ARTIFICIAL INTELLIGENCE AND THE GOOD SOCIETY

The Search for New Metrics, Governance and Philosophical Perspective

David Bollier
Rapporteur

A Report on the Third Annual Aspen Institute
Roundtable on Artificial Intelligence

THE ASPEN INSTITUTE

Communications and Society Program
Charles M. Firestone
Executive Director
Washington, D.C.
2019

To purchase additional copies of this report, please contact:

The Aspen Institute
Publications Office
P.O. Box 222
2014 Carmichael Road
Queenstown, Maryland 21658
Phone: (410) 820-5326
Fax: (410) 827-9174
E-mail: publications@aspeninstitute.org

For all other inquiries, please contact:

The Aspen Institute
Communications and Society Program
2300 N Street, NW
Suite 700
Washington, DC 20037
Phone: (202) 736-5818

The Aspen Institute
2300 N Street, NW
Suite 700
Washington, DC 20037

Published in the United States of America in 2019
by The Aspen Institute

Printed in the United States of America

19/007

Contents

This report is written from the perspective of an informed observer at the Aspen Institute Roundtable on Artificial Intelligence. Unless attributed to a particular person, none of the comments or ideas contained in this report should be taken as embodying the views or carrying the endorsement of any specific participant at the Roundtable.

Foreword

The suite of computing techniques commonly referred to as artificial intelligence has had and will continue to have profound affects on our society. AI's benefits not only include increased efficiencies across societal sectors but also a transformational change in knowledge generation, communication and personalized experiences. At the same time, these advances can have counterweights in certain uses, unintended consequences, or control by bad actors. This includes the potential to disrupt fundamental societal values and norms as well as exacerbate existing systemic issues such as inequality and inequity. Balancing AI innovation against these potential harms is both critical and necessary for the future of human progress.

In February 2019, the Aspen Institute Communications and Society Program convened twenty-four leaders across industry, academia and civil society to begin chipping away at this vexing challenge. Participants of the Roundtable on Artificial Intelligence engaged in extensive and meaningful dialogue around the need for a coherent philosophical approach towards AI, for both quantitative and qualitative metrics and for new systems of governance and accountability.

The following report, "Artificial Intelligence and the Good Society: The Search for New Metrics, Governance and Philosophical Perspective," authored by David Bollier, reflects on these discussions and debates. It highlights the promises and perils of AI systems and captures a robust debate regarding the proper methods for measuring progress or setbacks.

The report is divided into four sections. First, "Moonshot Visions of AI," features the enormous power, speed and scale of AI systems to positively impact our daily lives, specifically in healthcare and employment. "The Perils of Artificial Intelligence" then lays out numerous serious risks and limitations of these systems, ranging from embedded bias to a lack of public understanding of AI.

The second-half of the report shifts focus to suggest two cornerstone issues for the future of AI. "Toward A Philosophy of AI Design and Governance" articulates the need for a cohesive values-driven or philosophical AI approach to better assess its impact on society. The issue,

addressed in "Envisioning New Metrics, Governance and Accountability for AI," is then to devise the appropriate evaluation metrics and governance mechanisms to steer AI in socially constructive directions.

At the end, whether the need is to enlist community review boards to provide oversight or to adopt certain metrics for public accountability, it is clear that there will be no single, universal solution. Instead, just as the technology itself is a suite of multiple computing techniques, there are multiple approaches to steer AI uses towards the good society. Ideally, this volume will help readers understand both the compass and maps to chart our way forward.

Acknowledgments

On behalf of the Aspen Institute Communications and Society program, I want to thank the Patrick J. McGovern Foundation for their support and leadership in developing this roundtable. Thanks, also, to David Bollier, our rapporteur, for capturing the various dialogues, debates and nuanced viewpoints of participants. As typical of our roundtables, this report is the rapporteur's distillation of the dialogue. It does not necessarily reflect the opinion of each participant in the meeting. Finally, I want to thank Sarah Eppehimer, Project Director; Dr. Kristine Gloria, Senior Project Manager; and Tricia Kelly, Managing Director, for their work on the conference and bringing this report to fruition.

Charles M. Firestone
Executive Director
Communications and Society Program
The Aspen Institute
June 2019

Artificial Intelligence and the Good Society

David Bollier

Artificial Intelligence and the Good Society

The Search for New Metrics, Governance and Philosophical Perspective

David Bollier

Introduction

The development of Artificial Intelligence, or AI, is surging ahead at remarkable speeds, promising to bring dazzling new breakthroughs and efficiencies to medical treatment, public infrastructure, workplaces, education, households and everyday life. Yet at the same time, respected observers point with alarm to the grave disruptions and risks that such powerful technologies may entail. Various AI systems are likely to undercut some fundamental premises and infrastructures of modern life. This includes potentially jeopardizing the physical safety of people, the integrity of democratic governance and culture, and expectations of economic opportunity, privacy and fairness, among other social values.

Given the power of AI technologies and uncertainties about their impact, perspectives on the future tend to be polarized or at least ambivalent. There is excitement at the enormous benefits that AI systems might yield—"the prize is very big"—and also legitimate fears about the unintended, far-reaching consequences of technologies that are still in early stages of development.

Addressing the many questions about AI is a difficult empirical issue, however. The field of research and development is sprawling and diverse, and much knowledge is regarded as proprietary or subject to state secrecy. Compounding these problems, existing philosophical frameworks have trouble assessing the ethical, social and political ramifications of many AI systems, in part because their impact is likely to be so transformational. In a piece for *The New York Times*, Taiwanese technologist and venture capitalist, Kai-Fu Lee predicts that the coming AI revolution "will disrupt the structure of our economic and political systems," and will provoke "an AI-driven crisis of jobs, inequality and meaning."[1]

Questions abound. Will AI systems largely extend the historic dynamics of modern capitalism to more people through economic growth and innovation? Or will they disrupt the economy and societal systems in dangerous, destabilizing ways—for example, by overriding traditional structures that assure individual freedom, privacy and democratic sovereignty? Do AI systems tend to strengthen authoritarian, centralized control, as can be seen in the surveillance and control of citizens in China? Or is this simply one manifestation of a broader spectrum of possibilities? Setting aside the larger geopolitical and economic questions, there remain many questions about how AI will affect American society, especially government, politics and culture.

To address the many concerns raised by AI, the Aspen Institute Communications and Society Program convened twenty-four leading entrepreneurs, academics, technologists, philanthropists, educators, law scholars and other AI thinkers for a conference in Santa Barbara, California, on February 11-13, 2019. The gathering sought to bring some focused intelligence and expertise from diverse perspectives to consider the promise and perils of AI, especially over the next decade.

Special attention was paid to the transformative "moonshot" possibilities that AI could enable, and to general scenarios in which AI could remake healthcare and employment. Discussion also focused on helpful changes that could be made in education and public understanding to facilitate the development of AI. A major portion of the conference dealt with the need for a more coherent philosophical approach to developing AI and for devising effective new systems of measurement, governance and public accountability.

The two days of discussion were moderated by Charles M. Firestone, Executive Director of the Communications and Society Program. This report, an interpretive synthesis of the most salient themes discussed, was written by rapporteur David Bollier.

Moonshot Visions of AI

In his opening presentation, Reid Hoffman, Co-founder of LinkedIn and Partner at Greylock Partners, described himself as a "techno-optimist, not a techno-utopian," who sees AI as providing "an amplifier effect [in today's economy] equivalent to the Industrial Revolution." This is mostly a matter of exponential increases in productivity, he said. Hoffman envisions AI bringing new technologies to scale very

rapidly, improving the practice of medicine through cheaper and better diagnoses and treatments, and rejuvenating rural economies where manufacturing has left.

"Amazon's robotized facilities actually employ more humans than they did before the introduction of robots," he said. "There is a higher output of goods per number of humans, so there are still productivity gains, but as more things have been automated, more humans are brought in to do a lot of other things." Hoffman sees this dynamic extending well beyond the workplace: "If we start thinking about unreachable areas of the world, from ocean floors to the moon and Mars, robots are going to be an important part of that, whether it's manufacturing new materials or developing new places to live."

Of course, this vision of the future will be accompanied by all sorts of dangers and bad actors, Hoffman noted. "Humans are a fractious lot," and so there will be plenty of competitive and inhumane behaviors associated with AI. But in the end, AI may be "the only answer" to such problems and dangers—bioterrorism or fake news on social media, for example—because of the technology's unparalleled speed in diagnosing, measuring and reacting.

AI is *many* things, not a single thing.

These technical capabilities cannot help us with many social and political questions, however, such as "Who has the power to make decisions, how are rewards distributed, whether control should be centralized or decentralized, and questions of human freedom, autonomy, and privacy." These are inescapable questions. But in thinking about AI, said Hoffman, it is important to remember that "the prize is large." Hoffman cited one of his favorite books, *Nonzero*, by Robert Wright, which explores the concept of "non-zero sum" in game theory and the idea of developing new, more complex social systems that move us beyond "zero-sum" tradeoff scenarios. "How do we take these technologies and construct non-zero games of the future?" Hoffman asked.

An early observation in the conference highlighted a point about AI that needs to be kept foremost in mind: AI is *many things*, not a single thing. Gary Marcus, Professor of Psychology and Neural Science at

New York University, said, "People talk about AI in generic terms, but it's not one magical technique. Our hopes and fears all rest on *what kind* of AI we're talking about." While one type of AI in a given context may pose serious risks—deep learning systems in driverless cars that misidentify objects, for example—*particular* AI risks are likely to be very situational and technology-specific. This is an important caveat to keep in mind when having more general discussions. Deep learning is a subset of machine-learning, and both are forms of AI—but there are many forms of AI that have nothing to do with machine-learning of any sort. Differentiating types of AI matters.

The real game-changer is the capacity of machines themselves to learn and incorporate new insight into their functionality…this represents a great leap forward over historic modes of computing that were more or less static and repetitious.

Another meta-issue that will affect our perception of AI technologies going forward is the role of language and culture in making things visible—or invisible. As the technology gets faster, comes online, and become pervasive, the term 'AI' may become obsolete, similar to the way in which the term "cell" in "cell phones" has been largely abandoned. AI will become more ubiquitous and many see this moment as the opportunity to set the right principles, concepts and criteria for talking about it.

In general, AI technologies have such enormous power and versatility that they offer the possibility of enormous benefits to humankind. The real game-changer is the capacity of machines themselves to learn and incorporate new insight into their functionality. Thus, instead of relying on the fallible and limited human brain to detect patterns within large repositories of data, AI algorithms can sift through mountains of data and emerge with potentially actionable insights. This represents a great leap forward over historic modes of computing that were more or less static and repetitious.

