Harvard Data Science Review • 1.1

Taking Up the Revolutionary Call: Principles to Guide a Purpose-Driven AI Future

Brendan McCord

Published on: Jun 23, 2019

Updated on: Dec 10, 2019

DOI: [10.1162/99608f92.ee412edd](https://doi.org/10.1162/99608f92.ee412edd)

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

Professor Jordan's article, “Artificial Intelligence—The Revolution Hasn't Happened Yet,” issues a timely call to action for the future of the field. Jordan dissects prior work in AI to show us that there is quite a long way to go and reminds us that many of today's advancements focus on one particular method: deep learning. He breaks apart the modeling of human cognition from the creation of a new engineering discipline. Jordan cautions us that the terms we use, such as “revolution” and “artificial intelligence”—and the recklessness with which we use them—may harm long-term progress.

Jordan has played an eminent part in the discipline as it stands today. Thus it is with understandable hesitation that I disagree with three of his core claims regarding the current state of AI.

1. **Jordan avows that an AI revolution has not happened while demonstrating that one is, in fact, occurring.** The structure of the AI revolution has depended less on linear progress from mathematical objectives and more on the basic laws of economics asserting themselves. Useful algorithms are being applied to the systems put in place by past revolutions, yielding societal-scale effects.
2. **Jordan's desire to circumscribe the use of the term "AI" is misguided.** While the multitude of perspectives flowing together in the modern AI discourse has indeed destroyed some knowledge and led to extremism, tribalism, and misunderstanding, on balance, the use of the single overarching term “AI” is helpful, precisely because of the breadth of discourse it stimulates.
3. **Jordan is overly concerned that the pursuit of "intelligence in silicon" modeled after the human brain constrains AI's ability to address humanity's pressing challenges.** In reality, the brain is clearly one evocative and useful source of inspiration, and we will need many; we should get inspiration from wherever we can.

While these critiques concern the field's present, the central question then arises: what is the right strategy for future progress? Of all possible frames, which are the most efficient, and to what end? What roles should each of us play in this important project?

Jordan's opening vignette provides his answer. With this personal story about data analysis issues in Down syndrome testing as a backdrop, he makes the urgent appeal for the need to “conceive of something historically new — a human-centric engineering discipline.” On this call to action, I wholeheartedly agree.

If the discipline is to steer toward a human-centered approach, however, structural changes must occur. A shift to an AI discipline that is more purpose-driven will strengthen the character of the AI revolution for good.

A revolution deferred?

Jordan asserts that the AI revolution “hasn’t happened yet,” then advances the counterargument for that claim, detailing how the rules of the game have indeed changed. He argues that it is now possible to build societal-scale, inference-and-decision-making systems that involve machines, humans, and the environment. Efforts are underway in domains ranging from medicine to commerce, transportation, finance, education, and defense. This new infrastructure has immense resources and vast implications for individual humans and societies. While the potential benefits are undeniable, such systems are already exposing serious conceptual flaws.

Stakeholders across society are grappling with this shift. For the first time in the careers of senior academics, AI is a topic that almost everyone feels is on the critical path for their sub-discipline. In industry, the AI technique of machine learning could power essentially any company in which decisions can be tied to large scale data and will spur the emergence of new products, business models, and industries.

Most would have no qualms describing such circumstances as the early stages of a revolution. Jordan holds conflicting ideas. On one hand, he grasps that we are functionally in the midst of a new paradigm, and this is precisely why we need a new engineering discipline; on the other hand, he insists there has been no “revolution.”

One explanation for this paradox is that Jordan is focused on technique.¹ He acknowledges spending much of his time inside an intellectual bubble of engineers and mathematicians; thus, despite his appreciation of the societal-scale effects of today’s AI, his main unit of analysis remains mathematical progress. Through this lens, he sees straightforward accumulation of facts dating back decades and is underwhelmed.

This ends up being a peculiar standard. Indeed, it ignores the structure of the AI revolution, which depends less on linear progress from mathematical objectives and more on the ways in which a mélange of multiple technological advancements—some recent and some delivered to us by prior revolutions—cause a powerful economic phenomenon to occur: namely, the dramatic drop in price of a foundational input to decision-making.² As was the case in other revolutions, the laws of economics can be far more important in shaping human history than those steeped in a purely mathematical way of thinking are prepared to accept (Shapiro & Varian, 2010).

The 2018 book *Prediction Machines* lays out the rationale for looking to durable economic principles versus looking through the lens of mathematical progress to judge the AI revolution (Agrawal, Gans, & Goldfarb, 2018). Consider the comparison case of computing, a technology which undoubtedly launched a revolution. Stanford economist Tim Bresnahan reminds us that “computers do arithmetic

and nothing more. The advent and commercialization of computers made arithmetic cheap. When arithmetic became cheap, not only did we use more of it for traditional applications of arithmetic, but we also used the newly cheap arithmetic for applications that were not traditionally associated with arithmetic” (Agrawal et al., 2018). By reducing the general cost of something important—in this example, computation—technology changed the world.

Similarly, developments in AI are transformative insofar as they are causing a dramatic drop in the price of something that is foundational to decision-making: useful prediction. This is true despite the reality that the technologies are a long way from being done.

Whatever may be said about purely engineered AI, it is demonstrably useful in an ever-widening space of problems. Because it so efficiently draws upon the infrastructure of past revolutions—from electricity, to the computer revolution, to the rise of the intensely networked era that began in the 1970s, known as the Internet revolution—a whole world of AI applications is emerging around us. The net effect is a massive price drop for useful prediction.

