The Social & Economic Implications of Artificial Intelligence Technologies in the Near-Term
July 7th, 2016; New York, NY
http://artificialintelligencenow.com

WORKSHOP PRIMER: INEQUALITY & AI

Contents

Brief Topic Description

Artificial Intelligence (AI) promises to reshape the contours of inequality in society in at least five significant ways.[1]

1. AI is already creating new ways to generate economic value while also affecting how this value is distributed. *If the ability to develop and implement AI systems is not available equitably, certain parts of the population may claim the lion's share of the economic surplus.* In this scenario, those who have data and the computational power to use AI will be able to gain insight and market advantage, creating a "rich-get-richer" feedback loop in which the successful are rewarded with more success.[2]

2. AI may help to expose and counteract the prejudices and biases that undergird persistent social inequalities along the lines of race, gender and other characteristics.

[1] As Russell and Norvig point out, the history of artificial intelligence has not produced a clear definition of AI but can be seen as variously emphasizing four possible goals: "systems that think like humans, systems that act like humans, systems that think rationally, systems that act rationally." Here we rely on the emphasis proposed by Russell and Norvig, that of intelligence as rational action, and that "an intelligent agent takes the best possible action in a situation." Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Englewood Cliffs, NJ: Prentice Hall, 1995: 27.

[2] Donella H. Meadows and Diana Wright, *Thinking in Systems: A Primer*, White River Junction, VT: Chelsea Green Publishing, 2008.

Alternatively, AI systems may inherit, even amplify, many of these same prejudices and biases. AI depends on the data it is given and may reflect the characteristics of such data, including any biases.

3. There are two ways in which data can be biased: one is because the data available is not an accurate reflection of reality; the other is because the underlying process itself exhibits long-standing structural inequality. The former type of bias can sometimes be addressed by 'cleaning the data' or improving the data collection process; the latter requires interventions with complex political ramifications.

4. With more data, analytics, and automation, the capacity to draw more fine-grained distinctions between people will increase, and this may substantially increase profits for industries like insurance. This has the potential to affect how risk is pooled in society. AI may endanger the solidarity upon which insurance and other collective risk mitigations strategies rely.

5. Finally, women and minorities continue to be under-represented in the field of AI. A lack of inclusion and "like me" bias may limit the degree to which practitioners choose to recognize or prioritize the concerns of other communities.

AI redistributes wealth, but to whom?

AI has the potential to generate new economic value by creating and expanding products and services, as well as by reducing costs through automation and other approaches. How this economic value will be distributed? This is a key concern for economic inequality.

If the vast majority of economic value is distributed to those already among the wealthiest today, then the expansion of AI technologies could widen existing wealth and income disparities. In such a scenario, AI would exacerbate inequality.

At the same time, if automation reduces the costs of creating certain goods and services (and these savings are passed on to consumers), AI could narrow the gap between the haves and have-nots (at least in terms of access to these goods and services).[3] In this case, AI would increase the standard of living overall, and could even have a progressive redistributive effect.

AI could also give rise to entirely new ways of making a living, either by allowing people whose jobs have been eliminated to seek other ways of obtaining resources, or by creating new jobs that ensure affected workers an easy transition to other paid labor. Rather than simply replacing workers or reducing workers' share of the profits from productive activity, AI could free people to pursue new forms of living and work, increasing overall welfare in society.

[3] Dean Baker, "Can Productivity Growth Be Used to Reduce Working Time and Improve the Standard of Living of the 99 Percent? The Future of Work in the 21st Century," Economic Analysis and Research Network, 2014, http://www.earncentral.org/Future_of_work/Baker%20Shorter%20Work%20Time%20Final.pd.

Of course, as some commentators have observed, AI could also render certain workers' skills redundant, leaving those who have been replaced by automation with few options for alternative paid employment.[4] Even if workers can find new employment, these jobs may be of lower quality, may pay less, and may provide less stability and security than the jobs eliminated by AI.[5] This is why understanding AI's potential impact on human employment is an important aspect of understanding its impact on economic equality.

Further, if learning new skills is prohibitively expensive, workers may find it impossible to pivot to a new profession. Under these circumstances, not only would AI increase inequality, it might push certain parts of the workforce into permanent unemployment and poverty.

