Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change

Kelly Joyce¹, Laurel Smith-Doerr², Sharla Alegria³, Susan Bell¹, Taylor Cruz⁴, Steve G. Hoffman³, Safiya Umoja Noble⁵, and Benjamin Shestakofsky⁶

Abstract
This article outlines a research agenda for a sociology of artificial intelligence (AI). The authors review two areas in which sociological theories and methods have made significant contributions to the study of inequalities and AI: (1) the politics of algorithms, data, and code and (2) the social shaping of AI in practice. The authors contrast sociological approaches that emphasize intersectional inequalities and social structure with other disciplines’ approaches to the social dimensions of AI, which often have a thin understanding of the social and emphasize individual-level interventions. This scoping article invites sociologists to use the discipline’s theoretical and methodological tools to analyze when and how inequalities are made more durable by AI systems. Sociologists have an ability to identify how inequalities are embedded in all aspects of society and to point toward avenues for structural social change. Therefore, sociologists should play a leading role in the imagining and shaping of AI futures.

Keywords
artificial intelligence, ethical technology, intersectionality, sociology of technology

High-profile debates about the societal impact of artificial intelligence (AI) are appearing ever more frequently. The first pedestrian was killed by a self-driving car in Arizona, raising concerns about safety and accountability (Wakabayashi 2018). During the coronavirus pandemic, headlines suggested that the risks associated with face-to-face interactions were spurring the replacement of human workers with AI and automation (Semuels 2020; Walsh 2020). Examples such as an algorithm that recommended that black patients receive less health care than white patients with the same conditions (Obermeyer et al. 2019) and a report showing that facial recognition software is less likely to recognize people of color and women (Buolamwini and Gebru 2018; Grother, Ngan, and Hanaoka 2019) showed that AI can intensify existing inequalities. Many have raised concerns about AI and the future of health care, inequalities, warfare, work, and more.

Despite these concerns, there is a push to rapidly accelerate AI development in the United States. On February 11, 2019, President Trump signed executive order 13859, announcing the American AI Initiative: the U.S. national strategy on AI (OSTP 2019). On June 30, 2020, the President’s Council of Advisors on Science and Technology (2020) released a report recommending that federal investment in AI research and development grow by a factor of 10 over the next 10 years. Private companies such as Amazon, Google, and Microsoft are also investing heavily in AI development, both within their firms and in partnerships with federal agencies such as the National Science Foundation through initiatives like its Program on Fairness in AI in Collaboration with Amazon (NSF 2020). Through these investments, coupled with the societal value attributed to technological solutions, AI is increasingly being used to organize data, predict outcomes, and manage social worlds. Deeply intertwined with society, such systems are what science and technology studies scholars call sociotechnical, a term that calls attention to how values, institutional practices, and inequalities are embedded in AI’s code, design, and use.

¹Drexel University, Philadelphia, PA, USA
²University of Massachusetts, Amherst, MA, USA
³University of Toronto, Toronto, ON, Canada
⁴California State University, Fullerton, CA, USA
⁵University of California, Los Angeles, Los Angeles, CA, USA
⁶University of Pennsylvania, Philadelphia, PA, USA

Corresponding Author:
Kelly Joyce, Drexel University, 3161 Chestnut Street, Philadelphia, PA 19104-2816, USA.
Email: kaj68@drexel.edu
This article is a call for sociologists to critically engage AI sociotechnical systems, as well as a call for AI practitioners and policy makers to more fully engage with sociological insights into AI, inequalities, and structural change. In this scoping, forward-looking article, we frame new fields of inquiry to increase sociology’s participation in AI research. We review two areas—(1) the politics of algorithms, data, and code and (2) the social shaping of AI in practice—in which sociological theories and methods have made significant contributions to the study of inequalities and AI. We contrast sociological approaches that emphasize intersectional inequalities and social structure with other disciplines’ approaches to the social dimensions of AI, which tend to have a thin understanding of the social and emphasize individual-level interventions. We also suggest future areas of research, using immigration, health, and divisions of labor as illustrative examples. Scholars are using sociological theories and methods to investigate the creation, use, and effects of AI sociotechnical systems in art (Sachs 2019), health care (Cruz 2020; Pugh 2018), policing (Brayne 2020; Brayne and Christin forthcoming), politics (Tripodi 2019), social movements (Donovan 2017), systemic racism (Benjamin 2019; Daniels 2009, 2018; McMillan Cottom 2020; Noble 2018), and work (Christin 2017, 2020; Hoffman 2017; Mateescu and Ticona 2020; Pugh 2020; Sachs 2019; Shestakovskiy 2017; Shestakovskiy and Kelkar 2020; Smith-Doerr et al. 2019; Ticona and Mateescu 2018). Sociologists have an ability to identify how inequalities are embedded in all aspects of society and to point toward avenues for structural social change. Therefore, sociologists should play a leading role in the imagining and shaping of AI futures.

**Situating AI**

There is a long-standing cultural fascination with imagining and creating agentic machines (Buchanan 2006; Haenlein and Kaplan 2019; Nilsson 2009). With the increasing availability of computers in the 1940s and 1950s, scientists began trying to materialize this imagination by exploring whether computer software systems could be programmed to exhibit agency and decision-making. Alan Turing ([1948] 2004), a British mathematician, famously asked whether it was possible for machines to demonstrate evidence of intelligent behavior. The term *artificial intelligence* was coined in the 1950s as a way to distinguish new research from the prevailing paradigm of cybernetics (McCarthy 1988). Under the banner of AI, scientists began to design software that could play checkers and solve math problems. They aimed to create software that was able to learn and act on its own.

