The AI Now Report

The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term

A summary of the AI Now public symposium, hosted by the White House and New York University's Information Law Institute, July 7th, 2016

artificialintelligencenow.com

Contents

Co-Chairs
Kate Crawford (Microsoft Research & New York University)
Meredith Whittaker (Google Open Research)

Core Research Team
Madeleine Clare Elish (Columbia University)
Solon Barocas (Microsoft Research)
Aaron Plasek (Columbia University)
Kadija Ferryman (The New School)

Program Committee
Ryan Calo (University of Washington)
Ed Felten (White House Office of Science and Technology Policy)
Eric Horvitz (Microsoft Research)
Terah Lyons (White House Office of Science and Technology Policy)
Helen Nissenbaum (New York University)
Mustafa Suleyman (DeepMind)
Latanya Sweeney (Harvard University)
Nicole Wong (former Deputy US Chief Technology Officer)

Version 1.0
Last edited September 22, 2016

AI Now Overview

Artificial intelligence (AI) refers to a constellation of technologies, including machine learning, perception, reasoning, and natural language processing. While the field has been pursuing principles and applications for over 65 years, recent advances, uses, and attendant public excitement have returned it to the spotlight.[1] The impact of early AI systems is already being felt, bringing with it challenges and opportunities, and laying the foundation on which future advances in AI will be integrated into social and economic domains. The potential wide-ranging impact make it necessary to look carefully at the ways in which these technologies are being applied now, whom they're benefiting, and how they're structuring our social, economic, and interpersonal lives.

AI Now was the final in a five-part series of convenings spearheaded by the White House Office of Science and Technology Policy and the National Economic Council.[2] Each event examined AI from a different perspective, ranging from law and governance,[3] to AI safety and control,[4] to the potential use of AI for public good,[5] to the potential futures of AI.[6] *AI Now* complemented these approaches by focusing on the social and economic impacts of AI over the next ten years, offering space to discuss some of the hard questions we need to consider, and to gather insights from world-leading experts across diverse disciplines. We asked questions such as: what are the present social and economic issues raised by the rapid deployment of AI, and how can we build a deeper understanding of AI in the contemporary moment that will help us create a fair and equitable future?

To focus and ground the discussion, *AI Now* looked at AI in relation to four key themes: **Healthcare**, **Labor**, **Inequality**, and **Ethics**.

In choosing these themes, we selected the spheres of labor and healthcare to represent critical social systems where early AI systems are already in use, and which present complex and urgent challenges that span many dimensions of society both domestically and globally. We chose the themes of social inequality and ethics because they are overarching concerns for the field. Will the deployment of AI systems increase or

---

1 As Russell and Norvig point out, the history of artificial intelligence has not produced a clear definition of AI but rather 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." In the context of this primer on labor, we are relying 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 Ed Felten, "Preparing for the Future of Artificial Intelligence," The White House Blog, May 3, 2016,

3 University of Washington public workshop, "Artificial Intelligence: Law and Policy," May 24, 2016,

4 Carnegie Mellon University, "Workshop on Safety and Control for Artificial Intelligence," June 28, 2016,

5 Computing Community Consortium special event, "Artificial Intelligence for Social Good," June 7, 2016

6 Stanford University, The Future of Artificial Intelligence, June 23, 2016

decrease social inequality, and can we craft ethical practices to ensure that the benefits of AI are widely dispersed?

By drawing attention to the near-term stakes and emerging uses of AI systems, and by inviting experts from across domains to lend valuable perspectives, *AI Now* began a crucial conversation focused on developing practices of assessment, study, and verification today, with the goal of ensuring beneficial AI in the future.

Key recommendations:

Based on the generous contributions of those who brought their insight and expertise to *AI Now*,[7] and based on the research literature across relevant disciplines, we make the following high-level recommendations. It is important to note that while these recommendations draw on the intellectual generosity of all *AI Now* attendees, they do not reflect the views of any particular *AI Now* participant, participating organization, host, or sponsor.

We offer these recommendations as practical steps that stakeholders engaged at various points in the production, use, governance, and assessment of AI systems could take to address near-term challenges and opportunities created by the rapid deployment of AI through social and economic domains.

1. **Problem**: Developing and deploying AI systems requires significant infrastructure and data resources. This limits the opportunities for AI innovation and derived insight to those with access to such resources. It also limits competition if costs are prohibitive for new entrants.
   Recommendation: Diversify and broaden access to the resources necessary for AI development and deployment – such as datasets, computing resources, education, and training, including opportunities to participate in such development. Focus especially for populations that currently lack such access.

2. **Problem:** Early AI systems are already augmenting human work and changing management structures across labor sectors. Evidence from participants, including Jason Furman, President Obama's chairman of the Council of Economic Advisers, suggests that low-skill jobs are at the greatest risk of replacement from automation and AI. This raises important questions about existing social safety nets and human resource distribution in a world where machines do more of the work.
   Recommendation: Update definitions and frameworks that specify fair labor practices to keep pace with the structural changes occurring as AI management systems are deployed into the workplace. Also investigate alternative modes of income and resource distribution, education, and retraining that are capable of responding to a future in which repetitive work is increasingly automated and the

[7] *AI Now* attendees and speakers list: https://artificialintelligencenow.com/schedule/workshop/attendees

dynamics of labor and employment change.

3. **Problem:** AI and automated decision making systems are often deployed as a background process, unknown and unseen by those they impact. Even when they are seen, they may provide assessments and guide decisions without being fully understood or evaluated. Visible or not, as AI systems proliferate through social domains, there are few established means to validate AI systems' fairness, and to contest and rectify wrong or harmful decisions or impacts.
   Recommendation: Support research to develop the means of measuring and assessing AI systems' accuracy and fairness during the design and deployment stage. Similarly, support research to develop means of measuring and addressing AI errors and harms once in-use, including accountability mechanisms involving notification, rectification, and redress for those subject to AI systems' automated decision making.[8] Such means should prioritize notifying people when they are subject to automated decision-making, and developing avenues to contest incorrect or adverse judgments. In addition, investigate what the ability to "opt out" of such decision making would look like across various social and economic contexts.[9]

4. **Problem:** Current U.S. legal restrictions, such as the Computer Fraud and Abuse Act (CFAA) and the Digital Millennium Copyright Act (DMCA), threaten to make illegal forms of fundamental research necessary to determine how the fairness and accountability of public and private institutional systems will be impacted by the increased deployment of AI systems.
   Recommendation: Clarify that neither the Computer Fraud and Abuse Act nor the Digital Millennium Copyright Act intend to restrict research into AI accountability.

5. **Problem**: We currently lack agreed upon methods for measuring and assessing the social and economic impacts of AI systems, even as they are rapidly deployed through core social domains, such as health, labor, and criminal justice.
   Recommendation: Support fundamental research into robust methodologies for assessment and evaluation of the social and economic impacts of AI systems deployed in real-world contexts. Work with government agencies to integrate these new techniques into their investigative, regulatory, and enforcement capacities.