Such concerns have driven growing interest in social safety net programs, including universal basic income (UBI). UBI is a redistribution framework that would provide all members of society with the minimum amount of money to cover life's necessities.[6] UBI's supporters see it as a way to address whatever ongoing unemployment AI might cause. Unlike need-based welfare programs, UBI would divert the same amount of money to all members of society. This eliminates, according to proponents, the complex and costly bureaucracy necessary to administer programs targeted at the poor.[7] Crucially, UBI would aim to provide an income floor beyond which no one would fall. But it would not aim to ensure a more equitable distribution of income in society overall. This means that inequality could still increase, even in a society that has adopted UBI. Some critics discount UBI as a way of simply making growing inequality more palatable by keeping the unemployed out of poverty, and potentially providing corporations with an effective subsidy that allows them to compensate human workers less, and reap more of the profits.[8]

The uneven availability of AI technologies may increase the developers or adopters' bargaining power, providing them a "crystal ball" of analytical insight unavailable to others.

For example, if AI techniques allow one party in an exchange to effectively infer the other party's sensitivity to price (say, the knowledge that the other party is unlikely to bargain for a lower price, or, will require a 20% drop in price before being willing to sign a

[4] Erik Brynjolfsson and Andrew McAfee,*The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies*, New York: W.W. Norton & Company, 2014. For more context, see the Labor & AI primer from *AI Now*, https://artificialintelligencenow.com/

[5] Henry Siu and Nir Jaimovich, "Jobless Recoveries," Third Way (April 2015), https://s3.amazonaws.com/content.thirdway.org/publishing/attachments/files/000/000/862/NEXT_-_Jobless_Recoveries.pdf?1428093868; Roosevelt Institute, "Technology and the Future of Work: The State of the Debate," Open Society Foundations Future of Work Project (April 2015), https://www.opensocietyfoundations.org/publications/technology-and-future-work-state-debate.

[6] See, for example, Nick Srnicek and Alex Williams, *Inventing the future: Postcapitalism and a World without Work*, New York: Verso Books, 2015; but also Sam Altman, "Basic Income," *Y Combinator Posthaven*, January 27, 2016, https://blog.ycombinator.com/basic-income.

[7] Eduardo Porter, "A Universal Basic Income Is a Poor Tool to Fight Poverty," *The New York Times*, May 31, 2016, http://www.nytimes.com/2016/06/01/business/economy/universal-basic-income-poverty.html.

[8] Jathan Sadowski, "Why Silicon Valley is embracing universal basic income," *The Guardian*, June 22, 2016, https://www.theguardian.com/technology/2016/jun/22/silicon-valley-universal-basic-income-y-combinator.

contract), the party with access to AI will likely be at an advantage.[9] Likewise, benefits may go to the party that can rely on AI to determine the other party's susceptibility to different modes of persuasion (say, the knowledge that the other party will likely buy a product if they are 'primed' with a story about puppies).[10] At the extreme, these techniques may lead to predatory practices.[11]

And yet, if deployed more equitably, AI could also address power dynamics in markets that already exhibit such information asymmetries, and it could enable people to make more considered choices and counteract tendencies to make poor, irrational, or impulsive decisions.[12] The means of providing such access, however, remains an open question.

AI: reproducing or correcting human bias?

Automated decisions will increasingly shape people's life chances.

As AI takes on a more important role in high-stakes decision-making, it will begin to affect who gets offered crucial opportunities, and who is left behind—from offers of credit and insurance, to the availability of job opportunities and parole. Some hope that AI will help to overcome the biases that plague human decision-making;[13] others fear that AI will amplify such biases, denying opportunities to the deserving and subjecting the deprived to further disadvantage.[14]

Either way, the underlying data will play a crucial role and deserves keen attention. This is because recent advances in AI have happened almost entirely in the subfield of Machine Learning (ML), where computers learn how to perform a task by finding patterns in a large dataset of examples—examples often taken from human activities in similar domains.

There is the hope that by giving AI "all of the data," and allowing it to detect complex and subtle patterns that escape human recognition, it could illuminate and help to tear down artificial barriers that have contributed to social inequality along the lines of race, gender, and age, among other characteristics.

At the same time, there is the risk that if the data used reflects these biases, AI trained on this data will replicate and magnify those biases. In such cases, AI would exacerbate discriminatory dynamics that create social inequality, and would likely do so in ways that

[9] Council of Economic Advisers, *Big Data and Differential Pricing*, February 2015, https://www.whitehouse.gov/sites/default/files/docs/Big_Data_Report_Nonembargo_v2.pdf.