The term *AI* has been used in many different ways, in part because what counts as AI differs across a wide range of contexts and in part because drawing definitional boundaries is a deeply interpretative, political process (Forsythe 2001; Hoffman 2015). Although AI is about the deployment of computing infrastructure and programming code to create systems expected to mimic, augment, or displace human agency (H. M. Collins 1990; Suchman 2007), AI designers draw upon an enormous diversity of computational techniques, including cybernetics, logic and rule-based systems, statistical and probabilistic AI, pattern recognition, adaptive behavioral models, and symbolic knowledge representation such as expert systems (Crevier 1993; McCorduck 2004; Nilsson 2009). AI applications are also diverse, encompassing task automation, algorithmic risk scoring, and robotics, with scientific and public observers alike expecting AI to increasingly stand in for human labor, relationships, judgment, and creativity (Renski et al. 2020; Mateescu and Elish 2019; Mittelstadt et al. 2016). In part, AI’s fluid definitional scope explains its appeal; its expansive yet unspecified meaning enables promoters to make future-oriented, empirically unsubstantiated promissory claims of its potential societal impact.

In recent years, programming communities have largely focused on developing machine learning (ML) as a form of AI. The term *ML* is now more commonly used among researchers than the term *AI*, although AI continues to be the public-facing term used by companies, institutes, and initiatives. ML emphasizes the training of computer systems to recognize, sort, and predict outcomes from analysis of existing data sets. AI practitioners, including computer scientists, engineers, and the new occupation of data scientists (Ribes 2019), input existing data in order to train ML systems so that they can make autonomous decisions. This decision making is then actively reintegrated back into the social world (Beer 2017; Burrell 2016; Pink et al. 2017), supporting or making decisions about, for example, which roads get repaired, who is entitled to welfare benefits, and who is likely to return to prison (Eubanks 2018; O’Neil 2016). ML systems may also draw on real-time data generated within the conduct of life itself, including the large volumes of data generated through everyday use of computer technologies and sensor devices such as smart phones and security cameras (Greenfield 2018; Neff and Nafus 2016; van Dijck 2014). Recognizing real-time data’s potential value, researchers have taken advantage of the abundant data sources on everyday life to further ML research and applications. When used within ML systems, both existing and real-time human data points are often perceived as neutral, representative, and truthful in their correspondence with the social world (boyd and Crawford 2016; Joyce et al. 2018; Kitchin 2014). AI practitioners may not be aware that data about X (e.g., ZIP codes, health records) may also be data about Y (e.g., class and race inequalities, socioeconomic status). This assumption has resulted in the acceleration and intensification of inequalities as ML systems are developed and deployed, further compounded by the general epistemic authority granted to what AI practitioners call data-driven logics across institutional, scientific, and administrative spheres. Given that *AI* is the outward-facing term used by companies, institutes, and government initiatives, we use *AI*, instead of *ML*, in this piece.
The Politics of Algorithms, Data, and Code

Since its origins in the nineteenth century, sociology has highlighted the misguided assumption that data about social life are neutral or objective. In contrast with AI practitioners who assume that data generated about human subjects represent a singular, objective external reality, sociologists have identified the constructed, polyvalent nature of human data, challenging the simplicity of data-as-authority while recognizing the multiple meanings of quantified accounts (EspeLand and Sauder 2016; EspeLand and Stevens 2008). Data are never neutral but are instead value-laden (Gitelman 2013; Merry 2016; Scott 1998). Classification itself is a deeply moral project often implicated in social stratification (Bowker and Star 1999; Fourcade and Healy 2013; Thompson 2016). Data from mortality statistics, for example, stand as more than aggregated counts of unconnected individuals. More than a century ago Durkheim ([1897] 2006) demonstrated that suicide rates correspond with social integration, and Du Bois (1906, p.89) viewed high infant mortality as “not a Negro affair, but an index of social condition.” Today, variation in crime rates across different ZIP codes and wage differences by gender continue to reflect social relations that are not fully explained by the raw data points themselves. Sociologists thus recognize that what counts as data is socialized, politicized, and multilayered, because data about humans are also often data about structural inequalities related to gender, race or class (MacKenzie and Wajcman 1999). A sociological understanding of data is important given that an uncritical use of human data in AI sociotechnical systems will tend to reproduce, and perhaps even exacerbate, preexisting social inequalities.

Beginning in the early 2000s, critical scholarship opened the black box of algorithms, code, and platforms, bringing these value-laden practices to public attention. Scholars, primarily in anthropology, communication, law, philosophy, and science and technology studies, identified algorithms as a site of humanist and social science inquiry. First, they showed that these systems are not neutral but instead have values embedded in their design, use, and output (e.g., Beer 2009, 2013; Gillespie 2011; Introna and Wood 2004). Such work challenged the objectivity ascribed to algorithms by showing, for example, that choices about trending topics and suggestions on social media reflect value-laden decisions (Gillespie 2011) and by examining how algorithmic decision making, social categories, and cultural contexts are intertwined (Hallinan and Striphas 2014; Introna 2015). Second, early work questioned the meaning of “algorithm” itself, asking what constitutes an algorithm and arguing that we need to approach algorithms as systems that are composed of decentralized, changing configurations of people and code (Barocas, Hood, and Ziewitz 2013; Seaver 2013). What to study was not self-evident, and intellectual work helped operationalize studies of these social worlds and artifacts. Finally, scholarship began to examine the politics of algorithms and platforms, connecting these technologies to issues of surveillance, power, and social control (Introna 2015; Introna and Nissenbaum 2000; Introna and Wood 2004; Pasquale 2015).