As was the case in computing, this means that not only are we going to start using a lot more prediction, we are going to see it emerge in surprising new places (Agrawal et al., 2018). Historical examples such as the Internet demonstrate that these transitions will continue to unfold over decades, but the AI revolution now underway demands both our best thinking and our prompt acknowledgment.

An almighty initialism

Terms and their connotations have differed over the discipline’s history, but have always been integral to progress, and therefore merit scrutiny (Kuhn, 2015). In addition to “revolution,” the other term with which Jordan takes issue is ‘AI.’

There is a pattern in science wherein two factions form during paradigm shifts: one of researchers and the other of laypeople. At first, the former uses or develops terminology unfamiliar to the latter (Slotten, 1977). As time goes on, both groups come to use the same terminology, and aspects of scientific knowledge are successfully shared.

By contrast, ‘artificial intelligence’ is not the classical case of an exoteric versus esoteric population. Instead, the phrase is simultaneously being “intoned by technologists, academicians, journalists, and venture capitalists alike,” with each group struggling to understand one another while contributing to an emerging thought collective. This explains people’s increased awkwardness or embarrassment in

employing the term, even while the capacity of the technology to live up to its magical connotations is rising.

Yet use of the single overarching term ‘AI’ is helpful, precisely because of the breadth of discourse it stimulates. ‘AI’ brings general circulation, and general circulation is essential to harnessing the wisdom of the crowds in connecting a general economic force to its field- and application-specific implications.

Here we have a class of technologies leaving the laboratory at a high rate and “growing into massive industrial relevance” on their way to plausibly impacting most facets of human endeavor over the next century. Using a term that entices diverse groups of people to contribute improves the likelihood that this transition will benefit humanity. To stretch the point, this is the logic that says that a government shaped not merely by elites but by the people will yield better outcomes, even though coordination costs are higher and average expertise is lower.

For researchers, there is an opportunity to embrace the new formulation. The resurgence of the term ‘AI’ coincides with the ushering in of new, non-superficial shifts in scope and responsibility. Ex-“optimization or statistics researchers” now find themselves part of a field that transcends not only their former academic discipline, but also the entire academic enterprise.

Success in AI requires breaking down barriers across sub-disciplines such as computer systems thinking, inferential thinking, human-centered design, public policy, ethics, and more. It demands curiosity for new challenges that emerge as these technologies enter the world. The reward for academics who embrace this incipient field will be the ability to explore meaningful new theoretical questions that are engendered and the chance to have an impact at “societal-scale.”³

A choice of inspiration

One of Jordan’s consistent critiques of the term ‘AI’ is that it nudges us toward the goal of emulating the human brain. This echoes a longstanding argument dating back to the 1950s (Simon et al., 2000).

However, there is no debate that the study of the brain has and will continue to inform AI. For example, breakthroughs such as convolutional neural networks and the hierarchical organization of many current architectures depended directly on neuroscience observations or the modeling of human behavior (Hassabis, Kumaran, Summerfield, & Botvinick, 2017). Leading researchers such as Geoff Hinton and Yann LeCun have fruitfully explored computational principles found in the brain for decades, and many of today’s AI techniques and exemplars reference neuroscience. What then 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. Therefore, it is no surprise that the analog of human intelligence rushes into the vacuum.⁴ Put another way, “the search space of possible solutions [to intelligence] is vast and likely only very sparsely populated...this therefore underscores the utility of scrutinizing the inner workings of the human brain—the only existing proof that such an intelligence is even possible.”

The same was true of the airplane: how would the Wright brothers have envisioned flight without thinking of birds? They chose what bits of inspiration to borrow and test, then solved the engineering challenges required for liftoff at human sizes and weights (McCutcheon & Baxter, 2013). Eventually, a new discipline of aerospace engineering developed around engineered flight, and billions of humans observed it in airplanes and rockets (Stanzione, 1989).⁵ We now look back at the early animal-inspired imaginings of mechanical flight with amusement.

Jordan is correct that today we still know very little about how the brain works (Beltramo & Scanziani, 2019).⁶ We should view our lack of knowledge as an opportunity to push forward in exploring the computational principles of the brain through AI or vice versa—research directions with the potential to further and validate both fields in the process. Even if we acknowledge the difficulty of understanding the human brain, we can pursue its characteristics, such as nuance and versatility, in the design of our human-centered AI systems.

In reality, while the brain is clearly one evocative and useful source of inspiration, we will need others. Researchers in natural language processing or document modeling may find inspiration through tree-based architectures in nature; investigators of autoencoders may find it in entropy or free energy and thermodynamics; and explorers of Bayesian probability theory may draw their inspiration from long walks or the turning of the crank of mathematics. Progress itself can be a form of kindling.

The multiplicity of inspirations, and the stochasticity of individual researchers choosing which path of exploration to pursue, help avoid the notion of effective local minima. This lowers the odds that progress will get stuck. We should follow the advice of physicist Richard Feynman: it does not do any good to “just [increase] the number of [researchers] following the comet head. [I]t’s necessary to increase the amount of variety” (Gleick, 2014).

Beyond these critiques regarding terminology and sources of inspiration, the crucial question then arises: what is the right strategy for future progress? There is growing evidence that our current approach will not work for the future.