So, what are some of the "moonshots" that AI might tackle? There are many. Gary Marcus, Professor and author of *Rebooting AI*, believes

that brain neuroscience should be a prime target. "The human brain has 80 billion neurons, connections between all of them, and a thousand proteins at every synapse," said Marcus. "The sheer mass of data is something that no human can understand." But, "We could use AI to understand the brain better," he said, "and that would help us build better AI."

…"combining the causal reasoning of human beings with the sheer computation of computers. That would be revolutionary." ***– Gary Marcus***

Such applications of AI could open up new revolutions in medical science more generally, said Marcus, where discoveries in certain fields have been stalled for decades. "We don't have any good new drugs for depression, for example, and we still don't have really great leads on Alzheimer's Disease," he said. There are so many proteins floating around and interacting in such complicated ways that we, as human individuals, cannot understand them. If we did AI well, however, it could help us come up with answers by "combining the causal reasoning of human beings with the sheer computation of computers. That would be revolutionary."

Neil Jacobstein, Chair of Artificial Intelligence & Robotics at Singularity University, believes that AI technologies have the capacity to "create much more wealth around the world than we used to have," and "vastly improve digital manufacturing with solar-driven, low environmental impact, open source processes." The surge of new wealth could allow societies to "grow bigger pies of material wealth rather than just squabble over how we divide up a fixed pie. However, solving the social problem of wealth distribution is as important as solving wealth generation. Otherwise, we may continue to see hyper-concentration of wealth." To hasten this process, he suggested several ideas: 1) Start a new competitive challenge for the utilization of AI, sponsored by different federal agencies; 2) Institute a system of high-quality, AI-powered educational courses available for free to anyone; and 3) use AI to build better infrastructure for "smart cities."

The ability of AI systems to learn and evolve will open up entirely new frontiers of science, said Steve Chien, Senior Research Scientist and Head of the Artificial Intelligence Group at the California Institute of Technology's Jet Propulsion Laboratory. In the areas in which Chien works, studying the environment, he sees AI helping to build more focused, precise, environmental modeling, such as in addressing natural hazards. "Flooding is the world's most dangerous natural hazard, affecting tens of millions of people and causing billions of USD damage every year," he said. AI systems that rely on larger datasets and machine learning could help improve the forecasting, mitigation and response processes. These benefits could apply to agriculture as well.

AI's capacity to analyze large datasets also has enormous potential benefits for evaluating decision-making processes. Terah Lyons, Executive Director of the Partnership on AI, a multi-stakeholder nonprofit dedicated to machine intelligence, envisions AI providing for measurement and continuous improvement of decision-making, especially in domains where audits of systemic decisions are rare, such as in healthcare and criminal justice. AI could be used to monitor "confidence estimates" for predictions, decision-making and accountability structures on a continuous basis. The goal would be more probing, reliable evaluations of outcomes than are generally possible today. This will also lead to more public accountability for decisions or patterns of decisions that otherwise may be challenging to understand, but may suffer from bias, injustice or negative structural dependencies. In this respect, AI takes on an enabling role that is both more subtle and context-embedded than the simple automation of human processes. AI can be used as a tool of inquiry to clarify the character of problems, direct our attention to the most salient facts, and improve critical judgments. For example, if AI systems could be used to analyze millions of data points about people's behaviors—say, eating habits or disease patterns—it could yield better public policy strategies.

Technology moonshots have historically been the province of governments, the only entities with sufficient institutional resources to organize and fund super-ambitious projects. Today, that has changed, noted Reid Hoffman. "We are now at this interesting place where businesses have the scale and capital needed for moonshots—hopefully to serve a social good." He said that tech businesses may have a better

"moonshot capability" than governments, at least within the AI context, despite the Trump Administration's announcement that it is going to do an AI moonshot as part of its "American First" policies.[2]

Given these realities, Hoffman believes that our best, most practical option is "to try to potentially shape any [corporate] moonshots in ways that have positive, inclusive impacts on the majority of humanity." He pointed to mechanisms such as new market incentives for AI developers, new accountability structures to influence corporate decisions, and leveraging trending public concerns to jawbone AI design and practices.

"There is an AI tech race underway…which political value system will be most embedded in AI systems and how they shape the future?"
– Reid Hoffman

Such considerations are obviously moot when it comes to authoritarian governments like China, which is aggressively developing AI to compete economically and rule over its population socially and politically.[3] "There is an AI tech race underway," said Hoffman, which is not just an economic competition but one for geopolitical dominance. A relevant question, he continued, is "which political value system will be most embedded in AI systems and how they shape the future?"

AI and Healthcare

In the course of the conference, breakout groups were asked to come up with general scenarios for how AI could plausibly improve life in three areas—healthcare, employment and governance. The scenarios were intended to build upon current trends and identify both positive and negative signposts as they might evolve over the next five to ten years.

The spokesperson for the healthcare group—Alix Lacoste, Vice President of Data Science at Benevolent AI, a firm that uses data and machine learning to advance biomedical discoveries—noted the need for more informed patients. "Patients are not really well-educated

[about their healthcare decisions] and do not necessarily participate in their care. So we would like to democratize expertise for both patients and providers. They don't seem to talk to each other that much about best care." Yet the group also wants to assure that patients' values and risk preferences are taken into account, so that specific treatments, resuscitation instructions, and the like, are available. "It's really about empowering patients to make better decisions," said Lacoste.

To achieve this, the healthcare group concluded that AI in this field should focus mostly on enhanced decision-making as it applies to diagnosis and treatment planning. The decision-making should reflect "explainable causal reasoning," said the group, by which it means that "AI should be a continuous learning system about how to provide best care. It should synthesize information, keep it up to date, and make it available, based on appropriate rights to information."

Meanwhile, Amazon, J.P. Morgan Chase and Berkshire Hathaway have come together to rethink the entire healthcare system for more than one million employees, on a nonprofit basis. This alone could have a catalyzing effect throughout American healthcare.[4] Tom Gruber, an AI product designer, said, "AI is giving us a new toolkit to think about playing a different kind of game. There are some new options on the table that may break the holds that we currently have."

The group identified a number of metrics to assess the positive impacts that AI could have on healthcare:

- increased discovery of new medicines
- increased access to public data for discovering medical treatments
- lower healthcare costs (via simulations of incentives in the healthcare system)
- healthier populations
- earlier diagnoses of diseases
- shifts toward preventative care
- increased efficiencies in hospitals while optimizing patient care, and
- decreased latency of information needed for care.

How might AI tools be used to improve the healthcare system as assessed by these metrics? The group made a number of suggestions.

- Internet of Things systems could continuously monitor and analyze data to diagnose and propose preventative treatments. They could also increase the frequency of sampling of data.
- An omnipresent AI system could be used to aggregate all of a person's health data and serve as a decision-support system. This function could be augmented by bringing medical literature into the analysis, providing new forms of patient and provider education.
- An AI agent could serve as a "health companion and assistant" by dispensing health advice and patient education.
- AI could also help with disaster relief by connecting affected hospitals and connecting patients with similar conditions.
- Compliance with medical instructions could be improved through greater use of specialized robots or even games.[5]
- A more ambitious goal might be to use AI systems to render hospitals less necessary except in extreme cases. After all, hospitals are expensive and dangerous places for patients. Could AI systems be used to keep ourselves sufficiently healthy that there would be less need for hospitals?

The healthcare group also identified a number of negative outcomes that should be tracked. It is possible that insurers and employers with extensive access to employees' health data could use AI to identify greater risks for individuals, prompting them to raise insurance rates or refuse employment; hence the necessity of strong privacy protections. Neil Jacobstein of Singularity University noted that machine learning can now use photographs of the retinal fundus of the eye to determine a person's age, sex, cardiovascular risk and whether they smoke. This is not just accomplishing something better, faster or cheaper than doctors; it is performing a task that ophthalmologists did not know was possible.

New York State recently passed a law allowing health insurers to use social media to set premiums.[6] This could have significant negative effects on who gets coverage and who doesn't, said Meredith Whittaker, Co-founder and Co-director of New York University's AI Now Institute. "I really worry about the 'datafication' of insurance and the very clear direction in which incentives are going," she said. In extreme cases, if machine learning were used to make highly granular calcula-

tions of individual risk, the very basis for health insurance—the pooling of risk—would be seriously undermined. This is one reason, among others, for assuring strong privacy protections.

Without global privacy safeguards, AI systems could easily be used to rifle through patients' medical data for all sorts of unauthorized commercial purposes. Or they could be used to detect mental health issues and order preventative measures without patient knowledge or consent. These possibilities point up the need for strong legal protections to assure that patient data is used only to enhance an individual's own or other people's care, with their informed consent.

Another negative signpost for AI development is the use of population statistics to make decisions about individual patients without evidence or causal reasoning. For example, a neural network algorithm looking at correlations in healthcare treatments once concluded that asthmatic pneumonia patients are actually at a *lower* risk of dying than other patients.[7] The algorithm based its judgment purely on correlations and not on evidence-based causality, and therefore it did not understand that such patients are immediately put into intensive care units, which is the real reason for their higher survival rate.

AI systems pose the risk of discriminatory, unequal treatment of people that might otherwise go undetected. Facebook has been accused of racial and gender profiling in its targeting of advertising for housing, for example. If AI were to make judgments about people based on their genetic makeup, or even help identify what genetic changes to implement in order to achieve desired traits, it could facilitate discrimination that would otherwise be illegal.

In some cases, unequal treatment could be a secondary effect caused by unequal access to healthcare data and bias in the data itself, either for individual patients or national populations. This could leave some segments of society without access to important medical information.

AI and Employment

A second breakout group offered a scenario that imagined how AI might affect employment in the coming decade and what might be done in response. The group made a number of assumptions about the future—that the primary societal goal is efficiency and equal opportu-

nity, not equal results. The group also assumed that 100% employment is not the goal, and indeed, that a net increase in jobs over the next ten to fifteen years is unlikely to materialize. "AI will create new job opportunities, but it may well also drive underemployment and unemployment. The *ratio* of new jobs to jobs displaced is key. The real issue is not just whether jobs are lost, but rather the pressure on middle-class incomes," said Neil Jacobstein, the spokesman for the group, citing a *Fortune* magazine special report on the shrinking middle class.[8] "We didn't assume that jobs are always the goal," he said, "because jobs are just one way to provide people with material well-being and a sense of purpose and self-esteem…. We asked what are the conditions that have to be met to help people flourish?"

What does "flourishing" mean with respect to AI? The group concluded that it means the ability of people to experience a sense of purpose and agency, whether through work, religion, art, sports or game, and to experience a sense of belonging to communities. Flourishing means access to material goods, knowledge, energy and healthcare, all of which could be more made efficient and affordable via AI (e.g., clean, inexpensive manufacturing; high-efficiency products; renewable energy; responsive healthcare systems).

As AI makes production more efficient, a big question is how to assure that the distribution of benefits can be fair and adequate to sustain households. If the ratio of jobs created to jobs displaced is unfavorable, the disappearance of millions of jobs could cause significant social unrest and a "lot of angry young people," said Jacobstein.