Since this early groundbreaking work, there has been a flurry of scholarship on algorithms, code, and the resulting boom in ML and AI. One of the first to focus on the structural, intersectional dimensions of inequality and algorithms, Noble (2018) drew on black feminist thought (P. H. Collins 1990; Crenshaw 1991; hooks 1992) to show how Google’s search engine recreates the sexism and racism found in everyday life. Although Google blamed users for contributing to racist, sexist search results, Noble demonstrated how the standpoint of Google workers affects their decisions about what to label racist or sexist, as do the broader social contexts of pervasive racism and sexism. Additionally, Google’s profit motive in a capitalist economy leads it to prioritize particular information rankings and retrieval systems, producing what Noble called “technological redlining” and “algorithmic oppression,” building on the concept of “digital redlining” (Gilliard and Culik 2016). Technological redlining reminds us that the racist redlining practices used by banks to treat clients of color differently than white clients are still present, although in new forms. In contrast to the perceived neutrality accorded to Google’s search results, oppression and inequality are implicated and amplified in the decisions that produce them. Calling attention to the global, structural dimension of algorithms and code, Noble wrote, “The problems of big data go deeper than representation. They include decision-making protocols that favor corporate elites and the powerful, and they are implicated in global economic and social inequality” (p. 29). Silicon Valley is part of an inequalities project that produces class, gender, and racial inequalities in contemporary life (Noble and Roberts 2019; Noble and Tynes 2016).

Sociologists have built on this structural, intersectional approach (Collins 2019; Hancock 2016) to understand the many ways inequalities are deeply embedded in AI socio-technical systems, as well as their connections to profit and capitalism. Benjamin (2019) showed how history and biography collide in AI systems by calling them the “new Jim Code.” For Benjamin, the new Jim Code refers to “the employment of new technologies that reflect and reproduce existing inequities but that are promoted and perceived as more objective or progressive than the discriminatory systems of a previous era” (p. 23). Daniels (2018) demonstrated how white supremacists flourish in AI systems: decision-making protocols make racist content readily available, and companies refuse to ban such content from their platforms. Williams, Bryant, and Carvell (2019) documented the emotional cost to people of color who have to negotiate digital spaces that are simultaneously white racialized spaces. Calling attention to the structural, global dimensions of AI, McMillan Cottom...
(2020:443) argued that platform capitalism is racial capitalism, in which software constitutes sociopolitical regimes that produce “new forms of currency (i.e., data), new forms of exchange (e.g., cryptocurrencies), and structure new organizational arrangements between owners, workers and consumers” that cross national boundaries. Although companies that produce AI systems hide behind the claim that algorithms or platform users create racist, sexist outcomes, sociological scholarship illustrates how human decision making occurs at every step of the coding process. The desire of companies to keep processes private and secret to protect intellectual property and deflect scrutiny, coupled with a lack of government oversight, allows such decisions to occur out of public view (Lazer et al. 2020; McMillan Cottom 2020).

In addition to examining the human decisions that structure how AI systems function, sociologists have also turned a critical eye toward the data being used to train such systems. Drawing on fieldwork at three ML research sites, Joyce et al. (2018) argued that computer scientists and engineers are not trained to understand the multiple dimensions of human data and how these relate to systems of inequalities. Although computer scientists and engineers often consider privacy when designing ML systems, understanding the multivalent dimensions of human data is not typically part of their training. Given this, they may assume that data in electronic health records represent objective knowledge about treatment and outcomes, instead of viewing it through a sociological lens that recognizes how electronic health record data are partial and situated. Clinicians may code strategically to ensure insurance reimbursement, for example, instead of using codes that reflect the conditions that they believe patients have. Patient outcomes are also not neutral or objective; these too are related to patients’ socioeconomic status and often tell us more about class and other kinds of inequalities than the effectiveness of particular treatments. Cruz (2020) critically investigated the idea that more data will lead to reduced inequality by studying what happens when social and behavioral domains are added to patient records at a large, public hospital. Sociological and public health research has shown that gender, sexuality, and health are interconnected, and are part of understanding the social determinants (or causes) of health and illness. Despite the hope that better data on social factors would lead to more justice in health care, Cruz, by following data in action, demonstrated that providers’ and staff members’ lack of understanding of social determinants and their connection to the collected data, in addition to providers’ need to gather and input a variety of data in compressed time slots, meant that gender and sexuality data were not taken up in any meaningful way. Calling attention to the need for structural change beyond data collection alone, Cruz wrote, “To truly advance justice, we need more than ‘just data’: we need to confront the fundamental conditions of social inequality” (p. 1).

**The Social Shaping of AI in Practice**

Sociological scholarship, joining other critical social science research, shows that there is not one way of doing AI but rather numerous ways that emerge from the situated intersection of economics, organizational contexts, politics, regulations, and values implicated in the production of AI systems. By examining the design and implementation of AI sociotechnical systems, sociological work brings human labor and social contexts into view. Building on sociology’s recognition of the importance of organizational contexts in shaping outcomes, this scholarship makes three contributions to a sociology of AI. First, funding sources matter. Hoffman’s (2017) ethnographic study of two academic AI laboratories examined the role of commercial versus federal funding in shaping research questions, and thus the knowledge that labs produced. Highlighting how actors make sense of and structure their work, Hoffman showed that the lab that pursued commercial funding was able to nimbly move in and out of new problem areas, eschew theory for practical problem solving, and produce commercially viable AI systems. In contrast, the lab that secured government funding produced deep but incremental research that spoke primarily to a narrow community of experts. Both labs enact particular knowledge regimes that serve particular economic and political interests, demonstrating how context matters. Within academic capitalism, there are diverse approaches that influence the types of AI systems created.

