificial Intelligence: The revolution hasn’t happened yet”**

EMMANUEL CANDÈS, JOHN DUCHI, CHIARA SABATTI  
Stanford University  
March 2019

We praise Jordan for bringing much needed clarity about the current status of Artificial Intelligence (AI)—what it currently is and what it is not—as well as explaining the current challenges lying ahead and outlining what is missing and remains to be done. Jordan makes several claims supported by a list of talking points that we hope will reach a wide audience; ideally, that audience will include academic, university, and governmental leaders, at a time where significant resources are being allocated to AI for research and education.

**The importance of clarity**

Jordan makes the point of being precise about the history of the term AI, and distinguishes several activities taking place under the AI umbrella term.

Is it all right to use AI as a label for all of these different activities? Jordan seems to think it is not and we agree. To begin with, words are not simple aseptic names; they matter, and they convey meaning (as any branding expert knows). To quote Heidegger: “Man acts as though he were the shaper and master of language, while in fact language remains the master of man.” In this instance, we believe that mislabeling generates confusion, which has consequences for research and educational programming.

Mislabeling and lack of historical knowledge obscure the areas in which we must educate students. Jordan argues “that most of what is being called AI today, particularly in the public sphere, is what has been called Machine Learning (ML) for the past several decades.” This is a fair point. Now, what has made ML so successful? What are the disciplines supporting ML and providing a good basis to understand the challenges, open problems and limitations of the current techniques? A quick look at major machine learning textbooks reveals that they all begin with a treatment of what one might term basic statistical tools (linear models, generalized linear models, logistic regression) as well as a treatment of cross validation, overfitting, and related statistical concepts. We also find chapters on probability theory and probabilistic modeling. How about engineering disciplines? Clearly, progress in optimization, particularly in convex optimization, has fueled ML algorithms for the last two decades. When we think about setting up educational programs, clarity is recognizing that statistical, probabilistic, and algorithmic reasoning have been successful, and that it is crucial for us to train researchers in these disciplines to make further progress and understand the limits of current tools.

At the research level, different fields of research (e.g., optimization, control, statistics) use similar tools. These research communities, however, have distinct intellectual agendas and work on very different problems; by all being in “AI,” we obscure what progress is missing and what still remains to be solved, making it harder for institutions and society to choose how to invest wisely and effectively in research.

Mislabeling also hides the fact that a self-driving car requires more than just a good vision system. It will require roads and all kinds of additional infrastructure. Mislabeling hides the fact that,

even when we write that an “artificial intelligence” system recommends a diet [6], it is not AI that performs a study of gut microbiomes, measures their variety, evaluates insulin and sugar responses to different foods, nor even fits the model, which in this case, is a gradient-boosted decision tree [7]. This mislabeling also hides that machine learning should not be an end to itself: just getting people what they want faster (better ads, better search results, better movies, algorithms for more addictive “handles” in songs) does not make us better. What would make us better is a deep investment in real world problems, collaboration between methods scientists (ML researchers) and domain scientists, for instance, studying the persistent degradation of our oceans and recommending actions, or investigating susceptibility to and effective treatments for opioid addiction.

An important confusion Jordan addresses is the sense of over-achievement that the use of the term AI conveys. Bluntly, we do not have intelligent machines. We have many unsolved problems. We particularly applaud recognition that much progress is needed in terms of “inferring and representing causality.” This is an area where the ingredients that have made AI very successful—trillions of examples, immense compute power, and fairly narrow tasks—have limited applicability. To recognize whether a cat is on an image or not, the machine does not reason. Rather, it does (sophisticated) pattern matching. Pearl describes “the ability of imagining things that are not there” as distinctive characteristics of human reasoning, and he sees this counterfactual reasoning as the foundation of the ability of thinking causally; this is absent from the current predictive machine learning toolbox.

### **The role statistics can play**

In contrast, counterfactual reasoning and imagining what is not there (yet might be) are not foreign to statistics. Statistics has grappled for many years with the challenge of searching for causal relations: emphasizing (sometimes stifflingly) how these cannot be deduced by simple association, developing randomized trial frameworks, introducing the idea of “confounders.” Consider the Neyman-Rubin potential outcomes model, which effectively asks: what would have been my response, had I taken the treatment? Or the statistical approaches to estimate the unseen numbers of species, the “dark figure” of unrecorded victims of a certain crime. And more generally, the foundations of statistical inference build precisely out of the ability to imagine sample values you might obtain if you were to repeat an experiment or a data collection procedure. Recognizing how statistics incorporates this fundamental characteristic of human intelligence makes us think about its potential in accompanying the development of our data-laden society; we enumerate a few directions in which we think statistical reasoning is likely to be fruitful.