Past: Government-backed discovery

Taking research and development (R&D) expenditure as a proxy for innovation, the U.S. government sponsored most technology of consequence during the period from 1940-1990 (“National Patterns of R&D Resources,” 2016). Technology that the Department of Defense (DoD) developed and used defined state-of-the-art.

In AI, DoD focused on enabling risky discovery through guiding fundamental research and backing the foremost thinkers in the field. Funding from the Defense Advanced Research Projects Agency (DARPA) in areas such as perception, natural language understanding, and navigation ultimately led to the creation of self-driving cars, personal assistants, and near-natural prosthetics, in addition to a myriad of other critical and valuable military and commercial applications (“Where the Future Becomes Now,” n.d.). To say that the partnership among DoD, industry, and academia succeeded in transforming technological practice would be an understatement; it transformed the world.

At the end of the Cold War, however, the locus of innovation in many arenas shifted to the global commercial technology sector. Federal R&D declined as a percent of gross domestic product, while market forces caused private sector R&D to surge upward. As of 2016, commercial R&D expenditures comprised roughly \$220 billion more than federal R&D, with outsized investments being made by a small number of leading AI firms (“National Patterns of R&D Resources,” 2016).⁷

Today: Commercial puzzle-solving

Today, the business needs and value networks of a few major Internet companies substantially drive AI development. Decadal shifts in R&D funding are part of the explanation, along with another, more recent contributing factor that has emerged: today, AI makes money.

In numerous settings, machine learning in particular has enabled performance levels that are above a usability threshold for valuable commercial products. Small improvements in models often translate into profits, which in turn fund further algorithmic development, and the flywheel spins onward (Turck, 2016).⁸

To seize on these effects, major Internet companies have channeled a portion of their profits towards the formation of elite AI teams and endowed them with unique business assets, such as large proprietary datasets and massive computing. Meanwhile, companies selling complements, such as AI hardware, companies whose offerings become more valuable as the size and number of datasets increase, have invested in further accelerating the dynamic.

The net effect is that commercial puzzle-solving—defined here as the handiwork of researchers solving specific engineering challenges to feed value networks in industry—guides the AI field more so than any other source of research inspiration.

The benefits of this phenomenon are clear: major research and systems-building successes from Google, Facebook, Microsoft, Amazon, Apple, Netflix, Baidu, Alibaba, Tencent, SenseTime, and more. These breakthroughs are in solution areas that are relevant to Internet company problems, such as document retrieval, text classification, fraud detection, recommendation systems, personalized search, social network analysis, planning, diagnostics, and A/B testing. Fortunately, these advances can be ported to other domains as well.

The pernicious effects of the current strategy are similarly apparent. A few of the aforementioned entities have adopted business models fine-tuned for the development and deployment of AI yet potentially dangerous to civil liberties (Kania, 2018). Additionally, the role of government in the field of AI and its capacity to address these challenges have diminished just as the potency of the technology has reached into more and more domains of public importance. As a result, big tech is assuming the mantle of societal stewards of the technology and is increasingly finding itself on shaky ground.

For example, in 2018, Microsoft called for a national conversation on responsible uses of facial recognition technology (Smith, 2018), Google was at the center of a polarized debate regarding its stance on national security uses of technologies ranging from computer vision to the cloud (Shane & Wakabayashi, 2018), and Facebook showed how its platform could be weaponized in a manner that has the potential to threaten our societal model (Isaac & Wakabayashi, 2017). These examples are early signs that the current flywheel of commercially-driven technological progress does not spin cleanly in a vacuum.

Perhaps the most concerning implication of the current strategy is the risk it poses to the future of academic computer science. In the past, public sector organizations such as DoD or NASA posed big problems of societal significance and marshaled the resources needed for researchers in academia to help solve them. Jordan's example of the prehistory of the famous backpropagation algorithm, which arose in the 1950s and optimized the thrusts of the Apollo spaceships as they headed toward the moon, is one such instance from the 20th century. With the motivation of a must-solve problem and the requisite marshaling and management of resources, the Apollo case yielded technological breakthroughs across multiple domains, inspired possibly millions of new researchers and engineers to work on hard problems of public purpose, and contributed to precisely the outcome Jordan seeks: the creation of a new engineering discipline (Cameron, n.d.).

With the funding model for AI shifted to consumer parts of the economy and the government on the sidelines, it is no surprise that young researchers are drifting out of academia. They can find use cases

in industry, along with billions of users, exaFLOPS of compute, attractive personal economics, and the fellowship of other brilliant individuals who have followed that same path. Professors also find enticement either to leave, split their time between universities and companies, or convert to tech employees through “dual affiliation” (Recht, 2018). Society benefits from a stable academic sector, and the world urgently needs new solutions for problems of societal significance.

Jordan’s penultimate paragraph is a clarion call for the reader: “In the current era, we have a real opportunity to conceive of something historically new.” This project entails nothing less than the creation of a new human-centric engineering discipline. Rather than refute Jordan, the following principles can help guide the way forward.

1. Develop the theory iteratively with practical action. The back-and-forth between theoretical and practical ideas in order to solve a problem has shaped nearly every modern engineering discipline, and can be a blueprint for AI’s future. Indeed, many theoretical questions only emerge as the technology translates into impact via integration with workflows, processes, and decision-making.

In the case of civil engineering, there was not a practical phase of bridge building followed by a principles phase. Instead, the theoretical and practical ideas co-evolved. Similarly, in electrical engineering, “Maxwell’s equations provided the theory, but ideas [such as] impedance matching came into focus as engineers started to learn how to build pipelines and circuits” in order to address real world applications such as digital computing (Jordan, 2015).