Second, sociological analyses of the introduction of AI systems to work sites demonstrate how their uneven uptake contributes to the social production of inequalities. Brayne and Christin (forthcoming) showed how organizational contexts relate to workers’ ability to resist the use of predictive AI. Examining an urban police department and a midsized criminal court, Brayne and Christin identified two strategies of resistance used by workers: ignoring AI and obscuring data by blocking its collection or adding more data to dilute findings. Although workers used both strategies, the hierarchical, managerial structure of the police department contributed to police’s using predictive analytics on a more regular basis than in the court. Through this analysis of AI in practice, Brayne and Christin demonstrated “the enduring role of discretion, power, and occupational cultures in shaping the social and political impact of technological change” (p. 15). Challenging the belief that more data will lead to more equity and justice, Brayne (2017, 2020) showed how the introduction of predictive analytics to the Los Angeles Police Department deepens inequalities. AI use caused more surveillance of people already considered suspicious and exacerbated spatial inequalities by subjecting low-income neighborhoods of color to risk assessment and intervention. The integration of nonpolice, external data into surveillance systems may further deter people from seeking needed resources (e.g., social services, health care) for fear of leaving digital traces that can then increase their visibility in surveillance systems.
Third, sociological scholarship empirically investigates the claim that robots are taking our jobs by using an intersectional lens to examine how and where jobs are automated (Smith-Doerr et al. 2019) and how new forms of labor are created (Dahlin 2019; Shestakofsky 2017). Shestakofsky’s (2017) research at a small software firm revealed that new AI systems do not necessarily replace workers but can rather result in new complementarities between people and technologies. Shestakofsky found that human workers take on two new roles: they provide “computational labor that supports or stands in for software algorithms and emotional labor aimed at helping users adapt to software systems” (p. 409). Future research should not assume that automation will replace workers or that the outcomes of technology use are monolithic; instead, it should identify the interconnections between people and technology and how these are related to organizational contexts, the spatial distribution of jobs, and systems of inequalities (Smith-Doerr 2020; Vallas and Schor 2020).

Why a Sociological Approach to AI Is Needed

The sociological approach outlined above stands in sharp contrast to other efforts to understand the social dimensions of AI sociotechnical systems. Contemporary AI scientists do recognize the existence of the social and use multiple strategies to account for it within AI. The typical practitioner approach includes expanded data collection on dimensions of social worlds, such as when information on patient demographics (race, ethnicity, sexual orientation, and gender) and other social determinants of health (housing status, food stability, social support) are included in electronic health records (Chen et al. 2020; Douglas et al. 2015; IOM 2014; Onukwugha, Duru, and Peprah 2017). AI teams also commonly include a domain expert (e.g., physician, urban planner, educator) who is expected to speak for the social world that will be affected by code. The domain expert provides the team with information about how that world operates in addition to anticipating user responses to AI implementation (Ribes 2019). AI teams also conduct users studies to better understand how future users might navigate AI interfaces and conduct literature reviews on the problems they seek to address to better understand the particularities of a relevant social world. In addition to trying to incorporate the social within AI systems, practitioners rhetorically highlight societal impact as a way to promote products, with several “AI for social good” initiatives targeting society’s biggest challenges in education, urban development, and climate change (Google 2020; Microsoft 2020; United Nations 2020). Although early AI researchers tended to “delete the social” by privileging technical factors over social ones in design practices (Forsythe 2001; Star 1991), practitioners today increasingly aim to account for the social dimensions of AI within their work.

Yet these AI practitioner-led efforts typically fail to recognize the full complexity of social life, especially when situated against the critical insights offered by sociologists. AI practitioners tend to incorporate the social in ways that position human data as singular, straightforward, and stable, missing an opportunity to fully reckon with how power dynamics, systemic discrimination, and social inequalities contribute to the meaning of social categories (Cruz 2020; Noble 2018). Data collection that ties individual records to elements of social context (e.g., race or housing status) consistently neglects to situate inequalities within the organizations and institutions that reproduce them (Phelan, Link, and Tehranifar 2010; Roberts 2012; Williams and Sternthal 2010). Data science approaches regularly fail to capture the embodied, interactional nature of difference and its shifting meaning across time and place (Lee and Bean 2010; Dubbin, Chang, and Shim 2013; Spencer and Grace 2016; Bakhtiar, Olafsdottir, and Beckfield 2018). Moreover, a small number of individual experts cannot adequately speak for complex social worlds composed of multiple professions, stakeholders, and social hierarchies (Bowker and Star 1999; Timmermans and Berg 2003; Whooley 2019). Despite well-intentioned efforts to incorporate knowledge about social worlds into sociotechnical systems, AI scientists continue to demonstrate a limited understanding of the social, prioritizing that which may be instrumental for the execution of AI engineering tasks but erasing the complexity and embeddedness of social inequalities.

Sociology’s deeply structural approach also stands in contrast to approaches that highlight individual choice. One of the most pervasive tropes of political liberalism is that social change is driven by individual choice. As individuals, the logic goes, we can create more equitable futures by making and choosing better products, practices, and political representatives. The tech world tends to sustain a similarly individualistic perspective when its engineers and ethicists emphasize eliminating individual-level human bias and improving sensitivity training as a way to address inequality in AI systems.