6. **Problem:** Those directly impacted by the deployment of AI systems within core social and economic domains rarely have a role in their design, or a means to alter their assumptions and uses.
   Recommendation: Work with representatives and members of communities

[8] For a discussion of efforts to increase the transparency of AI systems, see Eric Horvitz, "Reflections on Safety and Artificial Intelligence," Exploratory Technical Workshop on Safety and Control for AI, Carnegie Mellon University, June 27, 2016

[9] For an example of early move in the direction of notice and opt-out, see especially Article 21 and 22: European Union. European Parliament, Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. Regulation. Brussels, 2016.

impacted by the application of automated decision making and AI systems to codesign AI with accountability, in collaboration with those developing and deploying such systems.

7. **Problem**: AI systems are being deployed within key social and economic domains, each with its own structure, history, and existing protocols. However, the focus of AI research is primarily technical, with fewer contributions from disciplines that prioritize human contexts, experience, and sociopolitical issues, such as the social sciences and humanities. Further, the computer science subfield of AI is heavily dominated by men with largely homogeneous racial and ethnic backgrounds. This can limit the perspectives and experiences of AI's creators, and lead to a greater possibility of "like me" bias.[10]
   Recommendation: Increase efforts to improve diversity among AI developers and researchers, and broaden and incorporate the full range of perspectives, contexts, and disciplinary backgrounds into the development of AI systems. The field of AI should also support and promote interdisciplinary AI research initiatives that look at AI systems' impact from multiple perspectives, combining the computational, social scientific, and humanistic.

8. **Problem:** Professional codes of ethics, where they exist, don't currently reflect the social and economic complexities of deploying AI systems within critical social domains like healthcare, law enforcement, criminal justice, and labor. Similarly, technical curricula at major universities, who recently emphasizing ethics, rarely integrates these principles into its core training at a practical level.
   Recommendation: Work with professional organizations such as The Association for the Advancement of Artificial Intelligence (AAAI), The Association of Computing Machinery (ACM), and The Institute of Electrical and Electronics Engineers (IEEE) to update (or create) professional codes of ethics that better reflect the complexity of deploying AI and automated systems within social and economic domains. Mirror these changes in education, making courses in civil rights, civil liberties, and ethics required training for anyone wishing to major in Computer Science. Similarly, update professional codes of ethics governing the professions into which AI systems are being introduced, such as those applied to doctors and hospital workers.

AI Now, across four key themes

We now turn to summaries of the four key topic areas on which *AI Now*'s discussion and debate focused. These provide a window into insights and recommendations shared by assembled experts, along with brief overviews of general challenges, opportunities, and potential interventions within each thematic area.

[10] Jack Clark, "Artificial Intelligence Has a 'Sea of Dudes' Problem," Bloomberg, June 23, 2016

1. Social Inequality

How might AI systems contribute to unfair bias and discrimination?

As AI systems take on a more important role in high-stakes decision-making – from offers of credit and insurance, to hiring decisions and parole – they will begin to affect who gets offered crucial opportunities, and who is left behind. This brings questions of rights, liberties, and basic fairness to the forefront.

While some hope that AI systems will help to overcome the biases that plague human decision-making,[11] others fear that AI systems will amplify such biases, denying opportunities to the deserving and subjecting the deprived to further disadvantage.[12]

Within this debate, data will play a crucial role and deserves keen attention. AI systems depend on the data they are given, and may reflect back the characteristics of such data, including any biases, in the models of the world they create. As such, considerations of the impacts of AI systems are closely tied to existing discussions around big data, such as those recognized in the 2014 and 2016 White House reports on the topic.[13]

Broadly speaking, bias in data takes two forms. One occurs when the data available is not an accurate reflection of the reality it's supposed to represent (due to inaccurate measurement methodologies, incomplete data gathering, non-standardized self-reporting, or other flaws in data collection); the second happens when the underlying process being measured for the purpose of data collection *itself* exhibits long-standing structural inequality (for example: if data on job promotions gathered from an industry that systematically promoted men over women were used to understand which characteristics predict career success). 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. It is important to note that while there are communities doing wonderful work on these issues, there is no consensus on how to "detect" bias of either kind.[14]

When data reflects biases of either form, there is the risk that AI systems trained on this data will produce models that replicate and magnify those biases. In such cases, AI systems would exacerbate the discriminatory dynamics that create social inequality, and would likely do so in ways that would be less obvious than human prejudice and implicit

---

[11] Future of Privacy Forum and Anti-Defamation League, *Big Data: A Tool for Fighting Discrimination and Empowering Groups*, Sep 11, 2014.

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

[13] The White House Office of Science and Technology Policy, *Big Data: Seizing Opportunities, Preserving Values*, May 2014; President's Council of Advisors on Science and Technology, *Big Data and Privacy: A Technical Perspective*, May 2014; The White House. *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*, May 2016.

[14] For example, the Fairness, Accountability, and Transparency in Machine Learning community

bias (and that could be justified as the work of "intelligent technology," and thus potentially more difficult to contest).[15]

With the introduction and expanded use of AI systems in domains of collective risk management, the capacity to draw more fine-grained distinctions between people will increase, potentially endangering the solidarity upon which insurance and other pooled social safety nets rely. The ability to precisely know who within a population will be more or less costly may allow companies to engage in far more effective "adverse selection," identifying particular populations and individuals that they would profit by avoiding or under-serving.[16]

In the case of health insurance, policyholders who possess traits or exhibit behaviors that AI systems recognize as predictors of the future onset of certain diseases or disorders may 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 such predictions are accurate and the insurer's behavior is rational, the effects are ultimately perverse.[17]

We may be undoing an important feature of insurance by giving organizations extraordinarily precise information that could allow them to make inferences that will deny care or limit coverage to people in need. Even the President's Council of Economic Advisors predicts that these dynamics will very likely apply to an array of arenas with risk-based price discrimination, and that competition will likely push insurers down this path.[18]

Ultimately, the use of AI systems within domains like insurance raises a basic question about the line that divides objectionable discrimination from the reproduction of inequality.

While normative principles enshrined in discrimination law might address these practices, such regulations might not be the most efficient, effective, or fair way to remedy what is ultimately a problem of economic inequality. Commitments to fairness by those who design and deploy of AI systems are also important. Nevertheless, discrimination law might be one important vehicle for ensuring that AI does not contribute to a more unequal society.

---

[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] For an examination of precision profiling and targeting in the online advertising context, see Joseph Turow, *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*, Yale University Press, 2013.

[17] Upturn, "As Insurers Embrace Big Data, Fewer Risks Are Shared," *Civil Rights, Big Data, and Our Algorithmic Future*, Sept 2014.

[18] Council of Economic Advisers, *Big Data and Differential Pricing* (2015).

Investigations into discrimination, fairness, and inequality in AI systems, however, may be impeded if legal frameworks limit or prohibit such inquiries. For example, in the U.S., aggressive interpretations of laws such as the Computer Fraud and Abuse Act (CFAA) and the Digital Millennium Copyright Act (DMCA) threaten to chill research in this space. Both are currently under challenge and in need of reform to ensure that necessary research is exempt from their prohibitions.[19]

Will AI systems benefit the many or the few?

AI systems are already creating new ways to generate economic value while also affecting how this value is distributed. To the extent such distributions favor certain incumbent entities or populations, AI systems are likely to perpetuate or exacerbate existing wage, income, and wealth gaps.

