



Some Reflections on Michael Jordan's article "Artificial Intelligence—The Revolution Hasn't Happened Yet"

Citation

Grosz, Barbara. 2019. Some Reflections on Michael Jordan's article "Artificial Intelligence—The Revolution Hasn't Happened Yet." Forthcoming.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:40564536>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Some Reflections on Michael Jordan's article "Artificial Intelligence—The Revolution Hasn't Happened Yet"¹

Barbara J. Grosz
Higgins Professor of Natural Sciences
Harvard SEAS (Computer Science)

In this article, Michael Jordan makes evident that many have lost sight of the full richness of human intelligence and have neglected to separate foundational understanding from engineering. Most importantly, he points out the need to develop an "engineering discipline . . . for the data-focused and learning-focused fields" and that the systems based on their methods "should be built to work as claimed". A distinguished machine learning (ML) insider, he speaks with authority, bringing insight to current discussions of promise of artificial intelligence (AI) and potential threats it raises for societal wellbeing. The article is, nonetheless, missing two important pieces of the story. One provides an important historical lesson. The other makes manifest a crucial dimension of the engineering challenges. I describe them in turn and then indicate ways, taken together, they should inform the engineering discipline Jordan envisions.

The historical lesson

In his otherwise excellent untangling of the "rebranding" of AI, which distinguishes between human-imitative AI and "Intelligence Augmentation," Jordan overlooks the long history of these two goals and the relationship between the fields that developed to meet each. In the early 1960's, contemporaneously with the emergence of AI research labs, Doug Engelbart founded the Augmentation Research Center at SRI (responsible for such innovations as the computer mouse, hypertext, and collaborative editing), with exactly the goal of using the power of computer systems to augment human intelligence (Engelbart 1962). This work lies at the root of the field of human-computer interaction (HCI). Unfortunately, from their founding, AI and HCI competed, with leading figures in each field denigrating the goals and methods of the other, to the detriment of both. It is only recently that researchers have bridged this divide, recognizing that for systems to behave intelligently, they require both AI capabilities and well-designed interaction methods.

The missing dimension: people and society

Jordan gives only brief mention to the importance of humanities and social sciences to this endeavor, considering them "perspectives" rather than central players in determining the values and principles that will form the foundation of the engineering discipline. Rights of both individuals and society are at stake. As Ece Kamar (2016) has persuasively argued, the AI/ML systems currently making headlines work only because they are hybrid human-machine systems. They are not only *of* humans (when the data are about people), but also *by* humans (who are involved in "human-in-the-loop" creation and curation of the data as well as evolution of the systems), and *for* humans (who use or work with them). It is thus crucial, as Fei-Fei Li (2018) has argued, that the rights, cognitive capacities and needs of people be considered in engineering these AI/ML systems. Gray and Suri's *Ghost Work* (2019) makes evident the individual and societal costs being paid by the hidden, often creative, human labor on which AI/ML systems

¹ The author's manuscript of this article, available at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:40564536>, is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

depend, and the unmet rights such labor has. Thus, in developing principles for the engineering discipline Jordan envisions for data- and learning-focused fields, it is crucial to consider the ethical challenges AI/ML systems raise and to integrate ethical reasoning into this discipline as well as into CS education and AI systems design (Grosz et al., 2019). It is likewise important that the principles related to empirical work in this new discipline reflect an understanding of the principles social scientists have developed for fairly and ethically involving people in their research.

Toward hybrid, collaborative intelligent systems

To meet the goals Jordan sets out for a "human-centric engineering discipline" will require remembering the historical lesson and including the important people-society dimension as a central element. Taken together, these pieces of the story indicate it would be unwise to disregard relevant research from "classical AI" simply because it is outside the scope of current work on data- and learning-focused methods. When matters of life and well-being are at stake, as they are in systems that affect health care, education, work and justice, AI/ML systems should be designed to complement people, not replace them. They will need to be smart *and* to be good teammates. Research in classical AI on teamwork provides a framework for developing engineering principles for hybrid, collaborative systems. It has also yielded many results that are relevant to data- and learning-focused research, even though the algorithms and methods are quite different. For instance, research on computational models of dialogue established that the structure of a dialogue mirrors the structure of the purposes of the participants in that dialogue. Dialogues are not simply linear sequences of utterances nor adjacency (question/response) pairs. These features not only make evident why open domains (e.g., social chatbots, Twitter) are harder than closed domains (travel assistants, customer service) and why short dialogues are easier than extended ("real") dialogue, but also could provide guidance for how to structure the representations used by machine learning algorithms (Grosz, 2018).

The bright flashes of AlphaGo and of speech and image processing systems in personal assistants and online searches have blinded many to the limitations of current AI/ML methods and algorithms. These limitations span individual intelligent behaviors—neither speech nor image processing is perfect nor as robustly flexible as human capabilities—and "general purpose AI" capabilities, which pale in comparison with human-level intelligence. News reports alternate between awe of AI-enabled systems and dismay when the societal implications of systems deploying ML and big data analytics become evident (of which bias is just one). Jordan's article points in a crucial direction for remedying this situation. It is an important read for data science and ML researchers, the developers of products using their methods, policy makers, and the general public.

References:

Englebart, D.C. (1962). *Augmenting Human Intellect: A Conceptual Framework*. Summary Report AFOSR 3223 under Contract AF 49(638)-1024, SRI Project 3578 for Air Force Office of Scientific Research. Menlo Park, Ca., Stanford Research Institute.
https://web.stanford.edu/dept/SUL/library/extra4/sloan/mousesite/EngelbartPapers/B5_F18_ConceptFrameworkInd.html

Gray, M. and S. Suri (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt Press.

Grosz, B. (2018). Smart Enough to Talk With Us? Foundations and Challenges for Dialogue Capable AI Systems. *Computational Linguistics* 44:1.

Grosz, B., D. Grant, K. Vredenburg, J. Behrens, et al. (2018). Embedded EthiCS: Integrating Ethics Broadly Across Computer Science Education. *Communications of the ACM (CACM)*, August 2019.

Kamar, E. (2016). Hybrid Intelligence and the Future of Work. In *Proceedings of the Productivity Decomposed: Getting Big Things Done with Little Microtasks Workshop at CHI*.

Li, F. (2018). How to Make AI That's Good for People, *New York Times*. March 7, 2018.