1. *Robustness*: As systems based on data interface more and more with the world, it is important that we build them to be robust. It is not sufficient to achieve reasonable performance on a hold-out dataset. We would like to retain predictive power when circumstances are subject to reasonable changes. (Think of high profile failures: in 2015, software engineer Jacky Alc  n   pointed out that the image recognition algorithms in Google Photos were classifying his black friends as “gorillas.”) Statistical reasoning and tools (for example, can we have “good enough” performance 99% of the time; can we be confident in our predictions; how confident are our predictions?) will be important.
2. *Validity of algorithmic inferences*: Algorithmic techniques to infer patterns and structure have had exceptional success recently in many areas of practical value. They can also be important,

even revolutionary, for science in many areas. Data as divergent as social media interactions or satellite and drone images may provide vital results through such algorithms.

However, the scientific validity of the results cannot be assumed. Conventional concepts such as random sampling of the intended population are rarely relevant. A deeper understanding of the data sources and the computations applied will be essential. Jordan’s anecdote on the probability of Down Syndrome is telling in this regard: a carefully designed system, taking into account statistical uncertainty—in this case, Jordan himself—identified a major flaw. Surely, we cannot expect Jordan to come along every time we have a doctor’s appointment.

3. *Fairness*: Beyond the scientific validity of inferences, the use of algorithmic results to recommend practical actions raises important questions of equitable treatment. While humans differ in a variety of ways, as a society we tend to believe that individuals should be treated as equals, have freedom of opportunity, “stand in relations of equality to others” [1]. As we aspire to create automated decision rules, we need to make sure they incorporate this principle; we have just begun to think about the challenges here. While an “algorithm” may be automatic, following prescribed rules, and will apply an identical recipe to everyone, this notion of consistent treatment is only as good as the data that one uses to train it. We strive for equal opportunity, not “as good as things have been.” There is a growing understanding that biased data collection yields biased results: when more data is available from a particular social group, algorithms are likely to do better for this group, which can in turn lead to a vicious cycle of minority group abandonment [2], yielding ever more bias. Here, researchers in machine learning have begun to develop properties algorithms should satisfy to guarantee “equitable treatment;” the statistical calculus of uncertainty, robustness, conditioning, population (and sub-population) quantities, and prediction errors have important roles to play.
4. *Privacy*: Numerous high-profile failures of privacy—Homer and colleagues’ de-identification of study participants from microarray data [3], the canceling of the second Netflix prize because data was linked across multiple domains [4, 5]—highlight the challenges of large-scale data analyses. As computing moves ever closer to peripheral devices (watches, phones, smart appliances), more privacy concerns arise. Indeed, a major challenge in large-scale health and genetics studies is sharing data securely and privately. Yet given the potential positive impacts access to such data would have—better understanding of biological bases for disease, better energy allocation, emergency monitoring—it behooves us to develop a methodology around privacy and concomitant statistical analyses. While a sophisticated literature of algorithmic techniques under privacy constraints is growing, we believe more carefully integrated statistical reasoning is likely to yield tremendous benefits.

We can summarize the points above with a slogan: cross-validation is not enough. It is critical to carefully quantify our decision-making algorithms, their fairness, their real-world consequences, and their confidence and robustness in predictions. These challenges should be a clarion call for statistical thinking.

### **It is not just an engineering program: further clarity is needed**

Jordan brings much clarity when he distinguishes human-imitative AI from other activities including ML, or when he explains why human-imitative AI has little to do with cybernetics, whose

“intellectual agenda has come to dominate in the current era.” After dismissing the idea of imitative AI as a guiding design principle—after all, we do not have feathered flapping airplanes—he suggests new disciplines of engineering around “Intelligence Augmentation” (IA) and “Intelligent Infrastructure” (II). (In passing, we personally appreciate the term Data Science as our ability to advance discovery, create new knowledge, and provide insights that suggest solutions to the world’s most pressing problems, as these will increasingly rely on our ability to learn from data.) Jordan names IA and II for what they are, helps us to recognize what is missing, and where progress needs to happen.