The composition of those able or allowed to develop and implement AI systems is a fundamental area where inequality may be exacerbated. AI is predicted to become a multi-billion dollar a year industry. Developing AI requires significant up-front investment, including sizeable computational resources, along with large amounts of data, both of which are costly. If the ability to develop and implement AI is not available broadly, a very narrow segment of the world will claim the lion's share of the economic surplus AI generates. 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.[20]

At the same time, if AI and automation systems reduce the costs of creating certain goods and services (and if these savings are passed on to consumers), they could narrow the gap between the haves and have-nots (at least in terms of access to these goods and services).[21] In this case, AI systems would increase the standard of living overall, and could even have a progressive redistributive effect.

AI systems 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 offer affected workers an easy transition to other forms of paid labor. Rather than simply replacing workers or reducing workers' share of the profits from productive activity, AI systems that ease the toil of labor could free people to pursue new forms of living and work, increasing overall welfare in society.

---

[19] https://www.aclu.org/cases/sandvig-v-lynch-challenge-cfaa-prohibition-uncovering-racial-discrimination-online; https://www.eff.org/press/releases/eff-lawsuit-takes-dmca-section-1201-research-and-technology-restrictions-violate; http://www.theatlantic.com/politics/archive/2015/04/aarons-law-reintroduced-as-lawmakers-wrestle-over-hacking-penalties/458535/.

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

[21] 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,

Still, as some commentators have observed, AI systems may render certain workers' skills redundant, leaving those who have been replaced by automation with few options for alternative paid employment.[22] Even if workers are able to 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 and automation systems.[23]

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 systems increase inequality, they might push certain parts of the workforce into permanent unemployment and poverty. This is why understanding AI systems' potential impact on labor is an important aspect of understanding their impact on economic equality, and preparing accordingly.

These 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 needed to cover life's necessities.[24] UBI's supporters see it as a way to address whatever ongoing unemployment AI systems 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.[25] Crucially, UBI would aim to provide an income floor beneath 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.[26]

Like all technologies before it, AI systems reflect the values of their creators, and there is hope that increased diversity among those developing, deploying, and maintaining AI systems may help create a future in which these technologies promote equality.[27] Currently, however, women and minorities continue to be under-represented in the field of AI particularly, and in computer science overall.

This lack of inclusion, and the "like me" bias that it can perpetuate, may limit the degree to which practitioners choose to recognize or prioritize the concerns of other

---

22 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.

23 Henry Siu and Nir Jaimovich, "Jobless Recoveries," Third Way (April 2015); Roosevelt Institute, "Technology and the Future of Work: The State of the Debate," Open Society Foundations Future of Work Project (April 2015).

24 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*, Jan 27, 2016.

25 Eduardo Porter, "A Universal Basic Income Is a Poor Tool to Fight Poverty," *The New York Times*, May 31, 2016.

26 Jathan Sadowski, "Why Silicon Valley is embracing universal basic income," *The Guardian*, June 22, 2016.

27 Kate Crawford, "Artificial Intelligence's White Guy Problem," *New York Times*, June 25, 2016.

communities. Amid growing recognition that heterogeneous groups outperform those that are more homogeneous,[28] there is also good reason to believe that increased diversity in the field of AI will help AI systems serve the interests of diverse populations. To address pressing issues of bias, discrimination, and inequality, the AI community will need to draw on a much broader range of perspectives.[29]

2. Labor

Too often, discussions of labor and AI systems have focused only on the fear of a jobless future. Current research demonstrates that more complex and more immediate issues are at stake, affecting not only labor markets in the abstract, but also employer-employee relations, power dynamics, professional responsibility, and the role of work in human life.

Many traditional economic researchers are closely tracking national labor markets and institutions as a means to examine the impact of AI systems.[30] This kind of research provides important qualitative data and facilitates an understanding of macroeconomic trends and the supply and demand of labor, i.e. how many jobs there will be. Complementing this view, social scientific research has examined *how* the shifting nature of jobs and work dynamics are changing people's everyday lived experience. Both perspectives are necessary to appreciate the social and economic impact of AI systems on labor in the near term.

Will AI impact the demand for jobs?

The role of automation[31] in the economy is far from a new subject of inquiry, and considerations of the impact of AI systems emerge from within long-standing debates.

While intuitively it may seem the demand for labor would decline as automation increases because there would be only a finite amount of work to be done – and some economists argue this – others don't see it this way, referring to this assertion as the "lump of labor" fallacy.[32] These dissenting economists suggest that as productivity rises in one industry (due to automation or other factors), new industries emerge along with new demand for labor. For example, in 1900 agriculture comprised 41 percent of the United States' workforce. By 2000, agriculture comprised only 2 percent. Labor economists David Autor and David Dorn point out that even with this dramatic change, unemployment has not increased over the long-term and the employment-to-population

---

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

[29] For example, Women in Machine Learning.

[30] Roosevelt Institute, "Technology and the Future of Work: The State of the Debate," White paper, Open Society Foundations Future of Work Project (April 2015).

[31] Automation is defined here as "a device or system that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator." At this broad level, technologies of automation and artificial intelligence are interconnected. Raja Parasuraman et al., "A Model for Types and Levels of Human Interaction with Automation," *IEEE Transactions on Systems, Man and Cybernetics* 30(3) (2000).

[32] Tom Standage, "Artificial Intelligence: The return of the machinery question," *The Economist*, June 25 2016.

ratio has in fact grown.[33] Other economists, such as James Huntington and Carl Frey, are more dire in their prediction that AI systems will dramatically reduce the number of jobs available.[34]

Still others debate whether the transformations and fluctuations in labor markets are related to technology at all, or instead caused by economic policy. Such arguments focus on what role existing institutions and regulatory mechanisms should play in relation to AI and automation systems. Robert Gordon, for example, argues that the current waves of innovation are not as transformative as they seem.[35] Contrary to this view, many are beginning to concur that labor markets are undergoing a consequential transformation due to technological change. These include Joseph Stiglitz and Larry Mishel, who argue that a keen focus must be maintained on regulation and other policy changes in relation to AI and automation systems in order to protect workers.[36]

Economists Autor and Dorn, among others, have observed a significant rise of "job polarization," in which middle-skill jobs decrease and leave some high-skill and more low-skill jobs. While new jobs may be created, these jobs are often low-paid and undesirable.[37]

For example, many of the jobs that support AI systems are, in fact, performed by humans who are required to maintain these infrastructures and tend to the "health" of the system. This labor is often invisible, at least in the context of the popular stories and ideas of what AI is and what it does. It is also, consequently, undervalued. Such jobs include janitors that clean offices, maintenance or repair workers who fix broken servers, and what one reporter termed "data janitors," those who "clean" the data and prime it for analysis.[38]

The questions to ask about the impact of AI systems on labor should include not only whether jobs will be created, but whether they will be *desirable* jobs that pay a living a wage.

---

[33] David Autor and David Dorn, "How Technology Wrecks the Middle Class," *New York Times* April 24 2013.

[34] Carl B. Frey and Michael Osbourne, "The Future of Employment: How Susceptible Are Jobs to Computerization" Oxford Martin School Programme on the Impacts of Future Technology Working Paper, (September 17, 2013).