The challenge in AI today is finding strategies that accelerate the cycle between theory development and practical action, particularly with respect to solving complex, large-scale, time-critical, and important problems. Examples include mitigating the impact of chronic diseases, improving education, addressing accelerating environmental degradation, and ensuring global and national security (Chui et al., 2018).

Consider the mission of humanitarian assistance and disaster relief. Worldwide, natural disasters kill 90,000 people and affect roughly 160 million people every year (“Natural Events,” 2012). Over the past two years, the costs of damage from disasters is estimated at \$510 billion (“Natural Catastrophe Review,” 2018).

Addressing this exigent global need would deliver practical benefits in the form of human lives and livelihoods. When disasters strike, critical operations need to be coordinated—civilians need to be evacuated, supplies transported, resources allocated, first responders deployed, victims located, and more. Despite recent successes, AI is not currently at the level of maturity where it can be used at scale to help people at the front lines make critical decisions, fight time, and ultimately save lives.

Applying AI toward challenges of this sort would not mean neglecting theory advancement, and may even lead to richer theory than if theory were developed in isolation. For example, how can human judgments integrate dynamically with AI models to enable them to generalize (e.g., from hurricanes to forest fires)? How should an autonomous vehicle react if it encounters a time critical scenario (e.g., a person is stuck in a structure where the water level is rising quickly) that does not match any of the test cases on which it was trained?

The iterative development of theory with practical action to address real world problems will be a critical step toward the ability for the AI discipline to meet crucial engineering guarantees. Namely, it will help establish the ability to turn the core inferential ideas of AI into engineering systems that work as claimed in the real world—under expected requirements and assumptions, alone or in combination with human decision-makers, and while exhibiting robustness, efficiency, and the minimization of unintended actions, side effects, and consequences.

2. Broaden the scope and disciplinary focus of AI research. If AI researchers are to remain relevant, they must be prepared to devote their efforts to addressing real-world problems. To do so will require crossing traditional disciplines and reaching outside of academia itself. AI researchers will need to understand the interactions among different technologies, public policy, international politics, ethics, and more, as well as the role of these interactions in the behavior of societal-scale AI systems. Researchers must treat the interactions between disciplines with the same rigor that the sciences have applied to the field’s predecessor disciplines over centuries; in other words, they must treat broader context itself as a discipline (Griffin, 2010).

Additionally, as academic research agendas progress, they naturally narrow. By contrast, addressing big problems “in the large” will require that researchers go beyond the pursuit of novel performance benchmarks and put more effort toward progress on other dimensions relevant to social value, economic value, and scientific advancement (Simonite, 2018).⁹ These efforts would transform the field (Martinez-Plumed et al., 2018).

3. Enlarge the discipline. If AI is a holistic discipline, if it is fundamentally concerned with purpose, and if breadth of viewpoint is important to attaining that goal, then AI must diversify as a field. There is real, quantifiable, engineering value in prioritizing diverse perspectives within teams that develop AI (Griffin, 2010). The consequences of not doing so in areas such as fairness, transparency, and ethics are clear (“Keeping an Eye on AI with Dr. Kate Crawford,” 2018).

Early progress advances in several respects. Educational, cultural, and socioeconomic diversity increases as millions of students around the world who would not otherwise have access to high quality training learn from AI leaders such as Coursera’s Andrew Ng and Daphne Koller or Udacity’s Sebastien Thrun. Disciplinary diversity grows through Fast.ai’s platform for getting “unlikely people”

involved in the future of AI, while models such as Lambda School empower individuals to switch careers, often from extreme outsider professions, to the tech industry. Efforts such as the Black in AI movement lead the way in addressing ethnic and racial diversity shortfalls (Snow, 2018).

More efforts such as these will be a central feature of the new discipline. By reshaping the sociology of AI, scientific judgment will change concerning “what kinds of problems we think are important, what kinds of research we think is important, and where we think AI should go,” which increases the likelihood that the technology will benefit broadly (Snow, 2018, p. 3). Today there remains a gap in programs that engage and equip the least technologically attuned. Yet it is civil society as a whole that must ultimately decide and respond to the ways we use or restrain technology, beyond engineering and research.

4. Proactively engage in norm-building and policy matters regarding the societal impact of AI.

Norms within a research community shape technology development and usage, and the system that delivers that technology to the world includes public policy. For these reasons, proactive engagement in both areas is necessary to ensure a positive direction forward.

A culture of value-driven norms is taking shape. In 2018, collective action gave voice to widely-held concerns or sentiments in the AI community (Simonite, 2018). Over time, new norms may disagree with aspects of the current zeitgeist but move us closer to a purpose-driven AI future.

For example, AI researchers grapple with the impact of their increasingly potent inventions on society (Metz, 2018). Several of humanity’s other high-consequence fields—such as nuclear physics, chemistry, and biology—experienced a renaissance of norms governing technology impact assessment and ethics. Leaders such as OpenAI’s Jack Clark and the Association for Computing Machinery’s Andrew Chien raise the awareness of AI as a dual- and omni-use technology (“Preparing for Malicious Uses of AI,” 2018) and highlight concerns regarding its diffusion to groups such as terrorists, criminals, or authoritarian governments (Chien, 2019).¹⁰

In a field where the mantra of the current era is openness, forming norms that are at odds with AI publishing paradigms will be a nuanced pursuit. It is important to retain the spirit of openness that has shaped the modern AI community while advancing the goal of responsible development.