Other tech thinkers, such as Kai-Fu Lee, have called for a bold, aggressive government role in dealing with AI and job displacement. "We can't know the precise shape and speed of AI's impact on jobs, but the broader picture is clear. This will not be the normal churn of capitalism's creative destruction, a process that inevitably arrives at a new equilibrium of more jobs, higher wages and better quality of life for all. Many of the free-market's self-correction mechanisms will break down in an AI economy," he writes, warning of the risk of "a new caste system, split into a plutocratic AI elite and the powerless struggling masses."[9] As a corrective, Lee proposed what he calls a "Social Investment Stipend," in effect a government salary for those who provide care work, community service or educational instruction. "There

are a lot of New Deal-style possibilities that could be explored and put in place fairly quickly. But it is also prudent to start experimenting with various forms of Universal Basic Income (UBI) possibly at the city level rather than the national level," he said. Lee envisions a UBI with not just handouts to people, but one that sets some sort of means-test for eligibility and requires people to do something in exchange for their money.

The group envisioned opportunities in part-time or occasional jobs whose wages could be matched by the government. There could also be many new jobs available in sports and entertainment, and public service jobs that provide childcare, eldercare and environmental improvements.

"People would also have access to free, high-quality, Web-based education in almost any area they choose, which is now a very real possibility," said Jacobstein. It would also be important to have retraining and reskilling programs ramp up in a very short time-frame—something that AI tutors could assist.

How could such a system be financed? The Employment group posited "an abundance economy scenario in which the potential for a ten-fold increase in wealth generation is possible through the growth of 'open source almost everything.'" In such a world of super-robust wealth-generation, taxation would not pinch so much, especially for high-flying industries, and in general would be eminently affordable. Even if raising taxes is not seen as desirable, it would be far cheaper than dealing with the social chaos or crime that could otherwise result.

Since the rapid development of AI is accelerating, the time-coefficient for developing a variety of effective responses is absolutely critical, said Jacobstein. It makes sense to focus on cities and best-practices at the local level as a way to increase resilience. Mayors are more likely to be innovative, accountable agents of change than national governments.

The Perils of Artificial Intelligence

While much of the conference explored the transformational potential of AI, the group was mindful of its serious risks and limitations as well. Two significant perils have already been mentioned: the disruptive impact of AI on various industries and work life, and the structural intensification of social inequalities.

AI and Inequality

A number of prominent civil rights groups, tech and business leaders have raised sharp and even extreme warnings about the threats that AI poses for societies. Entrepreneur Elon Musk has famously called super-intelligent AI systems "the biggest risk we face as a civilization," comparing their creation to "summoning the devil." Even respected business commentators worry that AI could trigger social and economic upheavals. For example, Kevin Roose of *The New York Times* has called AI a tool by which the world's business executives will "transform their businesses into lean, digitalized, highly automated operations," creating new private concentrations of wealth at the expense of workers and the public.[10]

A 2013 report by the Aspen Institute Communications and Society Program explored this theme in depth, focusing on the rise of a "power-curve distribution" of wealth and income that are associated with network platforms. The term refers to a power-law distribution in which a small number of people reap a disproportionate share of benefits while the bulk of participants receive very modest gains. This so-called winner-take-most dynamic, or 80/20 rule (in which 20 percent of participants reap 80 percent of the gains) appears to be a structural feature of network-based activity because well-positioned business players are able to realize most of the productivity gains that materialize as economic "frictions" are radically reduced at an extremely rapid rate.[11] This is displacing middle-class jobs at an accelerated rate, leaving people reeling from the pace of change and government scrambling to solve market disruptions using archaic policy architectures.

In the report, Kim Taipale, Founder and Executive Director of the Stilwell Center for Advanced Studies, said that the paradoxical result of network effects is that "freedom results in inequality. That is, the more freedom there is in a system, the more unequal the outcomes." This stems in part from the self-reinforcing benefits that accrue to the "super-nodes" of a network, a phenomenon sometimes called "preferential attachment." Players that function as super-nodes capture a far disproportionate share of rewards relative to their effort, while hard-working smaller players and individuals find it very difficult (for structural reasons) to increase their share of benefits. Because of this dynamic, said Taipale, "The era of bell curve distributions that sup-

ported a bulging social middle class is over, and we are headed for the power-law distribution of economic opportunities. Education *per se* is not going to make up the difference."

Skeptics of this analysis respond that these outcomes are not inevitable. We must recognize that we are living through a period of historic economic transition, which will eventually result in greater prosperity, widely distributed, if the economy is allowed to pursue its course.

Embedded Biases in AI

Inequality is not just a result of network effects or the winner-take-most dynamic outlined above. It is sometimes embedded in the very algorithms and data that are used to drive AI. Deciding what information shall be collected in the first place amounts to a bias, one that might be amplified by biases in the sampling methods used. "As someone with a background in large-scale systems measurement, I have learned that data can tell you a lot about the world, but there is no unbiased data," said Meredith Whittaker of the AI Now Institute. "At some point you're making a methodological decision that says this data means this, and not that; that we're going to measure something this way and not that, based on whoever is in the room; and that this is the particular way we are going to represent it quantitatively."

"…data can tell you a lot about the world, but there is no unbiased data." ***– Meredith Whittaker***

Steve Chien of the Jet Propulsion Laboratory suggests that a similar problem arises when you build an autonomous system to track natural phenomena. In monitoring the weather, for example, "Instead of tracking the entire world, or all of the storms in the world, you want to be more selective and have smart measurement of that subset that best allows you to predict and model—such as a critical storm front—at a higher resolution. However, the challenges are (a) to figure out what are these key parts; and (b) the algorithms used to control any sensors are introducing bias into the data because you're not collecting world-

wide data on a 24/7 basis, you're only collecting a tiny subset based on what the algorithms tell you are the most important data." Scientific models may therefore be limited at the start by the implicit biases of these AI selective data acquisition algorithms and resulting data. Weather data, for example, may implicitly adopt the baseline of certain types of storms (e.g., ones with more ice and less wind) characteristic of certain regions of the world, and not others, if those are used to design the sensing (data acquisition) algorithms.

…we ought to see AI as a mirror that reflects our own limited perceptions and social and political biases. *– Meredith Whittaker*

Expanding upon this idea, scientist Gary Marcus of New York University stressed that AI currently is a very data-driven paradigm, one that may or may not reflect causal relationships in reality. "In machine learning, we tend to have algorithms that at some level try to mimic the data that they've seen before without having a deep understanding of the causal laws that generate those data. But that's a design choice that happens to be the easiest way to go right now; it's not the only way that we can build algorithms." Indeed, he continued, "It might be possible to build deeper models of human interaction that reflect the underlying psychology that causes the behavior." But that's not possible now, which is why AI designers must be mindful of the causal models implicit in their algorithms and data. Deeper models for AI might emulate the process of language acquisition in children, said Marcus: "Children are not direct slaves to the data. When they learn language, for example, they're learning an abstract grammar that they can use in a lot of different circumstances, so they don't have to exactly mimic everything that their parents say. But the algorithms that we have right now are very much blind mimics, and that accounts for some of the bias."

Meredith Whittaker suggested that we ought to see AI as a mirror that reflects our own limited perceptions and social and political biases. Instead of using AI systems uncritically as simple "solutions" to

problems, we ought to see AI as a set of "diagnostic technologies" that can help reveal the embedded biases in processes, much as AI-driven recommendations for criminal sentencing help reveal racial and political biases in our judicial system.

It became clear in discussion that many of the conversations about AI focus on how it currently functions, while what we really need is a larger conversation about how AI *could be*. "We focus a tremendous amount on learning from data, but do almost nothing about how to build better causal inferencing," said John Seely Brown, Independent Co-Chairman of the Deloitte Center for the Edge. "Causal inferencing is complicated and can take a lot of time to disentangle the underlying contexts and to explore recursive trails of 'whys' and we don't yet have good enough techniques," he said, "but I think we're beginning to see a real shift in some fields toward building stronger causal models into our systems." Brown suggested that we need to take abduction—a logical and plausible argument stemming from the major premise—more seriously than limiting ourselves to just deductions, for example. We also need to find ways to mimic the process of learning, which sometimes requires constructing new types of causal models.

Alix Lacoste of BenevolentAI agreed that AI must find better ways to "marry the data-driven with a hypothesis-driven way of finding novel insights," and to navigate the gap between bias and expertise. One useful strategy would be to use AI-driven "attention mechanisms" to make certain patterns of data more salient to machine learning algorithms, but ensure that human expertise is then called upon to render more refined, subtle judgments. In this fashion, attention mechanisms can serve two objectives: one is to help bring the most relevant information forward; the other to help with interpretability by leveraging experts to review what the machine learning models decide to pay attention to.

Peter Norvig, Director of Research at Google, noted that there is a famous paper in the computer science literature, "Attention Is All You Need,"[12] which argues that attention mechanisms are an effective way to elicit insights from complicated, deep learning networks. The term "attention mechanism" does not necessarily relate to the ways that humans direct their attention, but rather to how computing systems can take account of relational factors in a given context (e.g., other words or phenomena) to provide guidance to humans. Computer sci-

entists continue to debate which attention mechanism or process may be the best way to achieve good results.

One problem in this entire debate about causality may be that we have narrowed our understanding of AI to machine-learning, said David Ferrucci, Founder and CEO of Elemental Cognition. "I don't think that's right or, in fact, good. To be provocative, I would say that human intelligence is not data-driven. It's rational, collaborative and communicative—and none of these is data-driven. Perhaps human stupidity and savantism are data-driven, but human intelligence? Decidedly not," he said.

"AI should ultimately help us understand stuff, but we have to realize that understanding is hard." ***– David Ferrucci***

"So when we think about what we want out of AI," Ferrucci continued, "we have to think about how AI ultimately can be communicative and collaborative with human beings, and engage our thought processes and the ways that we think and communicate. AI should ultimately help us understand stuff, but we have to realize that understanding is hard. We cannot understand causal relations between real-world phenomena by statistically analyzing data that may only weakly reflect the underlying phenomena. It's one thing for Google to bring us gazillions of pages; it's another thing to understand what they say and how to synthesize that knowledge."

One apparent general solution to this challenge is to make sure that we "keep humans in the loop" for any process involving AI, as one participant noted. But Meredith Whittaker hastened to point out that the idea of "autonomous learning" is a bit misleading in the first place. Deep learning systems generally rely on vast quantities of labeled data, in any case, she said, "which requires very low-paid people to label and classify data. So we need to keep in mind the full stack on which AI is being built, and keep in mind which humans AI is serving. The term 'autonomous' or 'automation' often hides a system's levels of dependency on precarious labor, and writes out the experiences of a lot of humanity that is enmeshed in these systems."