[10] Ryan Calo, "Digital Market Manipulation," *George Washington Law Review* 82, no. 4 (October 3, 2014): 995-1051.

[11] Upturn, *Led Astray: Online Lead Generation and Payday Loans*, October 2015, https://www.teamupturn.com/reports/2015/led-astray.

[12] Richard T Ford, "Save the Robots: Cyber Profiling and Your So-Called Life," *Stanford Law Review* 52, no. 5 (2000): 1578-1579.

[13] Future of Privacy Forum and Anti-Defamation League, *Big Data: A Tool for Fighting Discrimination and Empowering Groups*, September 11, 2014, https://fpf.org/wp-content/uploads/Big-Data-A-Tool-for-Fighting-Discrimination-and-Empowering-Groups-Report1.pdf.

[14] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Crown, 2016.

would be less obvious than human prejudice and implicit bias (and that could be justified as the work of "intelligent technology," and thus potentially perceived as more neutral, and more difficult to contest).[15] AI presents the possibility of both removing and replicating the bias in decisions that materially shape people's lives.

Such concerns were previously recognized in the 2014 White House reports on big data[16] and have become a major issue for the civil rights advocacy community.[17] These awareness-raising efforts have pushed the issue of discrimination to the center of current debates about big data, and the concerns are amplified in the context of AI.[18] The White House recently issued a follow-on report specifically on the topic of "Algorithmic Systems, Opportunity, and Civil Rights," examining the risks, benefits, and challenges of using data in automated decision-making for credit, employment, education, and criminal justice.[19]

While these reports emphasize that AI may help to advance civil rights and diversity, they also acknowledge that simply reconfiguring decision-making to rely more on AI and less on ad hoc human judgment may not eliminate human bias from the decision-making process or radically transform the composition of key institutions. AI may, potentially, even amplify such bias or undermine efforts to increase diversity. There are many reasons for this caution. AI requires data and, as Solon Barocas and Andrew Selbst point out, "data is frequently imperfect in ways that allow machines trained on it to inherit the prejudices of prior decision-makers or reflect the widespread biases that persist in society at large."[20] Other researchers have documented cases of bias in the wild in such varied applications as online advertising for both criminal background checks[21] and employment services,[22] web search,[23] pre-trial detention,[24] and facial recognition.[25]

[15] Tal Zarsky, "The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making," *Science, Technology & Human Values* 41, no. 1 (2016): 118-132.

[16] The White House Office of Science and Technology Policy, *Big Data: Seizing Opportunities, Preserving Values*, May 2014, http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf; President's Council of Advisors on Science and Technology, *Big Data and Privacy: A Technical Perspective*, May 2014, http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf

[17] See The Leadership Conference, *Civil Rights Principles for the Era of Big Data*, 2014, http://www.civilrights.org/press/2014/civil-rights-principles-big-data.html, and the recurring conference on Data & Civil Rights: http://www.datacivilrights.org/

[18] The Federal Trade Commission, *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues*, January 2016, https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

[19] The White House. *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*, May 2016, https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

[20] Solon Barocas and Andrew Selbst, "Big Data's Disparate Impact," *California Law Review* 104, no. 3 (June 2016): 671.

[21] Latanya Sweeney, "Discrimination in Online Ad Delivery," *Communications of the ACM* 56, no. 5 (2013): 44-54.

[22] Amit Datta, Michael Carl Tschantz, and Anupam Datta, "Automated Experiments on Ad Privacy Settings," *Proceedings on Privacy Enhancing Technologies* (2015): 92-112.

[23] Safiya Umoja Noble, "'Just Google It': Algorithms of Oppression," University of British Columbia, December 8, 2015, https://youtu.be/omko_7CqVTA.

[24] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," *ProPublica,* May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[25] Clare Garvie And Jonathan Frankle, "Facial-Recognition Software Might Have a Racial Bias Problem," *The Atlantic*, April 7, 2016,
http://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/.

AI paved with best intentions

When AI leads to discrimination, it is most often unintentional. The people who create the algorithms, the people who input the training data, and the people who maintain AI systems rarely *want* to create a biased outcome. And yet, properties of the data can lead to biased results, best intentions notwithstanding.