But of course, it is not just a matter of engineering. How AI, IA, II, and data sciences will develop and what our society will do with them depend on multiple aspects. Jordan’s piece touches at times on some of these larger questions; we selectively bring up a few here to emphasize the need both for these debates and greater clarity in these areas.

Jordan writes “humans are not very good at some kinds of reasoning.” Where do we go from here? What sorts of decisions should we outsource to algorithms? It seems important to qualify what we want computers to do and how we want to receive help to make decisions. The current AI framework compares our situation with that of many others and gives us an answer that seems best for “people like us.” Over time, this encourages us to be more like these other people, and erodes our individuality. There are domains where this might be appropriate; we do not care about a radiologist’s personal bend when interpreting an image, but desire the most accurate reading, as there is an underlying truth we seek. In other domains, this may not be the case. We have political opinions, but society cannot afford to have our personal beliefs be forever reinforced until different points of view are (to us) moral outrages. On the lighter side, there is no single food I should order tonight. However, if we let the machine make recommendations on the basis of a series of healthy eating parameters, religious restrictions, previous choices, cost considerations, and other “mood indicators,” we will be divided into a few disjoint groups eating monocultural food. We are malleable, gullible, and have a tendency to follow the crowd. The influence of the crowd via recommendation systems can be truly overpowering. Even if AI systems may allow us to avoid some mistakes, it is not clear that we want the machine to take over. Making choices is difficult and history is full of unfortunate attempts to abdicate to higher powers this defining human act. We need to cultivate this trait of ours, and keeping it exercised with simple tasks is generally a well proven strategy

Elsewhere, Jordan writes that we “must bring economic ideas such as incentive and prices into the realm of the statistical and computational infrastructure that link humans to each other and to valued goods.” Recently, the governor of California has stated that the state’s “consumers should also be able to share in the wealth that is created from their data.” We must have a debate about how individuals control the data they generate and who is entitled to monetize their value. A “free market” where each one is free to sell their own data is one of the options, but care must be taken, as markets often provide socially detrimental solutions when there are participants with very limited agency (as a single individual is likely to be here).

To make progress on these questions, we need the participation of many, and as statisticians and ML researchers, we have a limited perspective and are poorly equipped even to outline the challenges. Still, we wish to emphasize that the “engineers of AI, IA, II” must engage in these debates, just as geneticists participate in panels discussing the ethical implications of gene editing. We are uniquely

aware of the merit and limitations of these engineering feats, and we have the duty to make them transparent to all.

## References

- [1] E. Anderson. What is the point of equality? *Ethics*, 109(2):287–337, 1999.
- [2] T. Hashimoto, M. Srivastava, H. Namkoong, and P. Liang. Fairness without demographics in repeated loss minimization. In *Proceedings of the 35th International Conference on Machine Learning*, 2018.
- [3] N. Homer, S. Szlinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, J. V. Pearson, D. A. Stephan, S. F. Nelson, and D. W. Craig. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genetics*, 4(8):e1000167, 2008.
- [4] S. Lohr. Netflix cancels contest after concerns are raised about privacy. *The New York Times*, 2010. URL <https://www.nytimes.com/2010/03/13/technology/13netflix.html>.
- [5] R. Singel. Netflix cancels recommendation contest after privacy lawsuit. *WIRED*, March 12 2010. URL <https://www.wired.com/2010/03/netflix-cancels-contest/>.
- [6] E. Topol. The A.I. diet. *The New York Times*, March 2 2019. URL <https://www.nytimes.com/2019/03/02/opinion/sunday/diet-artificial-intelligence-diabetes.html>.
- [7] D. Zeevi, T. Korem, N. Zmora, D. Israeli, D. Rothschild, A. Weinberger, O. Ben-Yacov, D. Lador, T. Avnit-Sagi, M. Lotan-Pompan, J. Suez, J. A. Mahdi, E. Matot, G. Malka, N. Kosower, M. Rein, G. Zilberman-Schapira, L. Dohnalová, M. Pevsner-Fischer, R. Bikovsky, Z. Halpern, E. Elinav, and E. Segal. Personalized nutrition by prediction of glycemic responses. *Cell*, 163(5):1079–1094, 2015.