Educational Barriers to AI Careers

The "hidden labor" that sometimes play a part in AI points to a related problem—the lack of diversity among AI designers and business people, and the limited educational opportunities for women, people of color and disadvantaged students. These issues matter not just as a matter of social equity, but as factors that shape the very design and deployment of AI. As seen in recent controversies about the racial limitation of facial recognition systems, the identities of AI business strategists, engineers and researchers can greatly affect the performance and social character of AI.

"I think the real AI moonshot is inclusion. We need to try to bring underrepresented populations into AI." – *Tess Posner*

"Fewer than 23 percent of people working on AI globally[13] these days are women, and the number for other underrepresented groups are abysmal. It's truly a crisis," said Tess Posner, CEO of AI4ALL. "I think the real AI moonshot is inclusion. We need to try to bring underrepresented populations into AI." This is why Posner is working with high school students to try to interest women, minorities and low-income kids in computer science and AI. Posner stressed that tracking diversity in these fields requires holistic interventions to address access, hiring, retention and growth of diverse teams and education needs to begin at the high school and middle-school levels, and even in elementary schools—well before post-secondary education.

"I feel as if the target age is really middle-school, and sometimes even earlier," said De'Aira Bryant, a second-year doctoral student at the School of Interactive Computing at the Georgia Institute of Technology. "That's when kids are figuring out what their favorite subjects are, and thinking about potential careers. By ninth and tenth grades, students are already in the preparatory classes for specific disciplines, so it might be a bit late by then."

Anita LaFrance Allen, the Vice Provost for Faculty and Professor of Law at the University of Pennsylvania, echoed these concerns as a university administrator. She noted that, "At the moment, we just have a very, very non-diverse set of technical trainees and people producing technical trainees. About 33 percent of our tenure-track faculty are women, and in the engineering school, only 17 percent. We have one black engineering professor and one black math professor."

Bryant noted the distinct lack of access to computer science in the small town in which she grew up, Estill, South Carolina. "My high school still doesn't offer computer science and very few of the schools in a twenty- or thirty-mile radius in the Low Country do. That's why I've been working with the South Carolina legislature to try to get CS [computer science] in every school in South Carolina, at least at the high school level. That's just CS, not even AI."

Across the state, there are very few teachers capable of teaching AP classes in computer science, in part because of the scarce resources and funding to retain qualified teachers. "There is a huge lack of women in AP CS programs across the state," said Bryant, "even in top schools in Greenville, Columbia and Charleston. And at the University of South Carolina, the AI course is an elective and at Georgia Tech, the AI courses on campus are always full." Bryant believes that getting the educational system more involved and reformed will be crucial to addressing the diversity problems that Posner mentioned.

Several participants agreed that a "revolution" is needed to improve access to CS and AI programs and, in turn, increase the diversity of trained AI graduates and engineers. Reforms must be attempted at middle-schools with outreach programs, on up to graduate programs, said Alix Lacoste of Benevolent AI. Lacoste continued, "We need to fully integrate diversity and inclusion into our recruitment process, and educate the general public and media."

To counter the stigma that is sometime associated with computer science and AI, Bryant suggested "using 'gamification' and other techniques that have been successful in keeping kids interested." John Seely Brown agreed, noting how the hands-on, participatory approach introduced by remix culture "really opened kids' minds about how to use media in new ways." If that sensibility could be brought to AI, it could really expand interest and engagement among young people. "You

would be shocked at what kids can figure out," said Brown. "If you can create a set of tools in the right kind of 'club,' it's amazing the kind of learning that can result." A prime example is the Hacker Dojo in Santa Clara, California, an open working space for software projects.

Tess Posner described a new effort by AI4ALL to improve access to basic AI education, through the AI4ALL Open Learning Program: "We've found that online education doesn't work, especially if you're trying to be inclusive for all types of learners and people who may not already be involved in tech," she said. "You need a peer learning community and adult follow-on support. A static, online program only works if you already have other resources built into the system, such as caring adults and peer groups and a sense of inclusive belonging." Posner said that AI4ALL's program can fit into existing STEM programs, high schools and community based organizations because it works as a kind of "guide on the side" teaching model that is enough to facilitate learning in peer group situations.[14]

But when a suggestion was made to provide federal credits or incentive to take AI-related online courses, Anita LaFrance Allen was skeptical suggesting that by providing more incentives for universities to create more online education would simply cater to their desire to make money, resulting in many low-quality courses. This is not to say that the federal government should play no role, Allen hastened to add, "The federal government has a symbiotic relationship with higher education, with funding coming from the National Institutes of Health, the Department of Defense, the Department of Energy, and so forth." This relationship needs to be leveraged to produce greater learning about AI and greater diversity of students and graduates, she said. At least one participant reacted by saying that "the hair on the back of my neck goes up when we talk about federal solutions, particularly at this moment. I think we could get better and quicker results from cities and states." Another participant suggested that philanthropy could play a particularly useful role in improving CS and AI education.

Meredith Whittaker reminded the conference that unequal opportunity is not confined to education; it is a problem within the tech industry itself: "We need to acknowledge that the cultures into which we are sending these folks have their own problems in terms of pay,

transparency, opportunity and social equity. Black women have the highest attrition rate as employees at Google, according to its public diversity report. People interested in AI careers are pushed out at every stage of a leaky educational pipeline, and then once they reach the premier academic labs and companies, they are pushed out in different ways. This is something we need to look at closely."

Public Understanding about AI

There was a general consensus that changing some of the embedded biases in AI and computer science education will require changes in public awareness. While many technophiles are thrilled at the coming applications of AI, other people are confused or fearful. Others, meanwhile, feel abandoned or victimized by "the system," said Father Eric Salobir, a Roman Catholic priest and President of OPTIC, a network that explores the field of digital humanities. He said that such people may feel wary about new technologies and whether they will personally benefit from them, or be exploited by them.

Measuring and changing public opinion on AI is likely to be difficult, however, because the technology is not a retail product used by ordinary people; it is usually a centrally managed technology invisibly embedded in other systems. Not surprisingly, the public is often ill-informed about the potential uses and risks of AI technologies.

Mainstream media coverage does not help this situation. A great deal of news stories and commentary about AI tends to be either sensationalist horror stories or in effect press releases from industry itself. In a survey of UK media coverage of AI, nearly 60 percent of stories were indexed to industry-driven news about products, announcements and research, according to the Reuters Institute for Study of Journalism. Coverage of AI as an emerging public issue was relatively modest, and non-industry voices such as academics, activists, politicians and civil servants were heard less often than industry sources.

…we need to pay more attention to the "cultural narrative" about AI. – ***Raina Kumra Gardiner***

Raina Kumra Gardiner, Director of the Omidyar Network, believes that we need to pay more attention to the "cultural narrative" about AI. "We have *Terminator* and *Black Mirror*," she said. Participants also suggested that foundations could play a constructive role in supporting the development of positive cultural narratives.

Toward a Philosophy of AI Design and Governance

In attempt to move beyond critique to problem-solving, this conference hosted a larger conversation about two issues that may be critical to the future of AI: 1) developing more astute, coherent philosophical approaches for assessing the design and social and political impacts of AI; and 2) devising new evaluation metrics and governance systems to assure that AI systems will be accountable and deliver positive outcomes.

Every major shift in economic history has been accompanied by step-changes in the philosophical and economic frameworks for understanding the world, one participant pointed out. Therefore, the questions we need to face are not simply a matter of "How can AI solve a given problem," but rather, "How can we develop a richer, larger and more appropriate understanding of the new situations that AI technologies engender?" Addressing this question forces us to consider what sorts of new social and political institutions, working metrics and policy architectures may be needed. AI is not just a technical domain, after all, but a set of tools that has the ambitions and capacity to design new worlds. Can we therefore develop a philosophy, or at least better working understandings, commensurate with the powers of the technology? This necessarily requires us to revisit first principles and fundamental ethical and perhaps religious notions of what a human being is and should be.

The "Executive Compass" as a Tool for Building a Good Society

As a first step in grappling with the clash of values that invariably arises when making decisions, the conference considered an ethical tool known as the Executive Compass, which was developed by business school professor James O'Toole in the 1990s for the Aspen Executive Seminar, a values-based leadership project started by philosopher Mortimer Adler. The Compass is an attempt to distill some of the complexities of modern philosophy into a simpler, practical tool for

thinking about the tensions among values and how to resolve them.[16] The Compass consists of four primary poles, with two sets of values in fundamental tension with each other. One axis counterpoises the value of *liberty* with that of *equality*, while the other axis sees the value of *community* in tension with *efficiency*. Ultimately, O'Toole sees each of the values potentially in tension with each of the others.

O'Toole regards these four polar forces as "tugging at all modern polities." Indeed, he writes, "The tensions among those values have provided the drama to political life in the West since the time of Hobbes. In particular, the choice between liberty and equality is said to be the most fundamental, and inescapable, of all the trade-offs facing society." To illustrate his point, O'Toole invokes the conflicting values of Alexander Hamilton in favoring "economic growth and technological advance," as opposed to the priorities of Thomas Jefferson whose primary concern was "communitarian values such as the quality of life."

A key goal of the Executive Compass is to make certain tensions among values more explicit, thereby triggering deeper discussion about what values are really important in a given situation. It is also meant to help identify and guide acceptable tradeoffs that an organization or society might make, so that a happy blending of value-priorities can be made. At the very least, however, the Compass seeks to make a productive discussion possible among people who disagree, helping them to understand the other's point of view.

In terms of its application to AI and its future, the Compass could help situate various value commitments within a larger framework and identify fundamental tensions that must be addressed. For example, the manifesto of Theodore Kaczynski, the Unabomber—"Industrial Society and Its Future," a reading for the conference—makes quite clear that Kaczynski prioritized liberty over everything else—to the extreme. The concerns for equality explored in the 2013 Aspen report on the "power-curve society" are juxtaposed against business interests in efficiency.

While the Compass may be useful as a point of entry to discussion, participants found it a limited tool. Marc Rotenberg, President of the Electronic Privacy Information Center (EPIC), said that "real-world interactions are more complicated than economic 'indifference curves,'" and that tensions in the political economy are dynamic, not

static. Follow-on action in complex systems are likely to be complicated and unpredictable. Rotenberg also questioned whether the two axes of values necessarily apply to a situation. In some societies such as China, AI appears to be creating a world in which there is neither liberty nor equality; that axis seems irrelevant.

...the more meaningful river of AI's future is the political economy, involving the companies that are developing AI, not just government.
– Michael Chui

Meredith Whittaker stressed that any consideration of values must address distributional questions: "Whose liberty? Whose efficiency? These contextual, historical frameworks matter," she said. Others agreed that it is important to bring other stakeholders into any value analysis. This discussion underscores the importance of *trust* as another core, independent value beyond the primary four. One party's liberty may enable it to amass great power and wealth, causing social distrust that unfettered markets are not likely to address. Even though Facebook has been plagued by many high-profile scandals involving user trust, for example, its stock prices have not suffered, noted one participant.

