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

Rebecca Willett

Published on: Jun 23, 2019

Updated on: Dec 10, 2019

DOI: 10.1162/99608f92.98280d4a

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

Michael Jordan's article "Artificial Intelligence—The Revolution Hasn't Happened Yet" is a timely and thoughtful take on the current state of AI as understood in the popular media, academia, and industry.

As Jordan details, the meaning of the term AI has evolved over time. Its original, relatively narrow, use focused on logic and mimicking human intelligence; this perspective is epitomized by the Turing test for machine intelligence, in which a successful AI system's behavior would be indistinguishable from that of a human. This use, however, does not reflect the broad scope of the modern common usage of the term AI.

Modern AI refers to computer systems that intelligently process information. This definition includes classical human-imitative AI as well as signal processing, machine learning, statistics, algorithms, uncertainty quantification, information theory, distributed systems, operations research, and control theory. Popular modern usage of the term AI encompasses the notions of Intelligence Augmentation and Intelligent Infrastructure outlined in Jordan's article. All of these areas focus on the 'intelligent' use of data. Some parts of modern AI strive to imitate or are inspired by human intelligence, while others can take a very different form. For example, speech processing (e.g., converting an audio recording to a text transcript or dividing an audio recording into different tracks for different speakers) methods are generally less human-imitative than methods for natural language processing (e.g., extracting meaning and understanding from a text transcript), yet both are examples of modern AI challenges.

Jordan argues that public dialog on AI focuses too heavily on human-imitative AI. Compare Jordan's article with the recent article by Garrett Kenyon in *Scientific American*, "AI's Big Challenge." This article clearly articulates several limitations of modern AI and attributes these failures to AI's "jumping over the long, slow process of cognitive development and instead focusing on solving specific tasks" (Kenyon, 2019). Kenyon issues a call for a renewed focus on systems that mimic human reasoning. Jordan does not echo Kenyon's call to mimic biology, which he claims diverts our attention from the comprehensive scope of modern AI. While humans provide valuable inspiration for what is possible, human mimicry may be highly suboptimal. Mimicking biological systems is certainly suboptimal in other contexts: only by understanding the fundamental principles of aerodynamics were we able to build airplanes that fly faster while carrying more weight than any known biological system. Similarly, in modern AI we aspire to use data in ways that surpass human limitations and biases, not imitate them.

Attention has not been diverted from these goals because of the historical emphasis on human-imitative AI. I am convinced that both scholarly communities and society in general are well aware of

the broader challenges and opportunities of AI. There is widespread, intense interest in leveraging AI in many fields where we aim to complement or exceed human capabilities: automatic recognition of image content (particularly in medical imaging or remote sensing), identifying health care best practices, improving agricultural yields, developing new materials, understanding how the human brain encodes information, and more.

Despite this interest, modern AI systems still face a myriad of critical limitations. First, many state-of-the-art systems lack interpretability (Bang, 2018). For instance, a machine might be able to accurately predict who is most at risk for a certain disease but not be able to provide insight into risk factors or preventative measures. There is a growing need for systems that understand their own limitations, know what they don't know, and give users a better sense of their confidence in the outputs. Second, these algorithms can be unexpectedly fragile (Goodfellow, Shiers, and Szegedy, 2015). Google demonstrated a system that correctly identified animals in pictures. When they then added tiny, carefully-selected perturbations to the pixels — imperceptible to the human eye — they found that their system made grossly incorrect identifications, such as misidentifying a panda as a gibbon. Third, machines, unlike humans, are slow to adapt to new settings and can fail to transfer knowledge gleaned in one domain to another. Furthermore, training data can be skewed, resulting in unexpected unfairness (Gibbs, 2015). For instance, if an algorithm analyzed training images only of dogs in the grass and cats indoors, the algorithm might “learn” that grass is indicative of dogs and be unable to recognize an image of a cat in the grass. These problems with training data can be difficult to identify. If not handled correctly, they may lead to biased outcomes (Newitz, 2016). Finally, the above challenges are compounded when we need machines not only to recognize patterns in data but also to select a sequence of actions, as when a system plays a game (Somers, 2018) or interacts with humans or its environment (Gaming, 2017).

None of these challenges can be addressed in isolation, and, as Jordan describes, the next phase of modern AI requires an engineering mindset. Engineering integrates fundamental scientific and mathematical disciplines, adding to and synergizing them, resulting in useful technology for human benefit. Consider the construction of an airplane. Different communities have built thousands of airplanes over the years, but not all planes have had the same efficient use of fuel, robustness to aging components, stability during extreme weather, and established safety standards. Our past successes and failures, accompanied by studying fundamental principles of aerospace engineering, have led to a deep understanding of and best practices for airplane construction, use, and maintenance.