Within the public policy realm, 2018 saw progress in the form of new opportunities for AI researchers to contribute directly through technology policy programs; the growth of AI policy teams; “how to” guides for navigating a career in AI policy; the creation of policy research fellowships; and opportunities for short-term government service.

However, on the role of AI in challenging public issues such as fairness in lending or the countering of violent extremist organizations, some AI researchers still defer responsibility, fall back to individual

views on the world and the current political context, or opt out altogether. Geoff Hinton recently remarked, “I’m an expert on trying to get the technology to work, not an expert on social policy” (Simonite, 2018, p. 10).¹¹ If taken to the extreme, a logic of separating research expertise from policy matters could lead to anomie in the AI community and the risk that the tools AI research produce will be unsuitable for expansive real-world problems or even harmful to society. Likewise, experts in public policy often lack essential technical knowledge. Problems of this type will only be solved when both sides of this equation invest in equipping themselves to effectively collaborate with one another.

A case study that combines the need for nuanced norm-building with constructive engagement on policy matters concerns the use of AI in national security contexts. In 2018, Google withdrew from this class of work with DoD, beginning with separating from a computer vision project called Project Maven and then extending its concerns to the cloud more generally (Shane & Wakabayashi, 2018). Google cited its AI Principles in both cases (Gregg, 2018).¹² Approximately 4,000 out of 100,000 Google employees petitioned to urge this outcome (Shane & Wakabayashi, 2018).¹³

While most would agree that armed forces and their use for deterrence, defense, and offense are definitely a part of our world and will continue to be for the foreseeable future, there is a multiplicity of views within the AI community on the technology’s role in security.

One perspective is that the armies of lawful, democratic countries, whose arsenal is designed for defensive purposes and for the protection of human rights, and who is held accountable, can have a positive influence on promoting peace and stability in the world, and require the responsible use of modern technologies such as AI in order to do so. Those who hold this view would acknowledge that these militaries have been necessary in the past—World War II comes to mind—and have grappled with ethical issues since the time of Cicero.

However, even for individuals whose views are in dramatic opposition to that perspective, rather than opting out entirely, a norm of engagement and constructive dialogue may be more productive.

It could increase the likelihood that their concerns get incorporated into national security policies, practices, and projects; mitigate the risk of perverse results from less-principled actors by filling the void and working on these challenges; and help avoid a world in which there is further factionalization, which would make the upsides of AI more difficult to access. Additionally, it would give these individuals a fuller picture of the sources of risk associated with AI in security contexts, which could help avoid misguided or counterproductive reactions. Constructive engagement is the optimal strategy to employ when addressing these issues.

5. Make purpose the core concern. AI will create new opportunities to address enormous public goods problems. Due to its disruptiveness and potency, AI could also risk aggravating the same. A

purpose-driven approach is necessary at all levels of abstraction in the discipline, from technical to societal.

At a technical level, this will require conducting research and engineering to ensure that AI systems are safe, robust, and secure; pioneering new approaches for AI testing, evaluation, verification, and validation; and creating better ways of understanding and explaining AI-driven decisions and actions so that users can understand, appropriately trust, and effectively manage AI systems. Considerable academic work is needed to specify objectives and quantify attributes of purpose-driven AI systems such as social value, academic value, and scientific advancement in ways which are meaningful to the engineering community (Russell, 2018).

The public sector has an essential role in making purpose the core concern. Advancing past a strategy in which commercial puzzle-solving dominates to a more optimal ratio of public and private goals requires governments to determine which problems are funded and solved. They should prioritize uses of AI to address global public goods such as security, education, and health. As commonly-shared interests, these domains reinforce the case for more collaborative governance models in AI. Further, it is time for new regulatory structures to reinforce purpose, as even well-intended businesses and academics will come up against their limits. In this work, constituents and the international community must ensure that governments are accountable to a purpose-driven path.

Purpose-driven actors in industry—those that pursue what is right and embrace a sense of mission beyond simple value maximization for shareholders—may ultimately be most important for addressing society’s big problems. However, businesses often treat the most vexing challenges as externalities whose solutions are the responsibility of the public sector (Henderson, n.d.).

Businesses must broaden their objectives in AI beyond quarterly earnings toward shared prosperity and long-term profits. Just as it is ultimately not in the corporate self-interest of businesses such as oil and gas companies to wantonly warm the planet, so too is it not in the interest of AI leaders to wield transformative technology without internalizing responsibility for aspects of the larger system upon which their long-term success depends.

The notions of ‘shared value’ or ‘stakeholder theory’ are not only exciting streams of research for the AI community, but also concepts that have been implemented with demonstrable success by leading firms in other segments of the economy (Parrish, 2018). This research suggests that there is an opportunity for AI companies to create competitive advantage by systematically organizing around challenges such as ethics, safety, or business model-specific implications of the technology.

Microsoft formed the “Aether committee” to address challenges associated with the “influences of AI on people and society (Bacchus, 2018).” Google shared its thinking on responsible AI practices,

articulated AI principles, and issued perspectives on issues in AI governance (Walker, 2018). Over 20 for-profit organizations have joined the partnership on AI to establish best practices and educate the public on AI (“Meet the Partners,” n.d.). These bodies and artifacts create shared values, give employees a tool to push back on management, and ensure that AI development and deployment aligns with a broader sense of purpose.