[35] Robert J. Gordon, "The Demise of U.S. Economic Growth: Restatement, Rebuttal and Reflections," National Bureau of Economic Research Working Paper, (February 2014)

[36] Lawrence Mishel, John Schmitt and Heidi Shierholz, "Assessing the Job Polarization Explanation of Growing Wage Inequality," EPI Working Paper Paper (January 2013). See also Joseph Stiglitz, *The Price of Inequality: How Today's Divided Society Endangers Our Future*, New York: W.W. Norton and Co, 2012.

[37] For instance, see Henry Siu and Nir Jaimovich, "Jobless Recoveries," Third Way, April 8 2015. See also Tom Standage,"Artificial Intelligence: The return of the machinery question," *The Economist*, 25 Jun 2016.

[38] Steve Lohr, "For Big-Data Scientists, 'Janitor Work' is Key Hurdle to Insights." *New York Times* August 17 2014. While most media discussions elide this kind of work or frame it as an obstacle that will soon be performed by computers, scholar Lilly Irani has pointed out that this kind of work is integral to AI systems, describing it as "the hidden labor that enables companies like Google to develop products around AI, machine learning, and big data." Lilly Irani, "Justice for 'Data Janitors," *Public Books,* (January 2015).

Moreover, while discussions about the future of AI systems and labor tend to focus on what are traditionally conceived of as low-wage, working class jobs like manufacturing, trucking, and retail or service work, research demonstrates that there will be impacts across sectors, including professional and expert work that requires specialized training or advanced degrees, such as radiology or law.[39] In this context, new issues of professional responsibility and liability will need to be addressed.

How will AI impact employer-worker relationships?

In recent years, researchers have begun to examine how AI and automation systems that rely on big data (from Uber to automated scheduling software used in large retailers to workplace surveillance) are changing the traditional dynamics between employers and employees.[40]

Findings suggest that while such systems could be designed to empower workers, there are substantial ways in which these technologies can disempower workers, entrench discrimination, and generate unfair labor practices.[41]

For instance, AI-driven workforce management and scheduling systems are increasingly used to remotely manage labor, contributing to the growth of the on-demand economy and the rise of what has been called the "precariat." Though some researchers have concluded that just-in-time scheduling provides valuable flexibility, by far more research has documented the strain and uncertainty experienced by workers subject to these systems.[42]

The adverse experiences of workers managed by such systems include chronic underemployment, financial instability, insufficient benefits and protections that are traditionally granted full-time employees, as well as the structural inability to plan for family or self care (or even search for another job given the round-the-clock availability that such jobs often demand from workers). Moreover, the workers most likely to be affected by these practices are disproportionately women and minorities.[43]

Adding to this, new modes of remote management via AI system can make it harder to hold employers accountable for decisions made by "the system" that materially impact employees. This can leave workers vulnerable to exploitation.

---

[39] Dana Remus and Frank Levy, "Can Robots Be Lawyers? Computers, Lawyers, and the Practice of Law," Working paper, Dec 30, 2015.

[40] For a review of recent work, see Min Kyung Lee et al., "Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers," *CHI 2015 Proceedings* (April 2015).

[41] See, for instance: Julia R. Henly, H. Luke Shaefer, and Elaine Waxman, "Nonstandard Work Schedules: Employer- and Employee-Driven Flexibility in Retail Jobs," *The Social Service Review* 80(4): 609-634 (2006); Leila Morsy and Richard Rothstein, "Parents' Non-Standard Work Schedules Make Adequate Childrearing Difficult," Economic Policy Institute Issue Brief (Aug. 6, 2015). Additional research and examples are discussed below.

[42] Farhad Manjoo, "Uber's Business Model Could Change Your Work." *New York Times* 28 June, 2015.

[43] Carrie Gleason and Susan Lambert, "Uncertainty by the Hour," Position paper. Open Society Foundation Future of Work Project (2014).

For example, platforms like Uber (powered by big data and AI) serve to remotely control routes, pricing, compensation, and even interpersonal norms – determinations and decisions that would traditionally be in the hands of human management.[44]

Beyond simply obscuring the face and reasoning behind a given decision, this type of remote management is very often not recognized as "employee management" at all.

Because these new modes of management do not fit cleanly into existing regulatory models, companies like Uber market themselves as technology companies, not managers of workers. Following this, such companies view themselves as a marketplace to facilitate connection, and thus not responsible to workers in the way a traditional employer would be. In this configuration, workers end up assuming the risks of employment without guaranteed benefits (such as decreased tax burdens, healthcare, and other workplace protections) or potential modes of redress.

3. Healthcare

AI systems in healthcare, like the current AI systems overall, rely on large datasets, using complex statistical modeling and machine learning to generate insights from such data. Existing (and growing) sources of health data – including electronic health records (EHRs), clinical and insurance databases, and patient-produced data from consumer devices and apps[45] – are already being used by a variety of AI implementations promising the potential to improve healthcare from diagnostics, to patient care, to drug discovery, to the production, organization, and communication of medical information.[46]

How will AI be integrated into medical research and healthcare?

The integration of AI systems into medical research presents exciting prospects for understanding the foundations of illness, developing new treatments, making more finely grained diagnoses, and even tailoring medicine to specific individuals. However, such research will require caution, as existing limitations and biases could impact the ability to deliver on these promises.

Such limitations include incomplete or inaccurate research datasets, which may exclude certain minority populations, along with the complex financial incentives especially prevalent in the US healthcare system that, for instance, favor the development of certain drugs over others, or incentivize one treatment option over another.

---

[44] Alex Rosenblat and Luke Stark, "Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers," *International Journal of Communication* 10 (2016): 3758-3784.

[45] Electronic Health Records (EHRs) are quickly becoming a part of dynamic digital systems that hold multiple forms of patient information, from providers' notes and clinical assessments to lab test results and medication prescriptions. EHRs can also include non-patient specific information such as clinical guidelines, medical research literature, and scientific news.

[46] For research exploring these potential benefits, see: Wiens, Jenna, John Guttag, and Eric Horvitz. "Patient Risk Stratification with Time-Varying Parameters: A Multitask Learning Approach." Journal of Machine Learning Research 17, no. 79 (2016): 1-23. And: Bayati, Mohsen, Mark Braverman, Michael Gillam, Karen M. Mack, George Ruiz, Mark S. Smith, and Eric Horvitz. "Data-driven decisions for reducing readmissions for heart failure: General methodology and case study." PloS one 9, no. 10 (2014): e109264.