Similarly, we are continuously building new AI systems, but these systems do not yet have the same efficient use of data and computational resources, robustness to different domains, stability with respect to changes in the input data, or established safety and ethical standards. Safety challenges faced by self-driving cars (“Uber settles with family,” 2018), biased outcomes in web job searches

(Gibbs, 2015), economic risks of automated credit pricing (Hale, 2019), and the fragility of AI in healthcare (Ross and Swetlitz, 2018) all give compelling illustrations of the lack of these qualities in existing systems. Addressing these challenges is a key focus of modern AI research. Rapidly growing bodies of research study efficient (Sun, 2018), robust (RobustML, 2019), and fair (FATML, 2019) algorithms, “safe” approaches that allow us to learn from large datasets while protecting individuals’ privacy (Ji, 2014), and interpretable methods that improve interactions with human operators (“Interpretable ML,” 2019). Thus, in contrast to Jordan’s claim that modern AI “cannot yet be viewed as constituting an engineering discipline,” I argue that the principles of efficiency, robustness, stability, and safety are becoming foremost concerns rather than afterthoughts, making the developing field of modern AI an exemplar of engineering.

Thinking of AI as an engineering discipline is also helpful with respect to the AI nomenclature issue emphasized in Jordan’s article. Specifically, he argues that “the use of this single, ill-defined acronym [AI] prevents a clear understanding of the range of intellectual and commercial issues at play.” In contrast, using a single, sometimes ill-defined name has not prevented electrical engineering, chemical engineering, biomedical engineering, or other engineering disciplines from developing a clear understanding of relevant technical, commercial, and societal challenges. Rather, engineers benefit from this breadth of scope and develop practical, integrated systems to address these challenges. The recent surge in new sources of data escalates the need for practical, integrated systems that scalably extract information from data; address the challenges of interpretability, fragility, adaptability, bias, and interactivity; share information rapidly among distributed users and computers; perform causal reasoning; and ensure that the resulting decisions are ethical, safe, and reliable. Successful communities of engineers do not silo their work, but rather form partnerships across academia, industry, and the government to develop new, safe, useful technologies that advance society. Modern AI has the potential to be engineering at its best.

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)
- [Andrew Lo](#) (MIT)

- [Maja Mataric](#) (USC)
- [Brendan McCord](#) (Tulco Labs)
- [Max Welling](#) (University of Amsterdam)

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

Acknowledgments

Thank you to Robert Nowak, Stephen Wright, Avrim Blum, Gregory Ongie, and Laura Balzano for thoughtful comments and suggestions during the writing of this response.

References

Bang, Seojin. Introduction to Interpretable Machine Learning.

<https://medium.com/@Petuum/introduction-to-interpretable-machine-learning-3a62870f2f37>.

Updated November 20, 2018. Accessed May 22, 2019.

FATML, <https://www.fatml.org/>. Accessed May 28, 2019.

Gaming, NaEr. Boston Dynamics' Atlas Falls Over After Demo at the Congress of Future Scientists and Technologists. <https://www.youtube.com/watch?v=TxobtWAFh8o>. Updated July 2, 2017. Accessed May 22, 2019.

Gibbs, S. Women less likely to be shown ads for high-paid jobs on Google, study shows. The Guardian Website. <https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study>. Updated July 8, 2015. Accessed May 22, 2019.

Goodfellow, I. J., J. Shlens, and C. Szegedy. Explaining and Harnessing Adversarial Examples. arXiv. <https://arxiv.org/abs/1412.6572>. Updated March 20, 2015. Accessed May 22, 2019.

Hale, T. How big data really fits into lending.

<https://ftalphaville.ft.com/2019/03/13/1552488421000/How-big-data-really-fits-into-lending/>.

Updated March 13, 2019. Accessed May 22, 2019.

Interpretable ML. <http://interpretable.ml/>. Accessed May 28, 2019.

Ji, Z., Z. C. Lipton, C. Elkan. "Differential Privacy and Machine Learning: a Survey and Review." arXiv. <https://arxiv.org/abs/1412.7584>. Updated December 24, 2014. Accessed May 28, 2019.

Kenyon, Garrett. AI's Big Challenge: To make it truly intelligent, researchers need to rethink the way they approach the technology. <https://blogs.scientificamerican.com/observations/ais-big-challenge1/>. Updated February 26, 2019. Accessed May 22, 2019.

Molnar, Christoph. "Interpretable Machine Learning."

Newitz, A. Facebook fires human editors, algorithm immediately posts fake news. <https://arstechnica.com/information-technology/2016/08/facebook-fires-human-editors-algorithm-immediately-posts-fake-news/>. Updated August 29, 2016. Accessed May 22, 2019.

RobustML. <https://www.robust-ml.org/>. Accessed May 28, 2019.

Ross, C. and I. Swetlitz. IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show. <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/>. Updated July 25, 2018. Accessed May 22, 2019.

Somers, J. How the Artificial-Intelligence Program AlphaZero Mastered Its Games. <https://www.newyorker.com/science/elements/how-the-artificial-intelligence-program-alphazero-mastered-its-games>. Updated December 28, 2018. Accessed May 22, 2019.

Sun, Yiting. More efficient machine learning could upend the AI paradigm. <https://www.technologyreview.com/s/610095/more-efficient-machine-learning-could-upend-the-ai-paradigm/>. Updated February 2, 2018. Accessed May 28, 2019.

Uber settles with family of woman killed by self-driving car. The Guardian Website. <https://www.theguardian.com/technology/2018/mar/29/uber-settles-with-family-of-woman-killed-by-self-driving-car>. Updated March 29, 2018. Accessed May 22, 2019.

This article is © 2019 by Rebecca Willett. 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.