‘AI for Good’ initiatives are on the rise (Sennaar, 2019). There are two opportunities to enhance the structure of these endeavors. First, we should not treat these efforts as wholly distinct projects from core business activities. Profits power AI for Good in much the same way that a billionaire CEO might direct a portion of her wealth to social causes; the core businesses remain unchanged. Second, the mindset of agile development leads firms to find and tackle proximal problems; tech companies are highly efficient apparatuses for addressing challenges of the ‘me right now’ form, but not the ‘us later’ variety.

The risk is that AI for Good as a philanthropic pursuit creates near-term products that can deliver a concrete moment of good—an approach that may be, as former Stanford President John Hennessy suggests, vulnerable to short-termism (Thompson, 2018). Should actors frame AI for Good more comprehensively and then figure out the bite-sized downstream actions needed to get there? Can the community figure out what AI looks like when numerous countries and corporations apply it at the international level in a resourced way over a long period of time?

Industry has yet to explore a much more expansive notion of AI for Good. What would truly purpose-driven AI firms look like? Could they survive? Could they change the world? Businesses acting as businesses, not merely as charitable donors, might be the most powerful force for addressing the pressing issues we face.

The duties of the revolutionaries

It is critical to think about what it means to articulate and instill a vision compelling enough to address this more comprehensive work. The world saw what a successful project that combines the efforts of industry, academia, and government looked like through the Apollo program. With all forces concentrated on a single object, and through vigorous development in areas of advanced technology, humans unlocked the exploration of outer space. This problem-solving yielded breakthroughs such as the computer and international telecommunication and inspired generations of researchers.

What would be the purpose-driven, democratic version of the AI megaproject in the modern era? Perhaps it is a new mission to create a support and response system for the world in times of natural

disaster, or a new platform for collaborative scientific research and inter-corporate cooperation, comparable to CERN for AI (Hill, 2018).

Across projects and organizations, what principles should guide purpose-driven work on big problems in the area of global public goods? DoD has begun to address those questions in the context of its security mission. This auspicious start included the articulation of the Department's first AI strategy, the drafting of *Defense AI Principles*, and the formation of the Joint AI Center as a focal point for bringing these ideas to life (Leung & Fischer, 2018).

This beginning represents just a tiny fraction of what needs to be done across the U.S. government and throughout civil society. How should the nation proceed during this AI-based technology transition? I urge more boldly – and in what David Ignatius reminds us used to be called, the American way (Ignatius, 2010), a term which has nothing to do with national geography and everything to do with protecting those values that came out of the Enlightenment.

While the AI revolution cannot be denied, the shift to a purpose-driven AI discipline will help to strengthen its character for good. This new AI discipline will be focused on advancing theory iteratively with practical action, enlarged by diverse perspectives, broadened in scope and disciplinary focus, and guided norms concerning societal impacts.

Jordan is right that the magnitude and importance of this task is our call to action. Success in designing, building, and operating efficient large-scale AI systems to serve crucial purposes in both public and private sectors is a defining requirement for societal advancement and economic prosperity in the world of today and tomorrow.

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)
- [Max Welling](#) (University of Amsterdam)

- [Rebecca Willett](#) (University of Chicago)

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

References

Shapiro, Carl, and Hal R. Varian. *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business School Press, 2010.

Agrawal, Ajay, et al. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press, 2018.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. The University of Chicago Press, 2015.

Slotten, Ralph. "Exoteric and Esoteric Modes of Apprehension." *Sociological Analysis*, vol. 38, no. 3, 1977, p. 185, doi:10.2307/3709801.

Simon, H.a., et al. "AI's Greatest Trends and Controversies." *IEEE Intelligent Systems*, vol. 15, no. 1, 2000, pp. 8–17., doi:10.1109/5254.820322.

Hassabis, Demis, et al. "Neuroscience-Inspired Artificial Intelligence." *Neuron*, vol. 95, no. 2, 2017, pp. 245–258., doi:10.1016/j.neuron.2017.06.011.

"Taking to the Skies: The Wright Brothers and the Birth of Aviation."

<http://www.sil.si.edu/ondisplay/flight/intro.htm>.

Stanzione, Kaydon Al (1989). "Engineering". *Encyclopædia Britannica*. 18 (15 ed.). Chicago. pp. 563–563.

Beltramo, Riccardo, and Massimo Scanziani. "A Collicular Visual Cortex: Neocortical Space for an Ancient Midbrain Visual Structure." *Science*, vol. 363, no. 6422, 2019, pp. 64–69., doi:10.1126/science.aau7052.

Gleick, James. *Genius The Life and Science of Richard Feynman*. Open Road Media, 2014.

"National Patterns of R&D Resources." National Science Foundation. May 21, 2016.

<https://www.nsf.gov/statistics/2018/nsf18309/>.

"Where the Future Becomes Now." *Defense Advanced Research Projects Agency*,

<https://www.darpa.mil/about-us/darpa-history-and-timeline>.

Turk, Matt. "The Power of Data Network Effects." January 4, 2016. <http://mattturck.com/the-power-of-data-network-effects/>.

Kania, Elsa B. "China's AI Giants Can't Say No to the Party." August 02, 2018. <https://foreignpolicy.com/2018/08/02/chinas-ai-giants-cant-say-no-to-the-party/>.

Smith, Brad. "Facial Recognition: It's Time for Action." The Official Microsoft Blog. December 6, 2018. <https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action/>.