Consider the predominant language of bias, fairness, transparency, and accountability used in tech analysts’ attempts to build more equitable AI systems. Scholars and programmers located in corporations (e.g., Microsoft, Google), along with those in nonprofits funded in part by technology companies (e.g., Data & Society, the AI Now Institute), have identified important harms caused by AI. One harm, identified as bias, occurs when ML system results disproportionately affect particular groups (Barocas et al. 2017; Crawford, Miltn, and Gray 2014). The evidence of bias in AI systems is often perceived as a vestige of the human bias coded into them by individual programmers who need better training. Google, for example, launched a fairness module in its ML Crash Course in 2018 and the ML Fairness Effort in 2019 (Dean 2019), while the Association for Computing Machinery (ACM) created the ACM Conference on Fairness, Accountability, and Transparency in 2018, all with the goal of creating less biased systems in part by training individual technologists to be
Taking a sociological viewpoint highlights that what counts attend to meaning-making, understanding that objects and are trained to identify the multiple meanings of sociotechnical work flows (Amisha et al. 2019; Powell 2020). Sociologists to improve treatment outcomes, patient safety, and clinical systems, and how electronic health records can be analyzed lic sources of funding, are studying the feasibility of sup-2017; UNHCR 2020). AI researchers, with private and pub- nate and manage refugee resettlement (United Nations and Dell, for example, are developing AI systems to coordi- other areas of social life. Companies including Microsoft and use, but more engagement is needed. The private and sociologists' contributions to the study of inequalities in algo- and Calo 2016) have understandably not been picked up. Critique from within for-profit companies is precarious and may not be sustained. Sociologists, along with critical social scientists (Eubanks 2018; Noble 2018), can advance the conversation on bias, fairness, transparency, and account- ability by transforming it into one of inequality, hierarchy, and structural social change.

Conclusion
Sociologists are contributing to the study of AI development and use, but more engagement is needed. The private and public investment in AI sociotechnical systems means that many more systems will be developed and integrated into organizations and societies in the coming years. Research investment aims to bring AI technologies to criminal justice, education, health care, social services, urban planning, and other areas of social life. Companies including Microsoft and Dell, for example, are developing AI systems to coordinate and manage refugee resettlement (United Nations 2017; UNHCR 2020). AI researchers, with private and public sources of funding, are studying the feasibility of supporting medical clinicians' decision making with algorithmic systems, and how electronic health records can be analyzed to improve treatment outcomes, patient safety, and clinical work flows (Amisha et al. 2019; Powell 2020). Sociologists are trained to identify the multiple meanings of sociotechnical systems, artifacts, and even data. As a discipline we attend to meaning-making, understanding that objects and interactions are always socially situated and multiple. Taking a sociological viewpoint highlights that what counts as AI differs across a wide range of contexts and that drawing definitional boundaries is itself a deeply interpretative, political, and fundamentally social process (Forsythe 2001; Hoffman 2015).

In proposing a sociology of AI, we have highlighted sociologists’ contributions to the study of inequalities in algorithms, code, and data, and how contexts shape AI design and use. Scholars are using sociological theories and methods to investigate AI sociotechnical systems in ways that illuminate intersectional systems of inequalities and the connection between individuals, organizations, and policies and regulations in global systems. When applied to computing, an intersectional approach demonstrates how AI systems contribute to the social production of inequalities. An intersectional sociological framework adds to the growing discussion of the unequal impact of AI, accounting for the global histories of enslavement, patriarchy, white supremacy, and capitalism, and the ways in which historical and contemporary iterations of these systems are replicated in, exacerbated by, and resisted in digital systems.

A sociology of AI must attend to where, when, how, and for whom AI is developed. By situating AI, sociologists can illuminate how AI processes contribute to or alleviate global and local inequalities, privilege, and vulnerability for individuals and nation-states; the political and social implications of AI for mobility and migration; and its intersectional effects on workers on a global scale. Sociologists can contribute to making AI’s division of labor visible by studying workforce diversity, organizational and interorganizational labor practices, and changes in the relationship between workers and employers. Although AI is conceptualized, designed, and used by people, the production, deployment, and use of AI systems, as well as the discourses about them, can hide important aspects of this work. For example, much of the AI we use is produced and distributed by for-profit companies in workplaces with raced, classed, and gendered divisions of labor. Furthermore, the deployment of AI systems does not usually live up to its promise to eliminate human labor. Instead, AI tends to reconfigure work flows and workplace divisions of labor, often in ways that systematically hide the expropriation of human labor (Ekbia and Nardi 2017; Gray and Suri 2019; Shestakofsky 2017).

AI’s emergence presents us with the opportunity to ask how sociotechnical change can help bring about a better and more just world. Drawing on our deep engagement with social structure and inequalities, sociologists can develop theoretical and empirical insights to inform our imagination of sociotechnical futures—and how they might take root—at the levels of individuals, institutions, and societies. How people deploy and interpret AI systems varies according to the institutional and organizational contexts in which they are implemented (Brayne 2017; Christin 2017; Shestakofsky 2017). Algorithmic models can be designed to optimize for efficiency or fairness—to exclude individuals from opportunity, or to identify and lend more accountable, fair, and unbiased. The language of bias and transparency comes in part from computer science, in which terms like underfitting (the system includes too few data to reflect the social world), and overfitting (the system includes too many random fluctuations to reflect the social world) are used to describe bias in data processing, and from psychology, which has long focused on individual-level perception and attitudes.