Medical research data often presents itself as objective and universal, while in reality its findings may be partial, temporary, or specific to only some communities or contexts.[47] AI systems analyzing and building models from such data could introduce, entrench, or extend incorrect assumptions. This outcome is not inevitable, however. AI systems using data that does not present these problems (assuming that can be assured), or where the framework for use has corrected for these issues, offer the potential to actually mitigate the biases that may be inherent in, for instance, the existing constitution of randomized control trials (RCTs) or other public health databases.[48]

Assuming accuracy and freedom from bias, one of the most hopeful prospects for AI systems' integration in healthcare research and practice centers on its ability to aid diagnostics, finding patterns in vast quantities of data that lead to early detection of tricky diseases. Indeed, AI systems are already accomplishing diagnostic feats, including detecting leukemia.[49] In screening and clinical care settings, AI systems also present opportunities to mitigate or even prevent misdiagnoses, which can be deadly as well as costly.[50]

In this capacity, AI systems are gaining a more central role in establishing and identifying what illness is. It is here that attention must be paid to the faulty assumptions that may underlie constructions of "normal" or "average" health.[51] We need only to look to a time before 1973, when The American Psychiatric Association listed homosexuality in its authoritative Diagnostic and Statistical Manual of Mental Disorders, to understand the stakes here.[52]

Similarly, as AI systems are integrated directly into patient care, where they may be used in diagnostics and clinical management, often interposed between caregiver and patient, it will be important to appropriately define the limits of their "expertise." A human surgeon goes to medical school and must pass rigorous exams to become board certified: how do we establish the competency of an AI system designed to augment or replace a

---

47 For example, research recently published in the journal *PNAS* has raised attention to issues with inferential statistical models and software bugs that likely invalidate many thousands of research papers within a particular field of fMRI-based research. Neuroskeptic blog, "False-Positive fMRI Hits The Mainstream," *Discover Magazine*, July 7, 2016.

48 For an overview of the potential limitations of data generated from RCTs see Peter M Rothwell, "Factors That Can Affect the External Validity of Randomised Controlled Trials," *PLoS Clin Trials* vol 1 iss 1 (May 2006).

49 Jon Fingas, "IBM's Watson AI Saved a Woman from Leukemia," Engadget, August 7, 2016.

50 AI-based approaches have demonstrated great potential to reduce the harms and costs associated with hospital readmissions for heart failure. Mohsen Bayati et al., "Data-Driven Decisions for Reducing Readmissions for Heart Failure: General Methodology and Case Study," *PloS One* vol 9 no. 10 (2014). For recent directions in screening using AI, see: Paparrizos, John, Ryen W. White, and Eric Horvitz. "Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results." Journal of Oncology Practice (2016): JOPR010504.

51 The history of researching and treating heart disease is a case in point. Numerous studies have provided evidence that standards around heart disease, from education to diagnosis to medication, have been based on male bodies to the detriment of the diagnosis and treatment of heart disease in women. If "normal" or "standard" are equated with only a subset of people, then those who fall outside this subset may not benefit from treatments based on those standards. Vidhi Doshi, "Why Doctors Still Misunderstand Heart Disease in Women," *The Atlantic Monthly*; Richard J McMurray et al. "Gender disparities in clinical decision making." *JAMA* 266 no 4 (1991): 559-562.

52 Drescher, Jack. "Out of DSM: depathologizing homosexuality." Behavioral Sciences 5, no. 4 (2015): 565-575.

certified human? This, even as AI systems are often presented as infallible expert authorities, assumed to be free from error or bias. Such reliance means that these systems could be subject to less scrutiny, both in terms of a priori assessments of competence and in measuring ultimate performance. This creates possibilities for new kinds of harm not covered within existing ethical frameworks.[53]

Moreover, looking at where and for whose benefit AI systems are being deployed within the healthcare domain should be a matter of attention. Despite a desire to make healthcare accessible and affordable to all, substantial evidence shows that access to healthcare and health outcomes are unequally distributed, with poor, non-white, and female populations often systematically disadvantaged.[54]

The introduction of AI systems will not automatically correct these systemic inequalities, and has the potential to amplify them. While AI systems could enable appropriately specialized care that benefits diverse populations, they could also leave behind traditionally ignored, underserved, or outlier populations. If these populations aren't given appropriate consideration, this could result in models built by AI systems, and reinforced with data from patients privileged enough to have access to such AI systems, that reflect the conditions *only* of the fortunate, and ultimately configure an aggregate understanding of health and illness that fundamentally excludes the marginalized.

This concern deserves attention, especially given the existing and complex economics of healthcare in the United States, which will certainly influence the deployment and impact of AI systems, as it has influenced the integration of medical technologies in the past.

Within this context, ongoing efforts to reduce costs while simultaneously promoting growth has led stakeholders (such as government actors, insurance companies, health institutions, pharmaceutical companies, employers, and others) to pin their hopes on large-scale data collection and now to AI systems to help sustain their economic models of research and care.[55] Many such efforts are focused on laudable goals, such as enabling more accurate clinical decisions or improving preventative care.

However, the significant training, resources, and ongoing maintenance required to integrate these information technology and AI systems into hospitals and other healthcare settings is not always supported, or clearly budgeted. This has already led to an unequal distribution of technological resources and capabilities.[56]

---

[53] Nicholas Carr, "Automation Makes Us Dumb," *Wall Street Journal*, Nov 21, 2014.

[54] Kaiser Family Foundation, "Health Coverage by Race and Ethnicity: The Potential Impact of the Affordable Care Act," Report, Washington, D.C., 2013; John Ayanian, "The Costs of Racial Disparities in Health Care," *Harvard Business Review*, Oct 1, 2015.

[55] Claudia Grossmann et al., *Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good*, Workshop Summary, Institute of Medicine of the National Academies, Washington DC: National Academies Press, 2010.

[56] For instance, see American Hospital Association, *Adopting Technological Innovation in Hospitals: Who Pays and Who Benefits*, Report. American Hospital Association, Washington, DC, 2006.

How will AI's need for increased data collection and patient observation impact privacy?

AI systems' reliance on large datasets and situational observations raises pressing issues of patient privacy, confidentiality, and security.[57]

Current expectations for AI systems depend on making an increasing amount of patient data available for use across various devices, platforms, and networks. This brings with it opportunities to conduct potentially intrusive surveillance, for profit or otherwise. While technologies such as homomorphic encryption, differential privacy, and stochastic privacy hold hope for allowing AI systems to "use" data without "seeing" it, these are still in their early stages, and general-purpose applications have yet to be developed. In the meantime, the economic emphasis on surveillance and consumption of sensitive health data will only increase with the recent moves to promote evidence-based medicine and the Affordable Care Act's shift from a fee-for-service to a pay-for-performance model.[58] Insurers may also feel increased pressure to justify cross-subsidization models in the context of AI systems. For example, despite prohibitions in the Genetic Information Nondiscrimination Act of 2008, there is growing interest in using genetic risk information for insurance stratification.[59] In fact, differential pricing has become one of the standard practices for data analytics vendors, introducing new avenues to perpetuate inequality.[60]

The shift to ubiquitous tracking and surveillance through "smart" devices and other networked sensors that accompanies AI systems' need for data is already pushing the limits of existing privacy protections, such as the Health Insurance Portability and Accountability Act (HIPAA).[61] Risks that patients will be re-identified from granular-level data or have their identities, illnesses, or other health information predicted through proxy data will only increase as AI systems become integrated into health and consumer products.[62]

Moreover, the software that powers these data-collection devices is often proprietary, rather than open source (i.e. open to external scrutiny and auditing).[63] While a

---

[57] Helen Nissenbaum and Heather Patterson, "Biosensing in Context: Health Privacy in a Connected World," in Dawn Nafus (ed.), *Quantified: Biosensing Technologies in Everyday Life*, Cambridge, MA: MIT Press, 2016.