Shane, Scott, and Daisuke Wakabayashi. "'The Business of War': Google Employees Protest Work for the Pentagon." April 04, 2018. <https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html>.

Isaac, Mike, and Daisuke Wakabayashi. "Russian Influence Reached 126 Million Through Facebook Alone." October 30, 2017. <https://www.nytimes.com/2017/10/30/technology/facebook-google-russia.html>.

"Meet Margaret Hamilton, the Scientist Who Gave Us 'Software Engineering'." IEEE Software Magazine. January 04, 2019. <https://publications.computer.org/software-magazine/2018/06/08/margaret-hamilton-software-engineering-pioneer-apollo-11/>.

Recht, Benjamin. "You Cannot Serve Two Masters: The Harms of Dual Affiliation." *Arg Min Blog*, <http://www.argmin.net/2018/08/09/co-employment/>. This phenomenon is particularly concerning, as it may conceal the looming stability crisis in academic computer science; it conjures an image of a host organism that appears to function for a short while after having been colonized by what is being termed "co-employment" or "dual affiliation."

"r/MachineLearning-AMA: Michael I Jordan." *Reddit*, www.reddit.com/r/MachineLearning/comments/2fxi6v/ama_michael_i_jordan

Chui, Michael, et al. "Applying Artificial Intelligence for Social Good." McKinsey & Company. November 2018. <https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good>.

"Natural Events." *Environmental Health in Emergencies*, World Health Organization, 24 Aug. 2012, www.who.int/environmental_health_emergencies/natural_events/en/.

"Natural Catastrophe Review 2018." *Extreme Storms, Wildfires and Droughts Cause Heavy Nat Cat Losses in 2018*, Munich Re, 8 Jan. 2019, www.munichre.com/en/media-relations/publications/press-releases/2019/2019-01-08-press-release/index.html.

Griffin M.D., “How Do We Fix System Engineering?”, IAC-10.D1.5.4, 61st International Astronautical Congress, Prague, Czech Republic, 1 Oct. 2010.

Simonite, Tom. “Google's AI Guru Wants Computers to Think More Like Brains.” *Wired*, Conde Nast, 12 Dec. 2018, <https://www.wired.com/story/googles-ai-guru-computers-think-more-like-brains/>.

Martinez-Plumed, Fernando. “Accounting for the Neglected Dimensions of AI Progress,” arXiv:1806.00610 [cs.AI], Jun. 2018.

“Keeping an Eye on AI with Dr. Kate Crawford.” Microsoft Research Podcast. May 23, 2018. <https://www.microsoft.com/en-us/research/blog/keeping-an-eye-on-ai-with-dr-kate-crawford/>.

Snow, Jackie. “‘We’re in a Diversity Crisis’: Cofounder of Black in AI on What’s Poisoning Algorithms in Our Lives.” *MIT Technology Review*, 14 Feb. 2018, www.technologyreview.com/s/610192/were-in-a-diversity-crisis-black-in-ais-founder-on-whats-poisoning-the-algorithms-in-our/.

Simonite, Tom. “2018 Was the Year That Tech Put Limits on AI.” *Wired*, Conde Nast, 21 Dec. 2018, www.wired.com/story/2018-was-year-tech-put-limits-on-ai/.

Metz, Cade. *Efforts to Acknowledge the Risks of New A.I. Technology*. The New York Times, 22 Oct. 2018, www.nytimes.com/2018/10/22/business/efforts-to-acknowledge-the-risks-of-new-ai-technology.html.

“Preparing for Malicious Uses of AI.” *OpenAI Blog*, OpenAI, 21 Feb. 2018, blog.openai.com/preparing-for-malicious-uses-of-ai/.

Chien, Andrew A. *Open Collaboration in an Age of Distrust*. ACM, 1 Jan. 2019, cacm.acm.org/magazines/2019/1/233509-open-collaboration-in-an-age-of-distrust/fulltext.

Bowerman, Niel. “The Case for Building Expertise to Work on US AI Policy.” 80,000 Hours. January 31, 2019. <https://80000hours.org/articles/us-ai-policy/>.

Gregg, Aaron. “Google Bows out of Pentagon's \$10 Billion Cloud-Computing Race.” *The Washington Post*, WP Company, 9 Oct. 2018, www.washingtonpost.com/business/2018/10/09/google-bows-out-out-pentagons-billion-cloud-computing-race/.

Russell, Stuart. “How to Make AI That Works, for Us.” *Science Focus*. November 16, 2018. <https://www.sciencefocus.com/future-technology/how-to-make-ai-that-works-for-us/>.

Henderson, Rebecca. “Reimagining Capitalism: Business, Purpose and the Big Problems.” Harvard College, college.harvard.edu/college-events/rebecca-henderson-reimagining-capitalism-business-purpose-and-big-problems.

Parrish, Christine. "Redefining Capitalism: Do The Right Thing, Make Money, Change the World." *The Free Press*, 12 Apr. 2018, freepressonline.com/Content/Home/Homepage-Rotator/Article/Redefining-Capitalism-Do-The-Right-Thing-Make-Money-Change-the-World/78/720/58009.

Bacchus, Arif. "Microsoft Gives up 'Significant Sales' over AI Ethics Concerns OnMSFT.com." *On MSFT*, 16 Apr. 2018, www.onmsft.com/news/microsoft-gives-up-significant-sales-over-ai-ethics-concerns.