Although it is laudable that tech companies and think tanks have tried to address AI’s disproportionate impacts, there are severe limitations to an approach that primarily locates the problem within individuals. The language of human bias focuses on training individuals to create less biased, more fair and transparent systems. Sociological research demonstrates, though, that bias is not free-floating within individuals but is embedded in obdurate social institutions. Transparency and fairness, in other words, are not easily achieved through the better training of AI practitioners. Moreover, the high-profile firing of Timnit Gebru, an AI ethicist at Google, after she coauthored a paper that showed uneven harm from Google’s large-scale AI systems, shows what happens when concerns about inequalities challenge profit motives (Hanna and Whittaker 2020). Calls from within industry to address structural analysis (Crawford and Calo 2016) have understandably not been picked up. Critique from within for-profit companies is precarious and may not be sustained. Sociologists, along with critical social scientists (Eubanks 2018; Noble 2018), can advance the conversation on bias, fairness, transparency, and accountability by transforming it into one of inequality, hierarchy, and structural social change.
support to those who would benefit from additional resources (Eubanks 2018; O’Neil 2016). Digital platforms can be owned and governed by profit-seeking corporations, by cities, or cooperatively by the people who use them (Scholz 2016). Regulations will shape and constrain applications of AI in various social domains, and public policy will determine how individuals and societies cope with the consequences of its design and implementation (Jacobs and Karen 2019; Thelen 2018; Viscelli 2018).

In this indeterminacy lies opportunity for a sociologically informed research agenda that sparks hope and imagination. As Tavory and Eliasoph (2013) noted, “Ghosts of many potential futures haunt any interaction: any negotiation of the future contains and manages uncertainty. People’s ways of managing and performing it are where the action is” (p. 910). How can algorithmic design support social movements that advocate for equity, justice, and structural change? In what contexts can emancipatory design processes and outcomes emerge? How can social movements organize to redefine the direction of sociotechnical change? The rise of AI invites us to look toward an uncertain future and to ask how we can come together to minimize the harms generated by new technologies while sharing the benefits more broadly. The Social Science Research Council, led by sociologist Alondra Nelson, created a funding stream called Just Tech to help foster more equitable, inclusive technological futures in 2020. Federal agencies, private companies, and public-private initiatives are funding new AI development, and sociologists should create and participate in these research teams. Our focus on structural change and inequalities will help keep these issues front and center. The creation of AI sociotechnical systems is not simply a question of technological design but fundamental questions about power and social order.

Acknowledgment

The authors thank Aaron Yates for excellent research assistance.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Science Foundation under grants 1744356 and 2015845. All findings and views are the authors and do not necessarily reflect those of the NSF.

ORCID iDs

Kelly Joyce https://orcid.org/0000-0002-2822-1548
Laurel Smith-Doerr https://orcid.org/0000-0002-5162-0491
Sharla Alegria https://orcid.org/0000-0003-4388-9797
Susan Bell https://orcid.org/0000-0002-2537-1890
Steve Hoffman https://orcid.org/0000-0002-2379-8081
Benjamin Shestakofsky https://orcid.org/0000-0003-0797-2729

References

Amisha, Paras Malik, Monika Pathania, and Vyas Kumar Rathaur. 2019. “Overview of Artificial Intelligence in Medicine.” Journal of Family Medicine and Primary Care 8(7):2328–31.
Bakhhtiari, Elyas, Sigrun Olafsdottir, and Jason Beckfield. 2018. “Institutions, Incorporation, and Inequality: The Case of Minority Health Inequalities in Europe.” Journal of Health and Social Behavior 59(2):248–67.
Barocas, Solon, Kate Crawford, Aaron Shapiro, and Hanna Wallach. 2017. “The Problem with Bias: Allocative versus Representational Harms in Machine Learning.” Presented at the 9th Annual Conference of the Special Interest Group for Computing, Information and Society, Philadelphia. Retrieved February 23, 2021. http://meetings.sigcis.org/uploads/6/3/6/8/6368912/program.pdf.
Barocas, Solon, Sophie Hood, and Malte Ziewitz. 2013. “Governing Algorithms: A Provocation Piece.” Retrieved February 23, 2021. https://ssrn.com/abstract=2245322.
Beer, David. 2009. “Power through the Algorithm? Participatory Web Cultures and the Technological Unconscious.” New Media & Society 11(6):985–1002.
Beer, David. 2013. “Algorithms: Shaping Tastes and Manipulating the Circulations of Popular Culture.” In Popular Culture and New Media: The Politics of Circulation. New York: Palgrave Macmillan.
Beer, David. 2017. “The Social Power of Algorithms.” Information, Communication & Society 20(1):1–13.
Benjamin, Ruha. 2019. Race after Technology: Abolitionist Tools for the New Jim Code. Malden, MA: Polity.
Bowker, Geoffrey C., and Susan Leigh Star. 1999. Sorting Things Out: Classification and Its Consequences. Cambridge, MA: MIT Press.
yboy, danah, and Kate Crawford. 2012. “Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon.” Information, Communication & Society 15(5):662–79.
Brayne, Sarah. 2017. “Big Data Surveillance: The Case of Policing.” American Sociological Review 82(5):977–1008.
Brayne, Sarah. 2020. Predict and Surveil: Data, Discretion, and the Future of Policing. Oxford, UK: Oxford University Press.
Brayne, Sarah, and Angèle Christin. Forthcoming. “Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts.” Social Problems.
Buchanan, Bruce G. 2006. “A (Very) Brief History of Artificial Intelligence.” AI Magazine 26(4):53–60.
Buolamwini, Joy, and Timnit Gebru. 2018. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” In Proceedings of Machine Learning Research. Retrieved February 23, 2021. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.
Burrell, Jenna. 2016. “How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms.” Big Data & Society 13(1).
Chen, Soy, Danielle Bergman, Kelly Miller, Allison Kavanaugh, John Frownfelter, and John Showalter. 2020. “Using Applied Machine Learning to Predict Healthcare Utilization Based on Socioeconomic Determinants of Care.” American Journal of Managed Care 26(1):26–31.
Introna, Lucas, and David Wood. 2004. “Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems.” *Surveillance & Society* 2(2/3):177–98.

