Comments on Michael Jordan's Essay "The AI Revolution Hasn't Happened Yet"

David Donoho

Published on: Jun 30, 2019

Updated on: Dec 10, 2019

DOI: 10.1162/99608f92.c698b3a7

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

The headline news is that, across a range of well-defined challenge tasks such as image classification and speech-to-text transcription, the past decade has witnessed machine learning systems transition from mediocre performance to what can be called human-level performance. This is a landmark series of events; we can only make such a transition once. History may one day identify this transition as a revolutionary moment, depending on how technology develops next.

The media imagines this transition and loves to portray it as a vindication of some new genius ideas. It is not.

Reaching human-level performance in speech and biometrics was a result of three convergent technology trends:

1. The rapid expansion of speech and image data libraries during the last decade, as humans uploaded more than 1 trillion images per year to iPhone and Android servers, while many billions of minutes of human speech passed through Skype and WhatsApp servers
2. The explosion of computational power, as GPUs and cloud computing powered overwhelming growth in the amount of computing that could be used for training machine learning systems. Indeed, the number of CPU cycles used by ImageNet winners for model construction has scaled by a factor of 300,000x since 2012
3. Ascendancy of the Common Task Framework (aka Prediction Challenges; aka Kaggle Contests; aka ImageNet-style contests; to read more see Liberman 2016 or Donoho 2018), under which: contestants share training data; submit models to be algorithmically scored on sequestered test data; and winners are chosen algorithmically

As Jordan implies, (1)-(3) are not new. In fact, they culminate trends from the 1980s, when DARPA began a series of biometric data compilation projects run by NIST, and associated prediction challenges. The DARPA/NIST efforts transformed natural language processing and biometrics – long before the recent decade's heavily hyped successes with ImageNet or deep learning.

Trend (3), the Common Task Framework, is the largely unacknowledged part of the convergence (1)-(3) behind today's successes. It depends heavily on the use of labelled data, in order for contestants to train their models and for referees to objectively score results. More specifically, the striking successes in language and vision depend on very large amounts of correctly labelled data produced by human experts.

In short, the success stories in language and vision processing came from scrupulously mimicking human intelligence. No new intelligence has been created; existing human intelligence is merely

recycled.

The last decade shows that humans can record their own actions when faced with certain tasks, which can be recycled to make new decisions that score as well as humans' (or maybe better, because the recycled decisions are immune to fatigue and impulse).

Jordan introduces the term IA – intelligence augmentation – to refer to such 'human-imitative AI.' Augmentation is a marketing term, not a logically correct, neutral description.

Recycled human intelligence does not deserve to be called augmented intelligence. It does not truly augment the range of capabilities that humans possess.

I consider myself a good speller, but I object to the ubiquity of spellcheckers in information technology. Spell checkers impose upon me the purported wisdom of crowds. They do not augment my intelligence; they often waste my time. Moreover, when I fail to catch their mistaken 'corrections,' embarrassing messages go out—mistakes I would never have made.

Spell checkers are examples of recycled intelligence a la machine learning. For me, life with a spell checker is like having a person unfamiliar with my research expertise read everything I write and impose revisions without asking, then stubbornly resist my repeated attempts to correct the damage they do. While spell checkers may one day improve, similar examples will repeat in the years to come as machine learning embeds itself ever more deeply in human experience.

Something revolutionary is happening, but Jordan tells us that machine learning—not AI—delivers these exciting results, and that these results are not based on new ideas. I ratify this, further crediting the convergence of long-standing technology trends, and then discuss what this convergence really gives us: intelligence recycling.

Laypersons will expect that artificial intelligence means true intelligence, yet produced by computers. So, what is true intelligence? Jordan implies that one marker of true intelligence is to not repeat the same old thing when circumstances have changed.

In his opening example, he points to a medical system recycling a decision rule (designed from low-resolution data) even after data characteristics had changed (because the sensor has transitioned to higher resolution). In this new setting, the old decision rule became seriously miscalibrated, leading to a dramatic increase in false alarms and unnecessary recommendations of risky procedures. By thinking about the meaning of the data, the origin of the purported expert decision rule, and its analysis, Jordan came to understand that the rule is best ignored. Good for him, and for his daughter!

Jordan's story makes me worry that our civilization, misled by the attribution of intelligence to systems that are actually recycling old judgements, will endlessly reuse irrelevant data to make decisions that

could be better made by thinking afresh.

In the Mike Judge movie *Idiocracy*, the hero time travels hundreds of years into the future to find a civilization which at first glance is like our own. He gradually discovers that he is far smarter than any human inhabitant of this future Earth. Those humans know how to push buttons and enjoy the outputs of black boxes, but they cannot cope with unforeseen circumstances. The hero solves a crisis threatening the planet using nothing more than today's average intelligence, by recommending simple practices that today's schoolchildren could recommend, but which humans of the future don't know about. Because they have become so dependent on their black boxes, these inhabitants of the future can no longer figure out a solution for themselves.