Given these risks, there have been calls for greater transparency and due process for automated decisions.[26] Understanding how bias creeps into AI systems is a difficult but critical challenge. Even being able to *recognize* that it is happening can be technically challenging, including for experts.[27]

Two complimentary communities of researchers have formed over the past few years to address these challenges. The first focuses on auditing algorithmic systems through "black box testing," carefully varying the information presented to an AI system and observing any apparent disparity in output according to race, gender, age, or other characteristics.[28] Rather than aiming to understand the inner-working on an AI system, these techniques involve creating a range of profiles for a website with personalized content, say, and seeing how the online experience differs for each.

The second focuses on correcting for bias in the training data, ensuring that models developed through machine learning meet some formally specified standard of fairness, prioritizing interpretability in the process of generating models, or introducing accountability mechanisms.[29] Both of these initiatives can play a crucial part in detecting and attempting to mitigate bias in AI, but they face a number of practical challenges, including uncertainty regarding the legality of some of the technical auditing methods.[30]

[26] Danielle Keats Citron, "Technological Due Process," *Washington University Law Review* 85 (2007): 1249-1313; Kate Crawford and Jason Schultz, "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms," *Boston College Law Review* 55, no. 1 (2014): 93-128.

[27] Sweeney, "Discrimination in Online Ad Delivery"; Datta, Tschantz, and Datta, "Automated Experiments on Ad Privacy Settings."

[28] See, for example, the Web Privacy and Transparency Conference, https://citp.princeton.edu/event/web/, and Auditing Algorithms From the Outside: Methods and Implications, https://auditingalgorithms.wordpress.com/.

[29] See Fairness, Accountability, and Transparency in Machine Learning, http://www.fatml.org/.

[30] Esha Bhandari and Rachel Goodman, "ACLU Challenges Computer Crimes Law That is Thwarting Research on Discrimination Online," Free Future, June 29, 2016, https://www.aclu.org/blog/free-future/aclu-challenges-computer-crimes-law-thwarting-research-discrimination-online; Ifeoma Ajunwa, Sorelle Friedler, Carlos E. Scheidegger, and Suresh Venkatasubramanian, "Hiring by Algorithm: Predicting and Preventing Disparate Impact," March 10, 2106, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2746078.

Making inequality while seeking profit

The lasting historical effects of discrimination mean that AI systems may do the opposite of what their designers intend.

The reason for this, again, goes back to data, and what computer scientists working on discrimination in data mining call "redundant encodings."[31] In especially rich datasets (those with many variables tracking all sorts of things—the kinds of datasets that large social media companies or credit reporting agencies may collect), characteristics like race, gender, or age are almost certainly reflected in other, seemingly benign data points. Take race as an example. In the most obvious cases, redundant encoding can involve fairly obvious proxies for race like zip codes (such as areas historically subject to redlining and segregation). In subtler instances, other variables can act as unintentional proxies. For instance, race may be redundantly encoded in the list of websites visited by a particular user, sites especially popular among the Latino population in the United States, or African American teens on the East Coast, or white men above the age of 45, etc.

However, if the same data that redundantly encodes race also serves as a reliable predictor of customer value (people who visit a given combination of websites are more likely to click on payday loan ads, say), a company could systematically disadvantage members of a certain race in an attempt to maximize its profits.

Because inequality is not random, this is not very surprising. In a society suffering from bias along lines of gender, race, and other characteristics, these groups are more likely to be economically disadvantaged. And simply reproducing existing economic inequality, as in the examples above, will have a disproportionately negative effect on these groups.

Rendering judgment on the fairness of such rational market optimization is complex and difficult, especially because these practices are part of a wider set of forces and dynamics in capitalist societies. At the extreme, efforts to mitigate against these will necessarily begin to blur the line between a policy designed to ensure procedural fairness and one aimed at distributive justice.[32]

Certain computer scientists have put a fine point on this problem by demonstrating that there can be "a quantitative trade-off between fairness and utility."[33] The only way to ensure that consequential decisions do not amplify social biases—even unintentionally—is to ignore certain *relevant* details that also happen to act as redundant encodings. The resulting decisions would be less "accurate" (have less "utility") because they would grant opportunities to some people who—by the model's original estimate—might not "deserve" them. And it would do this at others' expense, effectively re-allocating the opportunities from one group to another. In some ways, this issue

[31] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel, "Fairness through Awareness," *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 2012, 215.

[32] Barocas and Selbst, "Big Data's Disparate Impact," 723–728.

[33] Dwork et al., "Fairness through Awareness," 215.

echoes debates about affirmative action versus the idea of a level playing field, and whether long-standing social inequalities should or should not be addressed by direct intervention in individual decision-making.