IOM (Institute of Medicine). 2014. “Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2.” Washington, DC: National Academies Press.

Jacobs, Jerry A., and Rachel Karen. 2019. “Technology-Driven Task Replacement and the Future of Employment.” *Research in the Sociology of Work* 33:43–60.

Joyce, Kelly A., Kendall Darfler, Dalton George, Jason Ludwig, Jacobs, Jerry A., and Rachel Karen. 2019. *The Algorithm at Work? Explanation and Repair*. New York: P. Peter Lang.

Joyce et al. 2019. *The Intersectional Internet: Race, Sex, Class, and Culture Online*. New York: Peter Lang.

NSF (National Science Foundation). 2020. “NSF Program on Fairness in Artificial Intelligence in Collaboration with Amazon.” Division of Information and Intelligent Systems. Retrieved December 22, 2020. https://www.nsf.gov/ funding/pgm_summ.jsp?pims_id=505651.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. “Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations.” *Science* 366(6464):447–53.

O’Neill, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Broadway.

Onukwugha, Eberuchekwu, O. Kenrik Duru, and Emmanuel Peprah. 2017. “Big Data and Its Application in Health Disparities Research.” *Ethnicity & Disease* 20(27):69–72.

OSTP (Office of Science and Technology Policy). 2019. “Summary of AI Executive Order, February 11, 2019.” White House Fact Sheets. Retrieved December 22, 2020. https://www.white- house.gov/ai/resources/white-house-fact-sheets/.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Phelan, Jo C., Bruce G. Link, and Parisa TehraniFar. 2010. “Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications.” *Journal of Health and Social Behavior* 51(1 Suppl.):S28–40.

Pink, Sarah, Shanti Sumartojo, Deborah Lupton, and Heyes La Bond, Christine. 2017. “Mundane Data: The Routines, Contingencies, Accomplishments of Digital Living.” *Big Data & Society* 14(1).

Powell, Alvin. 2020. “AI Revolution in Medicine.” *The Harvard Gazette*, November 11.

President’s Council of Advisors on Science and Technology. 2020. “Recommendations for Strengthening American Leadership in Industries of the Future: A Report to the President of the United States.” Retrieved February 23, 2021. https://science.osti.gov/-/media/~/media/pcast/202006/PCAST_June_2020_Report.pdf.

Pugh, Allison J. 2018. “Automated Health Care Offers Freedom from Shame, But Is It What Patients Need?” *The New Yorker*, May 22.

Pugh, Allison J. 2020. “An Inequality of Being Seen.” *Noema Magazine*, June 12.

Renski, Henry, Laurel Smith-Doerr, Tiamba Wilkerson, Shannon C. Roberts, Shlomo Zilberstein, and Enobong H. Branch. 2020. “Racial Equity and the Future of Work.” *Technology & Architecture + Design* 4(1):17–22.

Ribes David. 2019. “STS, Meet Data Science, Once Again.” *Science, Technology, & Human Values* 44(3):514–39.

Roberts, Dorothy E. 2012. *Fatal Invention: How Science, Politics, and Big Business Re-create Race in the Twenty-First Century*. New York: New Press.

Sachs, S. E. 2019. “The Algorithm at Work? Explanation and Repair in the Enactment of Similarity in Art Data.” *Information, Communication & Society* 23(11):1689–1705.
Scholz, Trebor. 2016. *Uberworked and Underpaid: How Workers Are Disrupting the Digital Economy*. Malden, MA: Polity.

Scott, James C. 1998. "Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed." New Haven, CT: Yale University Press.

Seaver, Nick. 2013. “Knowing Algorithms.” Paper presented at Media in Transition 8: Public Media, Private Media, Cambridge, MA, May 3–5.

Semuels, Alana. 2020. “Millions of Americans Have Lost Jobs in the Pandemic—And Robots and AI Are Replacing Them Faster Than Ever.” *Time*, August 6.

Shestakofsky, Benjamin. 2017. “Working Algorithms: Software Automation and the Future of Work.” *Work and Occupations* 44(4):376–423.

Shestakofsky, Benjamin, and Kelkar. 2020. “Making Platforms Work: Relationship Labor and the Management of Publics.” *Theory & Society* 49:863–96.

Smith-Doerr, Laurel. 2020. “Hidden Injustice and Anti-science.” *Engaging Science Technology and Society* 6:94–101.

Smith-Doerr, Laurel, Shlomo Zilberstein, Tiamba Wilkerson, Shannon Roberts, Henry Renski, Venus Green, and Enobong H. Branch. 2019. “HTF (the Future of Work at the Human-Technology Frontier): Understanding Emerging Technologies, Racial Equity, and the Future of Work.” Retrieved December 20, 2020. http://rbr.cs.umass.edu/htf/NSF-Workshop-Report.pdf.

Spencer, Karen Lutfey, and Matthew Grace. 2016. “Social Foundations of Health Care Inequality and Treatment Bias.” *Annual Review of Sociology* 42:101–20.

Star, S. L. 1991. “The Sociology of the Invisible: The Primacy of Work in the Writings of Anselm Strauss.” *Pp. 265–83 in Social Organization and Social Process: Essays in Honor of Anselm Strauss*, edited by D. Mames. New York: Aldine de Gruyter.

Suchman, Lucy. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge, UK: Cambridge University Press.