While the Compass aspires to help government or business leadership come to better value-based decisions, some participants questioned the apparent premise that centralized sources of power can be effective nowadays. "To me, the Compass assumes that government both understands and has power over technology," said Terah Lyons, Executive Director of the Partnership on AI. "But in our version of democracy, government isn't necessarily the arbiter of these conditions any longer." One statistic from 2016 makes this alarmingly clear, she said: U.S. government investment in AI in unclassified settings currently amounts to one-eighth of the amount invested by the top five companies operating in the AI ecosystem. "Public policy is not really in the driver's seat," said Lyons—a fact that is underscored by the relatively slow pace of government, law and policymaking. Michael Chui,

Partner at the McKinsey Global Institute, agreed: the more meaningful river of AI's future is the political economy, involving the companies that are developing AI, not just government.

The Relevance of Philosophy for AI

One takeaway from this discussion was the need for more serious philosophical reflection and debate about the design and deployment of AI technologies. "What I find really interesting," said Anita LaFrance Allen of the University of Pennsylvania, "is how few big visions are being created by intellectuals for the kind of world we'd like to see exist. Philosophers have been pretty silent about AI and the digital world. Because of that, there is something missing in our discussions. To me, that's a sad loss. We've got Mill in the utilitarian tradition, standing for voice, and Montesquieu in the Aristotelian tradition, and Kant standing for the Enlightenment. But these traditions don't take us far enough. We've got to go deeper, and force ourselves and our colleagues in relevant disciplines—political science, philosophy, law—to help us mine more deeply. We shouldn't abandon the canonical ideas, but take them forward."

There is a singular lack of "Big Think" about the social and human implications of AI. *– Vilas Dhar*

There is a singular lack of "Big Think" about the social and human implications of AI, said Vilas Dhar, a Trustee for the Patrick J. McGovern Foundation:

> Humanity has lost sight of a vision of what's possible. The people in the forefront of AI are not just defining the iterative technological process, they are in control of the massive social changes that come with it. There is almost no time spent on the fundamental questions of how you design the system. How do you break it apart and build it back up? What are the philosophical and political thoughts behind it? Our lawmakers are no longer equipped to ask these questions because of the

> increasing technical and moral complexity of these topics. So, the mantle falls to people in rooms like this one, and let's hope they don't fall victim to the master-of-the-universe syndrome.

Many participants agreed with this general sentiment. Neil Jacobstein of Singularity University said the historical divide between the sciences and humanities is causing friction, and limiting our imaginations, adding that "humanity is estranged from its authentic possibilities." He said interdisciplinary thinking is one reason why DARPA, the Defense Advanced Research Projects Agency, is funding "world models" and other programs that seek to integrate causal modeling with the deep learning AI algorithm.[17] An interdisciplinary approach to the development could help AI engineers deal with technical and design challenges, and it could also illuminate social and political implications and enable us to deal with them pro-actively.

"I sometimes wonder if we realize that we're going to have to invent a new kind of literacy. For the past hundred years or so, we've focused on content, but as we move into a world in which *context* matters – culture, history, economics, politics – we don't have very good ways to honor context…." ***– John Seely Brown***

At the design level, for instance, machine learning needs to be able to "read context," said John Seely Brown. "I sometimes wonder if we realize that we're going to have to invent a new kind of literacy. For the past hundred years or so, we've focused on *content*, but as we move into a world in which *context* matters—culture, history, economics, politics—we don't have very good ways to honor context. Yet so much sense of agency has to do with being able to read the context."

Seen from this perspective, designing AI involves some profound "epistemic challenges," said Brown. We are currently locked into notions of "optimization, which usually implies a form of reduction-

ism. But most problems have externalities; they can't be separated from their context. So we need new tools to unpack some of those externalities, each of which is entangled with others in very powerful ways. In some sense, our real challenge is how to disentangle a profoundly entangled system." Brown pointed out that public policy tends to put problems "in a box," as if core issues can be dealt with in isolation and through optimization strategies. But problems have a tendency to leak out of those boxes so fast, said Brown. "We kid ourselves into thinking we've solved a problem, when in fact we are making a bigger mess of the world."

Meredith Whittaker ascribed this problem to teams of narrowly focused tech experts making decisions, rather than cross-disciplinary teams. "We have tech people making quantified, reductionist determinations for domains without drawing upon the expertise of people in those very fields," she said. One example is Epic electronic medical records that may or may not reflect the full medical dimensions of a patient that a nurse identifies, but instead the crude taxonomy of billing codes. A similar reductionist logic can be seen in IBM's Watson for Oncology supercomputer, marketed as a superior way to make cancer diagnoses and treatment, said Whittaker. The AI system was aggressively marketed as a superior tool for cancer diagnoses and treatment, but its actual capabilities were quite limited, according to the medical publication *STAT*.[18] These types of stories point to the need for stronger interdisciplinary work on AI, said Whittaker, and for greater sensitivity to context and the philosophical assumptions behind AI design.

Neil Jacobstein noted that this same sort of thinking—solving specific and narrow problems without regard for context—prevailed among agricultural/biotech companies in the 1960s. They did not really think much about the second- and third-order consequences of pesticides on ecosystems. He said that a useful corrective to this kind of thinking can be found in a seminal 1971 essay by systems scientist Jay Forrester on the counterintuitive behaviors of social systems.[19] Jacobstein suggested that "we could better understand context and second- and third-order consequences if we combined pattern recognition, modeling and simulation."

Structural Imperatives Driving AI Development

The discussion about the importance of context spurred a broad conversation about structural and institutional imperatives driving AI design and deployment. Some observers worry that AI's enormous efficiencies, capacity for continuous learning, and reliance on centralized repositories of data make it a perfect tool for autocrats and authoritarians. This theme was previewed in a reading, "Why Technology Favors Tyranny," in *The Atlantic*,[20] in which author Yuval Noah Harari explains how AI has the potential to empower dictatorships:

> We tend to think about the conflict between democracy and dictatorship as a conflict between two different ethical systems, but it is actually a conflict between two different data-processing systems. Democracy distributes the power to process information and make decisions among many people and institutions, whereas dictatorship concentrates information and power in one place. Given 20th century technology, it was inefficient to concentrate too much information and power in one place…. However, artificial intelligence may soon swing the pendulum in the opposite direction. AI makes it possible to process enormous amounts of information centrally. In fact, it might make centralized systems far more efficient than diffuse systems, because machine learning works better when the machine has more information to analyze.

Tim Hwang, Director of the Ethics and Governance of AI Initiative, a joint project of the MIT Media Lab and the Harvard Berkman-Klein Center, agreed with this general analysis. One could easily make a "strong techno-determinist argument" that AI favors autocrats, said Hwang, because it takes a lot of money and institutional power to build large compute centers and acquire the massive quantities of data needed. Only a few large tech companies such as Apple, Google and Facebook—and the Chinese government—control sufficiently large quantities of personal data. Centralized players, whether autocrats or big companies, have strong motivations to build AI systems, for both surveillance and marketing purposes, and to use the psycho-social dynamics of online information. In recent years, there have been powerful efforts to manipulate voters through phony information sent to

precise demographic groups. The Russian Internet Research Agency (IRA) was able to influence the 2016 U.S. elections by targeting African Americans through online platforms.[21] As Harari notes, we may soon have to deal with "hordes of bots that know how to press our emotional buttons better than our mother does, and that use this uncanny ability, at the behest of a human elite, to try to sell us something—be it a car, a politician, or an entire ideology."

Hwang thinks that the control of chokeholds in AI infrastructure will be a key factor in whether centralized or decentralized control will prevail. To illustrate his point, he cited an historical comparison, commercial grain shipping in 19th century Chicago, as described in the book *Nature's Metropolis*.[22] A major economic and political transition occurred when the shipping of grain shifted from boats to railroads. "It turns out that once you shift to a railroad-based form of transportation, a relatively small number of people have a large amount of control over these markets," said Hwang. "I sometimes wonder, What is the 'railroad for AI?' And at what point do you implement certain types of infrastructure to assure that people have access to the technology?"

A related question, said Hwang, is, "Could you pull off the development of strong machine learning systems with a lot less data? If you could do that, suddenly the barriers to entry change quite a bit, which could shape up the potential for competition." Hwang believes this is the debate we need: "Can we actually achieve this goal in practice? The answer could influence whether or not there will be a technology lock-in or not in the future." This issue is important, said Vilas Dhar of the Patrick J. McGovern Foundation, because "AI may operate outside the boundaries of self-correcting behavior. The first-mover advantage may allow the aggregation of serious financial and technological resources, creating a threshold that prevents other people from being able to access the technology."

A number of participants expressed concern about the AI-driven concentrations of power to persuade and control. David Ferrucci, CEO of Elemental Cognition, noted, "It seems unfair when pockets of power have greater access to a given channel of persuasion than others, especially if that channel, powered by AI, is far more efficient in directing messages and persuading people than the conventional channels." Marc Rotenberg of EPIC put it more bluntly: "These systems do tend toward centralization and monopoly control."

What is so interesting, he added, is that computing in the Sixties and Seventies was largely centralized in large companies—and then the PC Revolution in the Eighties decentralized that power by pushing computing, applications and data out to individual consumers and businesses. Now we could be undergoing a "counterrevolution" that is re-aggregating computing power, he speculated. In any case, Rotenberg raises a tantalizing question: "Is there a current model under which AI authority could be genuinely distributed in the way that the early personal computer Revolution was? Is an alternative architecture viable?"

It was suggested that perhaps companies such as Amazon Web Services and the open source TensorFlow application represent a model for democratizing access to AI technology, despite ownership by a giant company. But Meredith Whittaker rejected that idea, pointing out that users of Amazon Web Services do not own the AI software or devices, nor are the systems easy to use.

However, for some business cases, this decision to rent-or-own on a cloud service may help offset costs for resources such as hardware for computing power, machine models and data. This model supports the idea of a "federated architecture" which allows for interoperability via a set of standards without having one central authority. Additional cloud service providers, such as Microsoft Azure, Google Cloud Platform, IBM Cloud, Salesforce cloud and others in the same market may even offer data sources at free or a low cost.

Amazon Web Services or federated technology is not necessarily helpful, said Patrick McGovern, Trustee of the Patrick J. McGovern Foundation, because of the sheer volume of data that you need—and control of data is only going to continue to get more consolidated. One outstanding question is whether companies who control this data will be good stewards of it.

For some participants, the future of AI development and decision-making will hinge upon whether we alter current structures of capitalism or not. Meredith Whittaker pointed out, "AI is controlled by a few large companies with the resources to build it. The technology is under the auspices of capitalist decision-making. If we are interested in applying AI to ends that would not be profitable, this is a political and deeply structural question. And so we would have to ask: What would be the incentives and mechanisms to drive that approach, and how would we do that in an ecology governed by the shareholder-value model?"