AI ultimately raises a basic question about the line that divides objectionable discrimination from the reproduction of inequality. Should normative principles enshrined in discrimination law address these practices? Even setting aside the likely reception of a policy to address economic disparities with discrimination law, such regulations might not be the most efficient, effective, or fair way to remedy what is ultimately a problem of economic inequality. Nevertheless, in societies where economic redistribution is limited and progressive policies might raise political objections, discrimination law might be one important vehicle for ensuring that AI does not contribute to a more unequal society.

The lifted veil

By offering more accurate predictions, AI promises to reduce economic uncertainties that currently result in cross-subsidies between different groups in society. For example, the inability to predict with perfect accuracy whether someone will fall sick has tended to ensure that people engage in risk-pooling through health insurance, with the result that the financial contributions of the healthy help to cover the costs of the ill.

To the extent that AI can reduce the uncertainty around the onset of some medical conditions, it could help manage these risks more effectively. But it could also reduce the willingness of people to participate in such programs if the cross-subsidization becomes more obvious, and people realize they personally won't reap the benefits. In other words, those who know they're healthy might balk at the idea of paying for the sick.[34]

As the President's Council of Economic Advisors notes, this will very likely apply in other economic arenas with risk-based price discrimination,[35] and competition will likely push insurers down this path. Policyholders who possess traits or exhibit behaviors that AI recognizes as reliable predictors of the future onset of certain diseases or disorders will be charged higher premiums. In such cases, those in already less favorable circumstances may end up bearing more of the economic brunt of poor health. This is why critics often charge that even if these predictions are accurate and the insurer's behavior is rational, the effects are ultimately perverse.[36]

The capacity to precisely know who within a population will be more or less costly may endanger the solidarity upon which insurance and other such social safety nets rely. AI will allow companies to engage in far more effective "adverse selection," identifying the particular populations that they would do better to avoid or under-serve. Companies will be better placed to focus their attention on those populations that will prove most

[34] Alistar Croll, "New Ethics for a New World," *O'Reilly Radar*, October 17, 2012, http://radar.oreilly.com/2012/10/new-ethics-for-a-new-world.html.

[35] Council of Economic Advisers, *Big Data and Differential Pricing*.

[36] Upturn, "As Insurers Embrace Big Data, Fewer Risks Are Shared," *Civil Rights, Big Data, and Our Algorithmic Future*, September 2014, https://bigdata.fairness.io/insurance/.

profitable.[37] We may be undoing an important feature of insurance by giving organizations extraordinarily precise information that could allow these companies to make inferences that will deny care or limit coverage to people in need.

What this reveals is that *the absence of error or bias does not make the treatment of individuals appropriate or desirable*. Here we see the well-recognized danger of "a computer-generated class system" that "could unfairly stratify consumers, covertly offering better pricing to certain people while relegating others to inferior treatment."[38]

This is what Oscar Gandy calls a 'cumulative disadvantage': even accurate inferences have their costs if they further entrench already disadvantaged populations into less favorable circumstances.[39] These are not worries about prejudice, bias, or error; they are broader concerns about the fairness of a system that, in rationally apportioning opportunities and costs, compounds existing structural inequities.

Ultimately, the structure of the market may determine how discerning we want AI to be. In a society with universal health care, for example, there are no opportunities to engage in adverse selection because one insurer covers all costs. As Paul Krugman remarks, health insurers in the United States essentially compete by trying to deny coverage to those who are most likely to need it;[40] drawing especially fine distinctions between people simply exacerbates these tendencies. In a society that does not rely on markets for the provision of health insurance, however, the same tactics could benefit both a sole insurer and patients because it would allow the sole insurer to more effectively manage risk and thus reduce the overall costs of health insurance.

AI needs women and people of color

As a field, computer science suffers from a lack of diversity. Women, in particular, are heavily underrepresented. The situation is even more severe in the subfield of AI. For example, while some AI academic labs are being run by women, only 13.7 percent of attendees were women this past year at NIPS, one of the most important annual AI conferences.[41]

A community that lacks diversity is less likely to consider the needs and concerns of those not among its membership. As Jane Margolis and Allan Fisher point out, the underrepresentation of women in AI can have very serious consequences:

> In an example from computer science, some early voice recognition systems were calibrated to typical male voices. As a result, women's voices were literally

[37] Bart Custers, *The Power of Knowledge: Ethical, Legal and Technological Aspects of Data Mining and Group Profiling in Epidemiology*, Nijmegen, Netherlands: Wolf Legal Publishers, 2004.