[58] David Blumenthal, Melinda Abrams, and Rachel Nuzum, "The Affordable Care Act at 5 Years," *New England Journal of Medicine* Vol. 372, Issue 25, (2015): 2453.

[59] Yann Joly et al., "Life Insurance: Genomic Stratification and Risk Classification," *European Journal of Human Genetics* 22 No. 5 (May 2014): 575–79).

[60] Jason Furman and Tim Simcoe, "The Economics of Big Data and Differential Pricing," The White House Blog, Feb 6, 2015.

[61] Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization," *UCLA Law Review* 57 (2010): 1701-1778.

[62] Computer scientist Latanya Sweeney has demonstrated that the vast majority of Americans can be identified using only three pieces of perceived "anonymized data": ZIP code, birthdate, and sex. Latanya Sweeney, "Simple Demographics Often Identify People Uniquely," Carnegie Mellon University, Data Privacy Working Paper 3, Pittsburgh: PA, 2000. See also, Bradley Malin, Latanya Sweeney and Elaine Newton, "Trail Re-identification: Learning Who You are From Where You Have Been," LIDAP-WP12. Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA, March 2003. Latanya Sweeney, "Matching Known Patients to Health Records in Washington State Data," Harvard University. Data Privacy Lab, White Paper 1089-1, Cambridge, MA, June 2013,

[63] For a discussion of the implications of such proprietary systems, see Frank Pasquale, *The Black Box Society: The secret algorithms that control money and information*, Cambridge: Harvard University Press, 2015.

recently-granted exemption to the Digital Millennium Copyright Act provides the opportunity to examine code for external medical devices, it may be even more important to examine internal medical devices, which are currently excluded from the exemption.[64]

Overall, experts have warned of grave security issues in the deployment of networked technologies across Internet of Things devices, many focusing on medical devices as specifically problematic.[65]

How will AI impact patients and healthcare providers?

The existing and potential applications of AI technologies have profound implications for the constitution of care-work and what it means to care for sick or vulnerable bodies.

Many visions for AI systems place them as a mediator of care work, potentially supplanting caregivers entirely. This move, while promising economic efficiencies, may change the relationships and norms between patients and doctors or other caregivers.

Common examples used to illustrate the potentials of AI systems to replace or supplement human caregivers range from robot surgeons to virtual avatars to companion robots.[66] Such examples have catalyzed important debates about the social implications of delegating care and companionship to non-human agents.[67] What kinds of equivalences are being made when machines replace, not just augment, human professionals? When we deem a machine capable of "care," what capacities are we assuming, and what definition of "care" are we using? Are these equivalences in the best interests of patients?

While companion robots are not supplanting human caretakers by any notable degree, at least not yet, AI-driven apps and networked devices promising patient empowerment and the ability for patients to take control of their health are proliferating, signaling the early stages of AI systems' direct interface with patients.

This direct interface brings with it both benefits, like healthier, more self-aware patients, and risks. Such risks include potentially misleading patients about the quality and precision of information they may be receiving, a concern the FTC has attempted to address in recent years.[68] In addition, such apps may serve to effectively shift the responsibility for care and monitoring from healthcare professionals to patients

---

64 Karen Sandler, Lysandra Ohrstrom, Laura Moy and Robert McVay, "Killed by Code: Software Transparency in Implantable Medical Devices," Software Freedom Law Center, New York, 2010.

65 A team at University of South Alabama demonstrated these issues in 2015, successfully hacking the iStan brand networked pacemaker and "killing" a mannequin that had the pacemaker installed. William Bradley Glisson et al. , "Compromising a Medical Mannequin," *arXiv preprint, arXiv:1509.00065,* 2015.

66 Laurel Riek, "Robotics Technology in Mental Healthcare," in D. Luxton (Ed.), *Artificial Intelligence in Behavioral Health and Mental Health Care*. New York: Elsevier, 2015: 185-203.

67 Nick Bilton, "Disruptions: Helper Robots Are Steered, Tentatively, to Care for the Aging," *New York Times* May 19, 2013.

68 FTC, "FTC Cracks Down on Marketers of "Melanoma Detection," Federal Trade Commission, Press Release, Feb 23, 2015.

themselves. This may disadvantage patients who do not have the time, resources, or access to technology to manage their own care.

What kinds of patients are favored in this new dynamic, and might patients not well-equipped to manage and maintain their own data receive substandard care? Moreover, what new roles and responsibilities do the designers and developers of such apps take on, and how do the ethical responsibilities at the heart of the medical profession get integrated into these differing design and engineering contexts?

4. Ethical responsibility

The deployment of AI systems will introduce new responsibilities as well as pose challenges to existing frameworks in areas such as professional ethics as well as research ethics or even public safety evaluations.

Recent discussions of ethics and AI systems have tended to prioritize the challenges posed by hypothetical general AI systems in a distant future, such as the advent of the "singularity" or the development of a superintelligence that might become an existential threat to humanity.[69]

Discussions of AI systems focusing on such speculative futures have tended not to focus on the immediate ethical implications of AI systems in the near- to medium-term, including the challenges posed by the enormous number of task-specific AI systems currently in use, and the means by which these could amplify existing inequalities, and fundamentally shift power dynamics.[70]

Contemporary AI systems perform a diverse range of activities, and poses new challenges to traditional ethical frameworks due to the implicit and explicit assumptions made by these systems, and the potentially unpredictable interactions and outcomes that occur when these systems are deployed in human contexts.

Given the profound implications of AI systems in terms of resource allocation and the potential to concentrate or restructure power and information, key questions need to be urgently addressed to ensure these systems are not harmful, especially to already marginalized groups.

---

[69] The ethical questions regarding the potential use of lethal force by autonomous robots need to be addressed, but are beyond the scope of this paper. For an overview of current debates see: Paul Scharre and Michael C. Horowitz, "An Introduction to Autonomy in Weapon Systems," Center for New American Security Working Paper (2015). While the use of lethal force by autonomous (or intelligent) systems is necessarily particular to the context of international relations and international law, the concept of "meaningful human control" is a concept relevant to other arenas in which autonomous or intelligent systems could potentially be making critical decisions. See below for a discussion of distributed, ethical control.

[70] "Task-specific" here does not imply any statement about complexity. The spell-checker in your word processor and the program that governs a "self-driving" car are both task-specific systems even though the complexity of the models used in these two tasks is very different. For context on superintelligences versus near-term ethical challenges, see Kate Crawford, "Artificial Intelligence's White Guy Problem," *New York Times*, June 25, 2016.

How do we delegate power and decision-making to AI systems?

The integration of AI systems within social and economic domains requires the transformation of social problems into technical problems, such that they can be "solved" by AI. This is not a neutral translation: framing a problem so as to be tractable to an AI system changes and constrains the assumptions regarding the scope of the problem, and the imaginable solutions.[71]

Further, such social-to-technical translations provide no assurance that AI systems will produce fewer mistakes than the existing system they are intended to replace. As Ryan Calo points out, while it is usually believed that AI systems (such as self-driving cars) will commit fewer errors than humans – and, indeed, they may commit *fewer errors of the kind that humans do* – for any system of even modest complexity, the AI system will inevitably produce new kinds of errors, potentially of a type that humans do not make, and thus, in many cases, that our current ethical frameworks may be ill equipped to address.[72]

Across many domains, ethical frameworks often require the production of a record, like a medical chart, a lawyer's case file, or a researcher's Institutional Review Board submission. They also provide mechanisms for redress by patients, clients, or subjects when these people feel they have been treated unethically.