Over the last century, IQ scores of human subjects increased with each generation, a phenomenon psychologists labeled the Flynn effect. This general rise in intelligence coincides with the increasingly widespread use of hypothetical reasoning among the educated, i.e., variants of "if A were the case, we would observe X; but we don't observe X, so we reject A," or which ponder "if B were the case, how bad would it be to assume A?"

People are smarter when they correctly use hypothetical reasoning. They access truths beyond those which can be inferred by simple direct measurement. They explore counterfactual situations they can imagine but not inhabit.

Statistical hypothesis testing, the cornerstone of statistical inference, is a prime example of hypothetical reasoning. Hundreds of thousands of scientific papers are published every year using hypothesis testing across all of science and medicine. Tremendous advances in agricultural productivity and in human lifespan arose over the last century, partly through rigorous application of statistical hypothesis testing in crop and medical research. So-called AI has not yet delivered the human benefit of comparable size and broad benefit. Most of today's claims for AI are forward-looking: we hope they may come true. The benefits of statistical hypothesis testing already exist. Today, humans eat better and live longer than we did a century ago.

Formal statistical models offer another powerful mode of hypothetical reasoning with data. Statisticians consider a family of models that might have generated the data, even though they know that only one (or perhaps none) can possibly be correct. They derive procedures assuming one specific model is exactly true but then study properties of those procedures under the assumption that some other model is actually true. Again, such modelling has succeeded for centuries in the scientific literature and has grown dramatically over the years.

Statisticians have been doing hypothetical reasoning for data analysis and interpretation for centuries, including two fields of special relevance:

1. Causal inference, which allows us to interpret data arising from the result of complex observation mechanisms, for example, biased sampling. Causal inference uses counterfactual reasoning to determine if the data we see might have been different if a certain hypothesized causal factor were silent.
2. Robust inference, which helps us correctly interpret data that may have been contaminated – perhaps even by a malicious opponent. Theoretical robust statistics uses a hybrid of worst-case reasoning and statistical modeling to envision consequences of such hypothetical contamination and protect against them.

Such hypothetical reasoning tools genuinely allow us to go far beyond the surface appearance of data and, by so doing, augment our intelligence. In Michael Jordan's personal story, his family and their medical advisers faced a machine-generated treatment recommendation that was risky to his unborn daughter. A prominent statistician, Jordan possessed a thorough knowledge of modern statistical theories and could reason hypothetically from the data of the problem and from background knowledge of the information systems that produced the risky recommendation. From Jordan's cultural grounding in causal inference and robust inference, he knew that machine recommendations might be systemically in error, as the data might not mean what they seem to mean. From knowledge of hypothesis testing and formal statistical models, he anticipated how properties of machine recommendations might change after changes in pixel density. He suspected that the false alarm rate of the recommendation system was seriously miscalibrated. His wife's doctor supplied qualitative confirmation. He and his wife ignored the machine's advice.

Look at all the hypothetical reasoning Jordan deployed in this instance! He had almost no data to go on, and he dared stand against the machine which supposedly had lots of data. In the end, his family avoided a risky procedure.

I see these themes running through Jordan's story:

1. Machines give us recycled intelligence, not true intelligence;
2. Relying on such recycled intelligence is risky; it may give systematically wrong answers;
3. True intelligence requires lots of data-free hypothetical reasoning about suspected causes, backed up by empirical checks.

We saw themes (1) and (2) already in my earlier objection to spell checkers. Theme (3) is the new element Michael Jordan implies.

Jordan's personal story shows that for AI to deliver what we think of as intelligence, it must transcend mere recycling [Theme (3)]. In the direction of true intelligence, Michael Jordan and other statisticians lead the way.

Discussion

Further commentary by:

- [Rodney Brooks](#) (MIT)
- [Emmanuel Candes, John Duchi, and Chiara Sabatti](#) (Stanford University)
- [Greg Crane](#) (Tufts 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)

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

Acknowledgements

Thanks to Xiaoyan Han (Cornell) and Vardan Papyan (Stanford) for many valuable comments. Thanks to Xiao-Li Meng for organizing a discussion of Michael Jordan's paper.

References

Donoho, D. (2017). 50 Years of Data Science, Journal of Computational and Graphical Statistics, 26:4, 745-766, DOI: [10.1080/10618600.2017.1384734](https://doi.org/10.1080/10618600.2017.1384734)

Lieberman, M. (2010). Fred Jelinek and the dawn of statistical machine translation. Computational Linguistics 36(4):595-599.

This article is © 2019 by David Donoho. 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.