While this poses a formidable challenge, Whittaker thinks that now is a ripe moment to make a broader re-evaluation. She suggests that the rise of driverless cars should provoke this sort of questioning: "Are we going to take the individual car ownership society that was essentially architected by Henry Ford and city planners, and just automate it? I love the idea of using this moment to think about how we might actually change structures, and not simply automate or make more efficient the structures we already have."

This is a particularly vexing challenge, however, because AI investments are driving AI development and thus the scenarios for its use, said Tim Hwang: "The problems that AI will solve are going to be defined by what the AI toolkit is good at doing. And this reflects the particular types of investments being made." If there is relatively little interest in trying to make AI systems take account of context and causal inference (to harken back to the earlier discussion), that is because there are "much more profitable ways of developing the field," said Hwang. "The actual scope of AI technologies is therefore quite narrow, in ways that I think are counterproductive." Whittaker agreed with this assessment, adding that profitmaking generally favors goals that are easier to measure and short-term; qualitative goals that pay off over the longer-term and benefit broader constituencies are less likely to be attractive to businesses.

"One could argue that AI, on balance, has not been so great for society so far because a lot of it is just about ad placement and manipulating eyeballs," said Gary Marcus, the New York University professor. Marcus said that Google and Facebook are not likely to invest in causal inference modeling, for example, unless there were a short-term commercial advantage in doing so. And therefore, "It may just be that we won't get to paradise unless there is some other means for funding research for long-term priorities."

Envisioning New Metrics, Governance and Accountability for AI

It may take some time to develop richer philosophical approaches to AI. In the meantime, there are many actions that can be taken to guide AI in socially constructive directions and govern its development. Interestingly, many major tech companies are publicly asking the U.S.

government to regulate certain forms of AI, such as facial recognition, surveillance and driverless vehicles, or at least provide serious guidance to industry. These systems pose some serious dangers that need to be anticipated and prevented, especially if such things as weaponized drones and intrusive surveillance technologies become widely accessible. There are also significant national security and economic security issues at stake, as well as issues of civil rights, privacy and fair elections.

Precisely *how* the government or other parties should regulate specific AI technologies remains something of an open question, however. Participants considered a number of useful approaches for improving the governance and accountability of AI. A first priority is arguably the development of reliable metrics and empirical monitoring of salient developments in AI. But this would serve mostly as an informational predicate to new forms of legal oversight and regulation, perhaps involving novel strategies such as AI review boards, impact assessments and other independent mechanisms.

What Metrics Are Needed to Guide AI Development?

At the most basic level, the field of AI could benefit from some consensus metrics to assess what is actually happening with AI worldwide. Such metrics, in turn, could help various parties assess how current AI trends are achieving (or failing to achieve) key social, economic and educational goals. As a first order of business, then, it is important to consider what metrics need to be invented to determine that AI technologies are proceeding on the right track.

Michael Chui, Partner at the McKinsey Global Institute, gave a presentation about this challenge, suggesting different touchstones for evaluating AI. There is a truism in business, often attributed to Peter Drucker, that "What gets measured, gets managed." But Drucker actually did say that, "Working on the right things is what makes knowledge work effective. This is not capable of being measured by any of the yardsticks for manual work." In other words, not everything that we care about can necessarily be measured quantitatively (*see* Types of Measurements).

TYPES OF MEASUREMENTS

- *Quantitative vs. qualitative metrics.* Numbers can capture many important realities but they can also be reductionist and fail to capture certain values.
- *Metrics vs. milestones.* Metrics can make articulate progress at granular levels, but milestones can act as salient markers or triggers about whether things are going in the right direction.
- *Input metrics vs. output (or process vs. outcome) metrics.* Input metrics are equivalent to "intermediate goods" in economics, which differ from final products or outputs.
- *Domain-specific metrics or cross-domain metrics.* Some systems should be judged within a special domain while others have traversal ramifications.

The choice of metrics can be quite consequential because it elevates certain priorities while ignoring other potentially important signposts. So the question is not only what metrics should we have, but which ones should we eschew? While metrics can help focus energies and coordinate the work of organizations and societies, they can also become empty totems. Or as researcher Joe Edelman has written, "Once metrics are defined, they're like parasites and undead spirits, and they take over human beings"[23] by inducing slavish attention rather than critical inquiry.

One important collective of metrics about the state of AI today is the AI Index, published by Stanford University,[24] and translated into Chinese, Japanese and Korean. The Index is an effort "to track, collate, distil and visualize data relating to artificial intelligence," and aspires to be a comprehensive resource of data and analysis for policymakers, researchers, executives, journalists and the general public. Another body that collects some metrics on AI is the Association for the Advancement of AI.

There was general agreement that there should be metrics to document the positive influences of AI. This, Neil Jacobstein of Singularity University said, "should really be seen as part of an overall effort for

regulating AI. Such metrics are needed to determine if we've regulated enough or too much, and in the right ways. We are not developing AI in a competitive vacuum globally. China and other countries are racing ahead. It is possible to regulate us into irrelevance." However, such a vision will require independently verified numbers or peer review, and not just industry-supplied numbers, said Meredith Whittaker. We have seen instances in which industry claims outpace the performance realities, she said. Unfortunately, continued Whittaker, because there is no "alternative AI production ecology," it might be difficult to develop reliable numbers.

In any case, there are many actual and potential applications of AI that are simply not tracked right now, Jacobstein said. We ought to be collecting data about the role of AI in addressing pandemic diseases, climate change, illiteracy, international conflict and in the progress being made in meeting the other United Nation's Sustainable Development Goals.

Steve Chien of the Jet Propulsion Laboratory suggested that the widespread use of AI to increase scientific progress in diverse scientific disciplines such as biology, chemistry, geology, etc. should be tracked, documented and reported. "There are a large number of traditional disciplines such as chemistry, biology and environmental sciences that are leveraging AI in their research, documented and passing the science peer review process in scientific journals sufficient tracking emphasis for the AI contribution," he said. "That is a significant impact that should be tracked." Chien cited several articles as exemplars including AI researchers (Kiri L. Wagstaff, David R. Thompson, Radio Science/Astrophysics; Umaa Rebbapragada, Astronomy; David R. Thompson, Greenhouse emissions) that document instances in which machine learning played a significant role in major scientific finding and models.[25] Additionally, there are other high profile exemplars where AI is playing a major indirect role, such as in the AI-based scheduling for the Orbiting Carbon Observatory 3 Space Mission.

A variety of additional "missing metrics" for AI were mentioned by participants: It would be helpful to have more extensive data about AI developments in critical places around the world, such as China and Scandinavia. It would be useful to have ongoing tracking of state legislation that affects AI, and on the ways that philanthropy is funding AI-related work. It would also be helpful to synthesize and update the

ethical frameworks for AI that are being developed around the world. Marc Rotenberg mentioned the work of Australian computer scientist Roger Clarke, who has catalogued about fifty AI ethics frameworks and tried to distill their most critical elements.[26]

To paint a richer picture of how AI is developing, it would be worthwhile tracking employment trends for graduating AI researchers, said Tim Hwang of the Ethics and Governance of AI Initiative. What employers are researchers choosing, and who is hiring what types of experts for which topics? he asked. "Where researchers choose to do their work has relevance for who controls AI developments and whether or not the public has access to that research," said Hwang. He considers such numbers a rough proxy for assessing the social good and which AI topics are being more intensively developed.

Similarly, it would be helpful to have numbers that reveal the diversity of genders, people of color and other minorities within AI fields now dominated by white males. These numbers will be "probative of the kinds of technical problems that will be prioritized within AI," said Hwang, "which is relevant to how machine learning develops." This in turn could have a "huge influence on the social impact of machine learning."

In terms of government policymaking, the lack of shared metrics for describing AI has serious implications for coordinating federal funding for AI and fostering multidisciplinary research, said Terah Lyons, who used to work at the White House Office of Science and Technology Policy. Lyons said that "there is not a shared taxonomy for how we think about measuring artificial intelligence." Indeed, many interagency meetings foundered because there was no shared language or policy categories among participants. "It's an extremely fundamental challenge, but it's still one that hasn't been addressed," said Lyons.

Steve Chien witnessed a similar challenges among federal agencies during congressionally initiated AI review directed by the National Defense Authorization Act in August 2018. "A national assessment of the state of AI was directed, but answering fundamental questions within the government, such as quantifying the NASA AI investment posed a tremendous challenge due to (a) multiple definitions of AI and (b) overlapping programs and organizations. These challenges are not unique to NASA, similar experiences were experienced at other Federal Agencies and even non-Governmental Entities."

Law and Regulation to Oversee AI

A lively conference breakout group considered a variety of ways in which greater AI oversight and governance might be established. A first, obvious approach is industry self-regulation, which could take place industrywide or through individual companies—and within a company, via specific parts of the organization (legal, marketing, research, etc.). Another approach is a set of universal guidelines for AI uses, or synthesized guidelines from decentralized practices and policies that may already exist. One such framework, the "Universal Guidelines for AI," has been endorsed by AI experts and international associations, including the American Association for the Advancement of Science.[27] Independent bodies might also instigate new forms of auditing and reporting about AI behaviors within companies.

A cross-cutting concern is whether any regulation should be specific to a type of AI, or more universal in coverage. Participants were divided on this issue. Some felt that there should be laws equivalent to HIPAA (Health Insurance Portability and Accountability Act) or FERPA (Family Educational Rights and Privacy Act) to regulate AI and the various contexts in which it might be used. Others felt that government laws and regulations would be too slow and therefore not effective, or that new laws are either unnecessary ("nothing's broken, so why fix it?") or redundant (legal regimes already exist to regulate AI). Yet there was agreement that certain areas of AI, such as facial recognition and social scoring, may require domain-specific legislation.

…it may be worthwhile to think about *layers* of governance…there should be less opacity, more due process and an accent on fairness.

That said, there was agreement that broader AI-related harms deserve to be addressed. These include racial or gender discrimination, consumer manipulation or fraud, breaches of trust, privacy invasions, political interference and social scoring, said Anita LaFrance of the University of Pennsylvania.

The potential means of AI governance are quite familiar: laws, regulations, liability rules, tort law, contract law and intellectual property law. Business behaviors might be "nudged" through various incentives created through tax law, civil and criminal liabilities, and even reporting requirements, which can be a form of governance. This list suggests that it may be worthwhile to think about *layers* of governance, ranging from self-regulation to informational disclosures and guidelines to federal law and regulation. The general sense is that there should be less opacity, more due process and an accent on fairness.