Contemporary AI systems often fall short of providing such records or mechanisms for redress, either because it's not technically possible to do so, or because the system was not designed with records and redress in mind (for instance, when such a system is the proprietary property of a given company).[73]

This means that AI and other predictive systems that are used to make claims about particular groups or individuals are often unreviewable and uncontestable by those affected. The inability to review or meaningfully contest determinations and decisions made by an AI system can accentuate various forms of power asymmetries, which is a key ethical concern.

When individuals impacted by a given automated decision are not able to review, contest, or appeal such decisions (assuming they are even aware that an automated

---

[71] For a discussion of the "two basic problems for any overarching classification scheme" (which always come into play at some stage in the development of an AI), see Geoffrey Bowker and Susan Leigh Star, *Sorting Things Out: Classification and its Consequences.* Cambridge: MIT Press, 1999, 69-70.

[72] Ryan Calo, "The Case for a Federal Robotics Commission," Brookings Institute, 2014, 6-8. Long-standing research in human factors engineering has demonstrated that automation should be understood to create new kinds of human error as much as it eliminates some kinds of human error. For example: Laine Bainbridge, "Ironies of Automation," *Automatica* 19 (1983): 775-779; Raja Parasuraman and Victor Riley, "Humans and Automation: Use, Misuse, Disuse, Abuse, " *Human Factors* 39 no.2 (June 1997): 230-253.

[73] See the work of Danielle Citron on the ways that algorithmic and automated systems essentially undermine the mechanisms of due process that have been developed in administrative law over the course of the 20th century. Danielle Keats Citron, "Technological Due Process," *Washington University Law Review*, vol. 85 (2007): 1249-1313.

system was behind a given decision or transaction) they are in a position of relative powerlessness.[74]

This poses the risk that AI systems will not only aggregate power by reducing the ability of the weakest to contest their own treatment, but also that those who design the systems will be able to define what counts as ethical behavior. Such power can take very subtle forms. For instance, we see that various kinds of automated systems are often used to influence or 'nudge' individuals in particular directions, largely dictated by those who design and deploy (and profit from) such systems.[75]

The ability to build AI systems from scratch, and thus potentially correct for such imbalances or achieve alternative goals, is itself subject to disparities. As mentioned above, building and maintaining AI systems requires vast computational resources, and vast quantities of data. Those that have the data and computing resources will have a strategic advantage over those who do not.

How do we integrate AI-specific ethical concerns within existing professional practices?

As AI systems become more integrated into professional environments, such as medicine, law and finance, new ethical dilemmas will arise across diverse professions.

For example, the use of AI systems in health contexts presents challenges to the core values upheld within the ethical codes of medical professionals, values such as confidentiality, continuity of care, avoiding conflicts of interest, and informed consent.[76] Challenges to these core values may arise in new and unanticipated ways, as diverse stakeholders promote various AI products and services within the context of the healthcare environment. How would a doctor uphold her oath to avoid conflicts of interest while using an AI diagnostic device trained on drug trial data owned by a pharmaceutical company with a vested interest in the prescription of specific drugs? While this example is hypothetical, it points to the type of thorny questions that must be addressed in the process of revising and updating professional codes of ethics.

Similarly, the professional associations that, broadly, govern the development and maintenance of AI technologies – the Association for the Advancement of Artificial Intelligence (AAAI), the Association for Computing Machinery (ACM) and the Institute of

---

[74] See Virginia Eubanks' discussion of algorithms becoming "decision makers" in cases of policing practices and public assistance. Virginia Eubanks, "The Policy Machine," *Slate*, April 30, 2015. In the former case, individuals were subjected to heightened police scrutiny for being placed on the Chicago Police Department's "heat list". In the latter case, an individual with a disability was unable to appeal an algorithmic decision to deny her public assistance despite her attempts to place an in-person interview to advocate on her own behalf. See also Citron, "Technological Due Process;" Danielle Citron and Frank Pasquale, "The Scored Society: Due Process for Automated Predictions," *Washington Law Review* 89 (2014): 1-33.

[75] Triston Harris, "How Technology Hijacks People's Minds, From a Magician and Google's Design Ethicist," *Medium,* May 18, 2016.

[76] American Medical Association, Code of Medical Ethics, "Opinion 10.01 - Fundamental Elements of the Patient-Physician Relationship."

Electrical and Electronics Engineers (IEEE) – also need to consider the creation of a codes of ethics (in the case of AAAI) or careful revision to their codes of ethics (in the case of ACM and IEEE). The existing codes from ACM and IEEE are over 20 years old,[77] and needless to say do not address the core issues of human agency, privacy, security, dignity, and harm prevention in the context of AI and automated decision-making systems. This is increasingly important as these technologies are integrated further into key social domains.

While more higher education institutions are beginning emphasizing ethics in their technical and scientific departments,[78] this emphasis is still early, and topics like civil rights, civil liberties, and ethical practice have yet to be standard requirements for graduation. Further, it is important to note that while a breach of the ethical code governing the practice of medicine carries penalties, including the potential of losing the ability to practice in the field, this is not the case in computer science or many other relevant domains. It is unclear whether most computer scientists would be able to recount the core tenants of the ACM or IEEE codes of ethics if asked, and also unclear whether, even if they were familiar, the incentives of employers or other pressures would not outweigh adherence to such non-binding codes.

Thus from a practical perspective, beyond simply redrafting and updating ethical frameworks, it's necessary to look at broader incentive structures, and ensure that they are formulated such that adherence to ethical codes is not simply an afterthought, but a core concern within relevant professional domains and an integral aspect of both learning and practice in the AI field.

Recommendations elaborated

Here we expand on the rationale behind the recommendations presented in brief above.

1. **Diversify and broaden access to the resources necessary for AI development and deployment – such as datasets, computing resources, education, and training, including opportunities to participate in such development. Focus especially for populations that currently lack such access.**

   As many noted during the *AI Now* Experts' Workshop, the means to create and train AI systems are expensive and limited to a handful of large actors. Or, put simply, it's not possible to DIY AI without significant resources. Training AI models requires a huge amount of data – the more the better. It also requires significant computing power, which is expensive. This limits fundamental research to those who can afford such access, and thus limits the possibility of democratically creating AI systems that

[77] Jake Metcalf and Kate Crawford, "Where are human subjects in Big Data research? The emerging ethics divide," *Big Data & Society*, June 3(1), 2016.

[78] https://ctsp.berkeley.edu/an-ischool-pledge-of-ethics.

serve the goals of diverse populations. Investing in foundational infrastructure and access to appropriate training data could help even the playing field. Similarly, opening up the development and design process within existing industry and institutional settings to diverse disciplines and external comment could help create AI systems that better serve and reflect diverse contexts and needs.