[38] Natasha Singer, "Your Online Attention, Bought in an Instant by Advertisers," *The New York Times*, November 17, 2012, http://www.nytimes.com/2012/11/18/technology/your-online-attention-bought-in-an-instant-by-advertisers.html.

[39] Oscar Gandy, *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage*, New York: Routledge, 2016.

[40] Paul Krugman, "Why Markets Can't Cure Healthcare," *The New York Times*, July 25, 2009, http://krugman.blogs.nytimes.com/2009/07/25/why-markets-cant-cure-healthcare/.

[41] Jack Clark, "Artificial Intelligence Has a 'Sea of Dudes' Problem," *Bloomberg*, June 23, 2016, http://www.bloomberg.com/news/articles/2016-06-23/artificial-intelligence-has-a-sea-of-dudes-problem.

> unheard…Similar cases are found in many other industries. For instance, a predominantly male group of engineers tailored the first generation of automotive airbags to adult male bodies, resulting in avoidable deaths for women and children.[42]

Amid growing recognition that heterogeneous groups outperform those that are more homogeneous,[43] there is also good reason to believe that increased diversity in the field of AI will help AI technologies to serve the interests of a diverse population.

Researchers note that the field's struggle to identify and confront issues of bias and inequality has been hampered by the current lack of diversity.[44] The concerns that shaped the debates about risk and AI have been driven by the interests and anxieties of wealthy, white men, rather than members of historically disadvantaged communities.[45] To address pressing issues of bias, discrimination, and inequality, the AI community will need to draw on a broader range of perspectives.[46] Like all technologies before it, AI reflects the values of its creators, and increased diversity may assist in a future for AI technologies that promotes greater equality.[47]

Questions to consider

- Will AI merely redistribute economic value in ways that favor the already wealthy, or grow the economy in new ways?
- How might we ensure that the economic benefits of AI technologies are widely distributed throughout society? Can market forces alone ensure that AI has a progressive redistributive effect, or are regulations or programs necessary to achieve these goals? Which mechanism is likely to be most efficient, effective, or most widely supported?
- How should we address the likely information and power asymmetries produced by AI technologies? Is making AI capacities available in non-proprietary modes enough to redress imbalances and produce a level playing field?
- How should we guard against bias in AI? Should discrimination law play a role? How might we make use of technical solutions in combination with institutional procedures and professional ethics?
- Is liability a sufficient incentive to guard against data errors and avoidable bias or is there a need for new laws?
- Are most consequential cases of bias in AI likely to come to light on their own or remain unnoticed? Who is best positioned to determine whether AI suffers from biases? How can that work be supported?
- What practical, institutional, and legal challenges do those working on auditing AI systems face and how should these be addressed?

[42] Jane Margolis and Allan Fisher, *Unlocking the Clubhouse: Women in Computing*, Cambridge, MA: MIT Press, 2003, 2-3.

[43] Scott Page, *The Difference: How the Power of Diversity Creates better Groups, Firms, Schools, and Societies*, Princeton, NJ: Princeton University Press, 2008.

[44] Kate Crawford, "Artificial Intelligence's White Guy Problem," *The New York Times*, June 25, 2016, http://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

[45] Ibid.

[46] For example, Women in Machine Learning, http://wimlworkshop.org/.

[47] Crawford, "Artificial Intelligence's White Guy Problem."

- To whom should the burden of correcting for bias fall? Are the designers/developers of AI responsible for purging bias from their systems, even if these biases reflect widely held social beliefs expressed in the historical data from which AI systems learn?
- How should we differentiate between cases where AI discriminates and cases where AI replicates economic inequality along, for instance, racial lines? Should designers of AI systems incur the costs of addressing inequality in society?
- Ultimately, what will “fair” AI aim to achieve? And how will such notions of fairness gain political legitimacy and practical buy-in?
- Will AI unravel insurance markets or improve the management of risk? What structures are necessary to ensure that AI can improve how institutions manage risk while not contributing to further inequality?
- How will greater diversity in the professional field of AI shift the values and norms embedded in these systems? What kinds of diversity should the community attempt to cultivate and how should it do this?
- How do we encourage people to consider the risks of AI while not discouraging people from adopting the technology in ways that might benefit disadvantaged communities?