There are signs that governance of privacy and AI issues may soon become more harmonized on an international scale. The Organization for Economic Cooperation and Development (OECD) is finalizing international guidelines for the design and use of AI. Similar OECD Guidelines for Privacy Protection have influenced national policies, industry practices and also helped resolve challenges for transborder data flows. In April 2019, the Trump Administration embraced the OECD initiative to develop the AI framework and to support related efforts by the OECD on privacy.[29] According to *The New York Times,* the White House was apparently concerned that the enactment of new state privacy laws and Europe's surging leadership on privacy protection could splinter domestic and international markets, to the detriment of U.S. technology companies. The OECD AI Guidelines are also in line with statements previously made by the White House regarding the protection of privacy, civil liberties and democratic values.[30] In a subsequent letter for *The New York Times*, Marc Rotenberg acknowledged the White House progress but also stated, "The United States must work with other democratic countries to establish red lines for certain AI applications and ensure fairness, accountability and transparency as AI systems are deployed."[31]

AI Review Panels, Impact Assessments and Certification

In a concluding presentation, Meredith Whittaker stressed that political choices lie at the heart of regulating AI. "When we ask who gets to determine which questions are relevant, what to measure and what to ignore, what gets funded and what research will be conducted, we begin to see that politics is going to define the scope of what AI means and its social impact," she said. "So these are decisions that we

should make with a great deal of intention and awareness." In assessing the future of AI, Whittaker urged that "we broaden the frame as wide as possible" so that we can take account of all factors—the labor costs behind AI, including precarious workers; the environmental impact of the technologies; the huge infrastructures that they entail; and the structural factors that determine AI affordances.

She suggested that we should be wary of relying too much on numbers: "Reducing life to numbers that can be managed by a few is a dangerous proposition, and AI offers a beguiling set of techniques that makes that seem very easy. But we are already seeing the potential consequences of that type of decision-making and the social asymmetries that can result." Asymmetrical power dynamics divide people into those who centrally control information and those who are the unorganized objects of information, she continued. AI intensifies this asymmetry because it mostly relies on "extractive processes that quantify and commodify our daily lives, personal interactions and our emotional signifiers."

Asymmetrical power dynamics divide people into those who centrally control information and those who are the unorganized objects of information…. AI intensifies this asymmetry.
– Meredith Whittaker

So what might be done beyond the industry self-regulation and government laws and regulations mentioned above? Whittaker offered a self-styled provocation for the group to consider: establish an AI review panel that would emulate the pioneering Cambridge Experimentation Review Board, which in the 1970s reviewed Harvard University's recombinant DNA research. The Board convened a representative cross-section of people who might be affected by the lab research—a nurse, teacher, parent, a scientist from another discipline, among others—and charged them with studying the issues, hearing arguments from all sides, and synthesizing a community consensus. Whittaker thinks that an AI review panel could host an intelligent conversation and interrogative process, and help build a common frame of reference in identifying and preventing social costs.

The virtue of this approach, said Whittaker, is that it provides "a participatory model for understanding that puts the burden on the experts to reach out to people who are potentially most at risk." A similar community panel, the Bronx Community Research and Review Board, was established by several hospitals in 1998 to make sure that their academic research practices are "fair, ethical and culturally appropriate" to the community.[32] Whittaker said such panels can help expand the definition of "what's important" in AI and avoid the rush to govern through numbers.

In a variant of this idea, Anita LaFrance Allen commended the idea of a national commission similar to the National Bioethics Commission established during the Obama Administration. That Commission assessed ethical problems raised by synthetic biology and served as a vehicle of "deliberative democracy" in formulating a consensus that might inform potential federal action. Even with no follow-through (the Trump Administration did not continue the commission), its dialogues focused the attention of affected parties and stimulated public discussion.

Some participants expressed skepticism at these ideas, however. A citizen review panel for AI would require a large investment of time and energy, and it could slow down or even stop certain AI initiatives. Alix Lacoste of Benevolent AI countered suggesting that in some cases, government and public intervention could in fact accelerate progress by enabling legislation, such as the Orphan Drug Act that sped up the drug approval process for medicines for rare diseases. In addition, Lacoste highlighted the potential positive role of government, philanthropy and review panels to help route AI research funds to scientific endeavors that may benefit society.

Reid Hoffman, the Co-founder of LinkedIn and Partner at Greylock Partners, emphasized that American tech companies are currently locked in a fierce race with China to develop AI, and various foreign intelligence agencies are trying to acquire American AI secrets. Hoffman said that citizen panels would get little public visibility and support, and if the government got involved, everything would move slowly, rendering any decisions ineffective.

A better approach than citizen panels, said Hoffman, would be to study a limited subset of AI, figuring out in advance what protections might be needed, and then to "re-factor" the oversight of AI later. "It's

a chimera to think you can actually get a real slowdown of AI given the nature of the competition and organizations operating here," he said. Whittaker replied that "framing the issue in terms of an arms race is implicitly xenophobic. If potential Chinese sovereignty in AI is raised as the bar against which we measure ourselves, I think we've already lost. It feels like a Red Scare narrative all over again."

Another form of oversight and governance to consider, said Marc Rotenberg of EPIC, is the idea of "impact assessments." The idea of rigorous, formal reviews of the likely impacts of a business project have long been used to ensure public accountability for the environment and privacy, he said. These are models that might be emulated. Indeed, the European General Data Protection Regulation has provisions for a "data protection impact assessment." Whittaker added that the AI Now Institute has in fact already produced an Algorithmic Impact Assessment framework, which it bills as "a practical framework for public agency accountability."[33]

There are also independent research and advocacy projects that might be worth creating or expanding. Neil Jacobstein suggested reviving the U.S. Office of Technology Assessment (OTA), which is still on the books, and until its defunding in the mid-1990s provided rigorous analyses of new technologies to the U.S. Congress. He said we need the OTA now more than ever. Additional third-party advocacy projects include the Algorithmic Justice League, which has documented racial biases in facial recognition software,[34] and the EU-funded AlgoAware project systematically reviews social and democratic issues raised by algorithms.[35]

It may be useful to have some type of organization that could act as an intermediary between AI projects and various constituencies, similar to the way that FINRA, the Financial Industry Regulatory Authority, mediates disputes among brokers, dealers and the investing public. One such example is a British think tank, doteveryone.org.uk, which the Omidyar Network has supported as an intermediary for transparency and fairness concerns in digital contexts.

Peter Norvig of Google suggested that perhaps a private, independent organization such as Underwriters Laboratories (UL) could help bolster public trust by certifying reliable AI services. In the early 1890s, when public distrust in the new technology of electricity was high, UL was founded to help reassure consumers about the safety of electrical

products. Various conference participants raised questions about the efficacy of certification programs, however, at least if applied to AI. It was pointed out that certification for AI would have to be domain-specific, not general. Yet even this approach would not necessarily prevent unauthorized "off-label" AI uses. Others questioned whether certification would actually change consumer decisions, especially when so much AI is managed at the enterprise level and is therefore invisible to consumers. Certification might also need government regulation as a backup regime if it were to be credible. Apart from these reasons, it was pointed out that since so many AI systems are still in formative stages, it is too early to identify the proper foci for certification or metrics.

Conclusion

Harnessing the immense power of artificial intelligence while controlling its potentially destabilizing consequences is indeed a wicked challenge. There are highly attractive breakthroughs that AI could deliver to humankind in terms of healthcare, scientific research and discovery, productivity, business innovation and wealth-creation. But there are also likely to be many complicated negative impacts—on employment, social inequality, democratic processes and possibly national security. There will be no universal solution—AI itself is too diverse and rapidly evolving—but clearly new modes of anticipating and controlling the unintended and/or catastrophic dimensions of AI are needed.

For a set of technologies that are still embryonic and evolving, and not necessarily even discussed with a common vocabulary within the U.S. government, this is a tall order. However, this Aspen Institute conference was encouraging in its own way because it surfaced some of the key vectors of engagement that must be joined: more cross-sectoral discussions, deeper philosophical inquiry, greater reflection on structural forces directing AI development. And most of all, how to prod AI development in the right directions—and what, indeed, *are* those "right directions?" These lines of exploration could be greatly aided by adopting new consensus metrics to assess AI and by establishing new governance mechanisms that can provide a greater measure of public accountability over the design and uses of the technologies. The challenge amounts to something of a koan, however: Can a technology that is inherently disruptive be made socially responsive, too?

Endnotes

1. Kai-Fu Lee, "The Human Promise of the AI Revolution," *The Wall Street Journal*, September 4, 2018, at https://www.wsj.com/articles/the-human-promise-of-the-ai-revolution-1536935115.

2. Office of Science and Technology Policy, "Accelerating America's Leadership in Artificial Intelligence," The White House. February 11, 2019, at https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/.

3. Besides its well-known censorship of the Internet, the Chinese government now pressures tens of millions of its citizens to use its propaganda/indoctrination app, "Study the Great Nation." See Javier C. Hernández, "The Hottest App in China Teaches Citizens About Their Leader – and Yes, There's a Test," April 7, 2019, at https://www.nytimes.com/2019/04/07/world/asia/china-xi-jinping-study-the-great-nation-app.html. The Chinese government is also using a secret system of advanced facial recognition technology to track and control the Uighurs, a largely Muslin minority. Paul Mozur, "One Month, 500,000 Face Scans: How China is Using A.I. to Profile a Minority," *The New York Times*, April 14, 2019, at https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html.

4. Nick Wingfield et al., "Amazon, Berkshire Hathaway and JPMorgan Team Up to Try to Disrupt Health Care," *The New York Times*, January 30, 2018, at https://www.nytimes.com/2018/01/30/technology/amazon-berkshire-hathaway-jpmorgan-health-care.html.

5. For instance, games like Re-Mission are educating users about their own bodies and helping them learn about treatments for diseases. See http://www.re-mission.net/.

6. Leslie Scism, "New York Insurers Can Evaluate Your Social Media Use - If they Can Prove What it's Needed," *The Wall Street Journal*, January 30, 2019, at https://www.wsj.com/articles/new-york-insurers-can-evaluate-your-social-media-useif-they-can-prove-why-its-needed-11548856802.

7. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. JAMA. 2017;318(6):517–518. doi:10.1001/jama.2017.7797.

8. Clifton Leaf, "'Unfinished Business': The Struggles of a Shrinking Middle Class," *Fortune*, December 20, 2018, at http://fortune.com/2018/12/20/shrinking-middle-class-report.

9. Kai-Fu Lee, *The Wall Street Journal*, September 4, 2018.

10. Kevin Roose, "The Hidden Automation Agenda of the Davos Elite, *The New York Times*, January 25, 2019, at https://www.nytimes.com/2019/01/25/technology/automation-davos-world-economic-forum.html.

11. David Bollier, *Power-Curve Society: The Future of Innovation, Opportunity, and Social Equity in the Emerging Networked Society* (Aspen Institute Communications and Society Program, 2013), at https://www.aspeninstitute.org/publications/power-curve-society-future-innovation-opportunity-social-equity-emerging-networked.

12. Ashish Vaswani, Noam Shazeer, et al., "Attention is All You Need," December 6, 2017, at https://arxiv.org/pdf/1706.03762.pdf.