Walker, Kent. "Google AI Principles Updates, Six Months In." Google, 18 Dec. 2018, www.blog.google/technology/ai/google-ai-principles-updates-six-months/.

"Meet the Partners." The Partnership on AI. Accessed February 10, 2019. <https://www.partnershiponai.org/partners/>.

Sennaar, Kumba. "AI for Good – An Overview of Benevolent AI Initiatives | Emerj - Artificial Intelligence Research and Insight." Emerj. January 31, 2019. <https://emerj.com/ai-sector-overviews/ai-for-good-initiatives/>.

Thompson, Nicholas. "John Hennessy on the Leadership Crisis in Silicon Valley." *Wired*, Conde Nast, 26 Oct. 2018, www.wired.com/story/leadership-crisis-in-silicon-valley-john-hennessy/.

Hill, Steven. "Opinion Piece: Why We Need a CERN for AI." *Ethics of Algorithms*, 22 May 2018, ethicsofalgorithms.org/2018/05/22/opinion-piece-why-we-need-a-cern-for-ai/.

Leung, Jade, and Sophie-Charlotte Fischer. "JAIC: Pentagon Debuts Artificial Intelligence Hub." *Bulletin of the Atomic Scientists*. August 08, 2018. <https://thebulletin.org/2018/08/jaic-pentagon-debuts-artificial-intelligence-hub/>.

Ignatius, David. "Can the Pentagon Build a Bridge to the Tech Community?" *The Washington Post*, WP Company, 24 Jan. 2019, www.washingtonpost.com/opinions/can-the-pentagon-build-a-bridge-to-the-tech-community/2019/01/24/39c0e3b2-2019-11e9-9145-3f74070bbdb9_story.html.

This article is © 2019 by Brendan McCord. 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.

Footnotes

1. Another possible explanation for his seemingly conflicting views is that Jordan is being ironic; he is fighting extravagance with extravagance. [↵](#)

2. ² Closely related to these technological advancements are the presence of more widely available inputs such as data and compute. [↵](#)
3. ³ While the importance of general circulation in addressing societal impacts argues for the use of a single, widely-accessible handle such as ‘AI,’ there remains a complementary opportunity to break apart the large constellation of techniques to which that term refers. Jordan succeeds in providing a coherent framework for doing so with three groups of largely distinct aspirations: “human-imitative AI,” “intelligence augmentation,” and “intelligent infrastructure.” I suspect any taxonomy of this space will need to be revised every few years, given the rate of change and uncertainty about where the technology is headed. [↵](#)
4. ⁴ Indeed, although Jordan’s three categories were intended to steer us away from brain-inspired intelligence, it is interesting to note that all three have strong linkages to the same. His first category, “human-imitative AI,” is the brain itself. His second category, “intelligence augmentation,” is analogous to how we have been designed by evolution as social primates that use our brains cooperatively to interact with each other and our environment. His third category, “intelligent infrastructure,” captures the emergent property of a collection of human brains that is a society, and the knowledge creation that occurs in that stratum. To a certain degree, his description of intelligent infrastructure also applies to the brain itself, which uses distributed representations and may itself be a “society of mind.” [↵](#)
5. ⁵ One shortfall of this analogy is that in aerospace engineering the low-level technology is the same between birds and planes (lift, drag, turbulence, and fluid mechanics, generally), whereas the high-level technology is completely different (commercial aircraft, air transportation systems, etc.). In the brain, the higher-level technology (e.g., massive interconnectivity of relatively simple computing elements) is similar, but the lower-level technology (e.g., neurons versus tensor algebra) is vastly different. [↵](#)
6. ⁶ For instance, earlier this year, this paper published in *Science* challenged our assumptions about one of the best understood areas of the brain, the visual system. [↵](#)
7. ⁷ As of 2016, the total R&D spend was \$510B, with federal R&D accounting for roughly \$122.9B and commercial R&D for approximately \$342.7B. Google’s parent company Alphabet on its own spent roughly \$21.4B annually in R&D in 2018. [↵](#)
8. ⁸ In network markets such as mapping, advertising, or search that have hundreds of millions of customers and tend to exhibit “winner take all effects,” the positive feedback from incremental algorithmic progress can mean massive and disproportionate rewards. This is a macro version of a phenomenon articulated by many, including Matt Turck. [↵](#)

9. ⁹ Geoff Hinton reflects on this situation: “if you want to get a paper published in machine learning now it’s got to have a table in it, with all these different data sets across the top, and all these different methods along the side, and your method has to look like the best one.” [↵](#)

10. ¹⁰ Of note, this is an important thrust with relatively little public sector leadership. In the US, the Congressional Office of Technology Impact Assessment was closed in 1995, an early victory of the anti-science movement. [↵](#)

11. ¹¹ This statement should not be taken out of context; Hinton frequently shares his perspective on issues concerning AI’s role in society. [↵](#)

12. ¹² Google decided not to bid on DoD’s landmark cloud computing initiative called the Joint Enterprise Defense Infrastructure (JEDI). [↵](#)

13. ¹³ Out of scope for this piece but helpful for some readers is the insight that the protest was based on a misunderstanding of the nature of Project Maven. This effort was about analyzing video to enhance situational awareness of things such as the placement or origin of vehicle-borne improvised explosive devices (IEDs) or the presence of hostile elements with automatic guns approaching environments such as schools or hospitals. Protestors concluded incorrectly that the project involved facial recognition for the purpose of drone-based killing. [↵](#)