2. **Update definitions and frameworks that specify fair labor practices to keep pace with the structural changes occurring as AI management systems are deployed into the workplace. Also investigate alternative modes of income and resource distribution, education, and retraining that are capable of responding to a future in which repetitive work is increasingly automated and the dynamics of labor and employment change.**

   During his address at the *AI Now* Experts' Workshop, Jason Furman, President's Obama's Chief Economist, noted that 83% of jobs making less than $20 per hour in the US will face serious pressure from automation. For middle-income work that pays between $20 and $40 per hour, that number is still as high as 31%.[79] This is a seismic shift in labor markets, and one that could lead to a permanent unemployed class. To ensure that AI systems' efficiencies within labor markets doesn't lead to civil unrest, or the dissolution of key social institutions like education (in a scenario where education is no longer seen as a pathway to better employment), alternative means of resource distribution and other modes of handling the introduction of automation should be researched thoroughly, and policy should make way for well-structured tests of various implementations before the seismic shift, and its potentially disastrous consequences, has taken hold.

   AI systems also have manifold impacts within labor markets, beyond "replacing workers." They shift power relationships, employee expectations, and the role of work itself. These shifts are already having profound impacts on workers, and as such it's important that an understanding of these impacts take into account the way in which fair and unfair practices are constituted as AI systems are introduced. For example, if companies that develop AI systems that effectively act as management are seen to be technology service companies, as opposed to employers, employees may be left without existing legal protections.

3. **Support research to develop the means of measuring and assessing AI systems' accuracy and fairness during the design and deployment stage. Similarly, support research to develop means of measuring and addressing AI errors and harms once in-use, including accountability mechanisms involving notification, rectification, and redress for those subject to AI systems' automated decision making. Such means should prioritize notifying people when they are subject to automated**

[79] Jason Furman, "Is This Time Different? The Opportunities and Challenges of Artificial Intelligence," transcribed remarks from the AI Now expert workshop, July 7, 2016, New York University

decision-making, and developing avenues to contest incorrect or adverse judgments. In addition, investigate what the ability to "opt out" of such decision making would look like across various social and economic contexts.

AI and predictive systems increasingly determine whether people are granted or denied opportunities. In many cases, people are unaware that a machine, and not a human process, is making life-defining decisions. Even when they are aware, there is no standard processes to contest an incorrect characterization, or to push back against an adverse decision. We need to invest in research and technical prototyping that will ensure that basic rights and liberties are respected in contexts where AI systems are increasingly used to make important decisions.

4. **Clarify that neither the Computer Fraud and Abuse Act nor the Digital Millennium Copyright Act intend to restrict research into AI accountability.**

   In order to conduct the research necessary for examining, measuring, and evaluating the impact of AI systems on public and private institutional decision-making, especially in terms of key social concerns such as fairness and bias, researchers must be clearly allowed to test systems across numerous domains and via numerous methodologies. However, certain U.S. laws, such as the Computer Fraud and Abuse Act (CFAA) and the Digital Millennium Copyright Act (DMCA), threaten to limit or prohibit this research by outlawing "unauthorized" interactions with computing systems, even publicly accessible ones on the internet. These laws should be clarified or amended to explicitly allow for interactions that promote such critical research.

5. **Support fundamental research into robust methodologies for assessment and evaluation of the social and economic impacts of AI systems deployed in real-world contexts. Work with government agencies to integrate these new techniques into their investigative, regulatory, and enforcement capacities.**

   We currently lack a rigorous practice for assessing and understanding the social and economic impacts of AI systems. This means that AI systems are being integrated into existing social and economic domains, and deployed within new products and contexts, without an ability to measure or calibrate their impact. The situation can be likened to conducting an experiment without bothering to note the results. To ensure the benefits of AI systems, it is imperative that concerted research be done that develops rigorous methodologies for understanding AI systems' impact, and which can help shape practices across sectors and within government when such methodologies are used. Such research and its results can be likened to an early warning system.

6. **Work with representatives and members of communities impacted by the application of automated decision making and AI systems to codesign AI with accountability, in collaboration with those developing and deploying such systems.**

In many cases, those living with the impact of AI systems will be the most expert on the context and outcome of AI systems. Especially given the current lack of diversity within the AI field, it is imperative that those impacted by AI systems' deployment be substantively engaged in providing feedback and design direction, and that these suggestions form a feedback loop that can directly influence AI systems' development and broader policy frameworks.

7. **Increase efforts to improve diversity among AI developers and researchers, and broaden and incorporate the full range of perspectives, contexts, and disciplinary backgrounds into the development of AI systems. The field of AI should also support and promote interdisciplinary AI research initiatives that look at AI systems' impact from multiple perspectives, combining the computational, social scientific, and humanistic.**

   As a field, computer science suffers from a lack of diversity. Women, in particular, are heavily underrepresented. The situation is even more severe in AI. For example, while a handful of AI academic labs are being run by women, only 13.7 percent of attendees were women at the most recent Neural Information Processing Systems conference, one of the field's most important annual gatherings.[80] A community that lacks diversity is less likely to consider the needs and concerns of those not among its membership. When these needs and concerns are central to the social and economic institutions in which AI is being deployed, it is essential that they be understood, and that AI development reflects these critical perspectives. Focusing on diversity among those creating AI is key. Beyond gender and representation of protected classes of people, it is also important that diversity include various disciplines beyond computer science (CS), creating development practices that rely on expertise from those trained in the study of applicable social and economic domains.

   Much of the expertise that will be needed to conduct thorough assessments of AI's impact within social and economic domains lies outside of CS and the AI subfield within CS. This since the many contexts within which AI is being integrated and used – be it medicine, labor markets, or online advertising – are themselves rich areas for study. To truly develop rigorous processes for assessment of AI's impacts, we will need interdisciplinary collaboration, and the creation of new research directions and fields.

8. **Work with professional organizations such as The Association for the Advancement of Artificial Intelligence (AAAI), The Association of Computing Machinery (ACM), and The Institute of Electrical and Electronics Engineers (IEEE) to update (or create) professional codes of ethics that better reflect the complexity of deploying AI and**

[80] Jack Clark, "Artificial Intelligence Has a Sea of Dudes Problem," Bloomberg, June 23, 2016.

automated systems within social and economic domains. Mirror these changes in education, making courses in civil rights, civil liberties, and ethics required training for anyone wishing to major in Computer Science. Similarly, update professional codes of ethics governing the professions into which AI systems are being introduced, such as those applied to doctors and hospital workers.

In professions such as medicine and law, professional conduct is governed by a code of ethics that dictates acceptable and unacceptable practices. Professional organizations like ACM and IEEE do have codes of ethics, however these codes are outdated, and do not adequately address the specific and often subtle challenges inherent in implementing AI systems within complex social and economic contexts. While doctors do follow professional codes governing their conduct with patients, the development of AI systems that, for example, assist doctors in patient diagnosis and treatment, pose ethical challenges which are not always addressed by existing professional codes of ethics. Professional codes and computer science training must be updated to reflect the responsibilities that AI system builders have to those that will suffer disproportionately adverse effects as a result of the implementation of these systems. In cases where AI is used to augment human decision-making, professional codes of ethics should include safeguards to identify when AI-systems are subject to conflicts of interest.