13. World Economic Forum, "Assessing Gender Gaps in Artificial Intelligence," 2018 Global Gender Gap Report. http://reports.weforum.org/global-gender-gap-report-2018/assessing-gender-gaps-in-artificial-intelligence/.

14. Other participants suggested several web resources that could be valuable points for engagement for AI education: the Udacity and Coursera websites (www.udacity.com and www.coursera.org), and SARA (Socially Aware Robot Assistant) developed by Carnegie Mellon University. Such resources are potentially valuable because so much AI education requires self-learning and peer learning in some form or another.

15. J. Scott Brennen et al., "An Industry-Led Debate: How UK Media Cover Artificial Intelligence," Reuters Institute for the Study of Journalism, University of Oxford, December 2018, at https://reutersinstitute.politics.ox.ac.uk/our-research/industry-led-debate-how-uk-media-cover-artificial-intelligence.

16. James O'Toole, *The Executive's Compass* (New York, NY: Oxford University Press, 1993), Chapter 4.

17. DARPA describes its World Modeler program as seeking "to develop technology that integrates qualitative causal analyses with quantitative models and relevant data to provide a comprehensive understanding of complicated, dynamic national security questions. The goal is to develop approaches that can accommodate and integrate dozens of contributing models connected by thousands of pathways—orders of magnitude beyond what is possible today."

18. Casey Ross and Ike Swetlitz, "IBM pitching its Watson Supercomputer as a Revolution in Cancer Care. It's Nowhere Close," STAT, September 5, 2017, at https://www.statnews.com/2017/09/05/watson-ibm-cancer.

19. Jay W. Forrester, "The Counterintuitive Behavior of Social Systems," *Technology Review,* January 1971, updated in 1995 and available at http://static.clexchange.org/ftp/documents/system-dynamics/SD1993-01CounterintuitiveBe.pdf.

20. Yuval Noah Harari, "Why Technology Favors Tyranny," *The Atlantic*, October 2018, at https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technologh-tyranny/568330.

21. Renee DiResta, Dr. Kris Shaffer, et al, "The Tactics & Tropes of the Internet Research Agency," New Knowledge, 2018 at https://disinformationreport.blob.core.windows.net/disinformation-report/NewKnowledge-Disinformation-Report-Whitepaper.pdf.

22. William Cronon, *Nature's Metropolis: Chicago and the Great West* (New York, NY: W.W. Norton & Co., 1992).

23. See https://medium.com/what-to-build/is-anything-worth-maximizing-d11e648eb56f.

24. Yoav Shoham, Raymond Perrault, et al, "The AI Index 2018 Annual Report," AI Index Steering Committee, Human-Centered AI Initiative, Stanford University, December 2018), at https://aiindex.org.

25. Cynthia Rudin and Kiri L. Wagstaff, "Machine Learning for Science and Society," *Machine Learning*, April 2014 95(1), pp. 1-9, at https://link.springer.com/article/10.1007/s10994-013-5425-9. https://doi.org/10.1007/s10994-013-5425-9.

26. Roger Clarke, *Principles for AI: A SourceBook* April 15, 2019, at https://www.rogerclarke.com/EC/GAIP.html.

27. Universal Guidelines for Artificial Intelligence (2018), https://thepublicvoice.org/ai-universal-guidelines/.

28. OECD moves forward on developing guidelines for artificial intelligence (AI), Feb. 20, 2019, at https://www.oecd.org/going-digital/ai/oecd-moves-forward-on-developing-guidelines-for-artificial-intelligence.htm.

29. Steve Lohr, "A.I. and Privacy Concerns Get White house to Embrace Global Cooperation," *The New York Times,* April 3, 2019, at https://www.nytimes.com/2019/04/03/technology/artificial-intelligence-privacy-oecd.html.

30. Donald J. Trump, Executive Order on Maintaining American Leadership in Artificial Intelligence, Feb. 11, 2019, https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/.

31. Marc Rotenberg, The Battle Over Artificial Intelligence, *The New York Times*, April 18, 2019, at https://www.nytimes.com/2019/04/18/opinion/letters/artificial-intelligence.html.

32. See http://www.bronxhealthlink.org/tbhl/programs.

33. Dillon Reisman, Jason Schultz, Kate Crawford, and Meredith Whittaker, "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability" April 2018, at https://ainowinstitute.org/aiareport2018.pdf.

34. See https://www.ajlunited.org.

35. See https://www.algoaware.eu.

APPENDIX

"Developing Goals and Metrics for the Good Society"

Santa Barbara, California
February 11-13, 2019

Roundtable Participants

Anita LaFrance Allen
Vice Provost for Faculty, and
Henry R. Silverman Professor
of Law and Philosophy
University of Pennsylvania Law
School

David Bollier
Rapporteur

John Seely Brown
Independent Co-Chairman
Deloitte Center for the Edge

De'Aira Bryant
Ph.D. Student
Computer Science: Intelligent
Systems
Georgia Institute of Technology
Human Automation Systems
(HumAnS) Lab

Steve Chien
Senior Research Scientist
and Head of the Artificial
Intelligence Group
Jet Propulsion Laboratory
California Institute of
Technology

Michael Chui
Partner, McKinsey Global
Institute
McKinsey & Company

Vilas Dhar
Trustee
Patrick J. McGovern Foundation

Suzanne DiBianca
Executive Vice President,
Corporate Relations
and Chief Philanthropy Officer
Salesforce

David Ferrucci
Founder and Chief Executive
Officer
Elemental Cognition

Charlie Firestone
Executive Director
Communications & Society
Program
The Aspen Institute

Raina Kumra Gardiner
Director
Omidyar Network

Note: Titles and affiliations are as of the date of the conference.

Tom Gruber
AI Product Designer

Reid Hoffman
Co-Founder, LinkedIn
Partner, Greylock Partners

Tim Hwang
Director, Ethics and Goverance of AI Initiative MIT Media Lab and Harvard Berkman-Klein Center for Internet and Society

Neil Jacobstein
Chair, Artificial Intelligence & Robotics
Singularity University

Alix Lacoste
Vice President, Data Science
BenevolentAI

Terah Lyons
Executive Director
Partnership on AI

Gary Marcus
Professor of Psychology and Neural Science
New York University

Patrick McGovern
Trustee
Patrick J. McGovern Foundation

Peter Norvig
Director of Research
Google

Tess Posner
Chief Executive Officer
AI4ALL

Marc Rotenberg
President and Executive Director
Electronic Privacy Information Center (EPIC)

Fr. Eric Salobir
President
OPTIC

Meredith Whittaker
Co-Founder and Co-Director
AI Now Institute, and
Distinguished Scientist
New York University

Staff:

Sarah Eppehimer
Project Director
Communications & Society Program
The Aspen Institute

Kristine Gloria
Senior Project Manager
Communications & Society Program
The Aspen Institute

Note: Titles and affiliations are as of the date of the conference.

About the Author

David Bollier is an American activist, scholar and blogger whose work focuses on *the commons* as a new/old paradigm for re-imagining economics, politics and culture. He pursues this work as Director of the Reinventing the Commons Program at the Schumacher Center for a New Economics, and as co-founder of the Commons Strategies Group, an international advocacy project.

Bollier has co-organized pioneering international conferences and strategy workshops, and consults regularly with diverse activists and policy experts in the U.S. and Europe. His blog, Bollier.org, is a widely read source of news about the commons, and his book *Think Like a Commoner: A Short Introduction to the Life of the Commons* (2014), has been translated into six languages. He and coauthor Silke Helfrich published *Free, Fair and Alive: The Insurgent Power of the Commons* in English and German in 2019.

Bollier is an editor or author of many other books, including *Patterns of Commoning* (2015) and *The Wealth of the Commons* (2012), both with co-editor Silke Helfrich; *Green Governance* (2013), co-authored with the late Professor Burns Weston; and *Viral Spiral* (2009), *Brand-Name Bullies* (2005), and *Silent Theft* (2002). In 2012, Bollier received the Bosch Berlin Prize in Public Policy from the American Academy in Berlin for his work on the commons. Bollier lives in Amherst, Massachusetts.

About the Communications and Society Program

www.csreports.aspeninstitute.org

The Communications and Society Program advances democratic values through communications and information technology policy. The Program convenes diverse global leaders and experts, frames issues for the exchange of insights on the societal impact of digital and network technologies, and catalyzes new policies and leadership that serve the public interest. It enables global leaders and experts to explore new concepts, exchange insights, develop meaningful networks, and find personal growth, all for the betterment of society.

The Program's projects range across many areas of information, communications and media policy, such as artificial intelligence, broadband and spectrum policy, race and media, institutional innovation, and diplomacy and technology. The Program has also run ongoing projects on trust, media and democracy, and on the future of public libraries.

Most conferences employ the signature Aspen Institute seminar format: approximately 25 leaders from diverse disciplines and perspectives engaged in roundtable dialogue, moderated with the goal of driving the agenda to specific conclusions and recommendations. The Program distributes its conference reports and other materials to key policymakers, opinion leaders and the public around the world. Its digital reports platform aims to inform and ignite broader conversations at the intersection of democracy and communications technologies.

The Program's Executive Director is Charles M. Firestone. He has served in this capacity since 1989 and is also a Vice President of the Aspen Institute. Prior to joining the Institute, Mr. Firestone was a communications attorney and law professor who argued two cases before the United States Supreme Court and many in the courts of appeals. He is a former director of the UCLA Communications Law Program, first president of the Los Angeles Board of Telecommunications Commissioners, and an appellate attorney for the U.S. Federal Communications Commission.

Previous Publications from the Aspen Institute Roundtable on Artificial Intelligence

Artificial Intelligence, The Great: Coming to Terms with AI-Driven Markets, Governance and Life, David Bollier

The Roundtable on Artificial Intelligence, held in August 2017, challenged the powerful narrative of AI's growing dominance and inevitable influence on today's society. To facilitate this thinking, the Roundtable encouraged participants to re-situate the human as the focal point, asking the question: In what ways do AI innovations enhance and or limit personal human autonomy? Discussions traversed the philosophical to the applied, and raised deep, fundamental questions on how to guide the trajectory of AI "with meaning and dignity for humans."

Artificial Intelligence Comes of Age: The Promise and Challenge of Integrating AI into Cars, Healthcare and Journalism, David Bollier

The Aspen Institute Roundtable on Artificial Intelligence (AI) is an annual private seminar of business executives, government leaders, and visionaries. It examines the implications of artificially intelligent technologies on societies, governments, communities and individuals, and the new leadership roles that are required. As its inaugural year (2016), the report concentrated on three artificial intelligence sectors: self-driving cars, AI and Medicine, and AI and the Media.

Reports can be ordered online at *www.aspeninstitute.org/publications* or by sending an email request to *publications@aspeninstitute.org.*