Tavory, Ido, and Nina Eliasoph. 2013. “Coordinating Futures: Toward a Theory of Anticipation.” *American Journal of Sociology* 118(4):908–42.

Thelen, Kathleen. 2018. “Regulating Uber: The Politics of the Platform Economy in Europe and the United States.” *Perspectives on Politics* 16(4):938–53.

Thompson, Debra. 2016. *The Schematic State: Race, Transnationalism, and the Politics of the Census*. Cambridge, UK: Cambridge University Press.

Ticona, Julia, and Alexandra Mateescu. 2018. “Trusted Strangers: Cultural Entrepreneurship on Domestic Work Platforms in the On-Demand Economy.” *New Media & Society* 20(11):4384–4404.

Timmermans, Stefan, and Marc Berg. 2003. *The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care*. Philadelphia: Temple University Press.

Tripodi, Francesca. 2019. “Google and Censorship through Search Engines.” U.S. Senate Committee on the Judiciary, July 16. Retrieved December 20, 2020. https://www.judiciary.senate.gov/imo/media/doc/Tripodi%20Testimony.pdf.

Turing, Alan. [1948] 2004. “Machine Intelligence.” In *The Essential Turing: The Ideas That Gave Birth to the Computer Age*, edited by B. Jack Copeland. Oxford, UK: Clarendon.

UNHCR. 2020. “Figures at a Glance.” Retrieved May 29, 2020. https://www.unhcr.org/en-us/figures-at-a-glance.h.

United Nations. 2017. “The International Migration Report.” UN Department of Economic and Social Affairs. Retrieved July 10, 2020. https://www.un.org/development/desa/publications/international-migration-report-2017.html.

United Nations. 2020. “AI for Good: Global Summit.” Retrieved May 29, 2020. https://aiforgood.itu.int/.

Vallas, Steven, and Juliet B. Schor. 2020. “What Do Platforms Do? Understanding the Gig Economy.” *Annual Review of Sociology* 46:273–94.

van Dijck, Jose. 2014. “Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology.” *Surveillance & Society* 12(2):197–208.

Viscelli, Steve. 2018. “Driverless? Autonomous Trucks and the Future of the American Trucker.” Center for Labor Research and Education, University of California, Berkeley, and Working Partnerships USA. Retrieved February 23, 2021. http://driverlessreport.org/files/driverless.pdf.

Wakabayashi, Daisuke. 2018. “Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam.” *The New York Times*, March 19.

Walsh, Bryan. 2020. “AI and Automation are Creating a Hybrid Workforce.” *Axios*, October 31.

Whooley, Owen. 2019. *On the Heels of Ignorance: Psychiatry and the Politics of Not Knowing*. Chicago: University of Chicago Press.

Williams, April, Zaida Bryant, and Christopher Carvell. 2019. “Uncompensated Emotional Labor, Racial Battle Fatigue, and (In)Civility in Digital Spaces.” *Sociology Compass* 13(2):e12658.

Williams, David R., and Michelle Sternthal. 2010. “Understanding Racial/Ethnic Disparities in Health: Sociological Contributions.” *Journal of Health and Social Behavior* 51(Suppl.):S15–27.

Author Biographies

Kelly Joyce is a professor of sociology and the founding director of the Center for Science, Technology, and Society at Drexel University. Her research is situated at the crossroads of medical sociology and science and technology studies; she conducts research on the social, cultural, and political dimensions of medical technology innovation. Results of this research have been published in her book *Magnetic Appeal: MRI and the Myth of Transparency* (Cornell University Press, 2008), her coedited volume *Technogenarians: Studying Health and Illness through an Aging, Science, and Technology Lens* (Wiley-Blackwell, 2010), and scholarly journals such as *Sociology of Health and Illness, Social Studies of Science, and Engaging Science, Technology, and Society*. Her research on the social dimensions of algorithms and smart textiles has been funded by awards from the National Science Foundation and the National Institutes of Health.

Laurel Smith-Doerr is a professor of sociology at the University of Massachusetts at Amherst. She studies organizational contexts for equity in science and technology and processes in collaborative knowledge production. She won a 2020 Fulbright fellowship to investigate AI knowledge production, equity, and recruitment of tech workers in Germany. She coedited the 2017 *Handbook of Science and Technology Studies* (MIT Press) and won the STS...
Infrastructure Award (Society for Social Studies of Science). Her article with Sharla Alegria, Kay Husbands Fealing, Debra Fitzpatrick, and Donald Tomaskovic-Devey in the American Journal of Sociology provides a new organizational-level theory of the gender pay gap and won the 2020 Devah Pager award for best article from the Inequality, Poverty, and Mobility Section of the American Sociological Association. She is principal investigator of the $3 million National Science Foundation ADVANCE grant at the University of Massachusetts at Amherst. Currently she serves as chair of the Science, Knowledge and Technology Section of the American Sociological Association and vice president–elect of the Eastern Sociological Society.

Sharla Alegria is an assistant professor of sociology at the University of Toronto. She earned her PhD in sociology with a certificate in women, gender, and sexuality studies and global health at the University of Massachusetts at Amherst in 2016. Two broad goals drive her research agenda: (1) to understand how inequalities persist when individuals and institutions publicly reject discrimination and (2) to understand the equity related consequences of the shift toward flexible workplace practices, especially in knowledge-based, globally interconnected work.

Susan Bell is a professor of sociology in the Department of Sociology at Drexel University, where she is a faculty affiliate in the Center for Science, Technology & Society and the Urban Health Collaborative, Dornsife School of Public Health. Her research examines patient cultures, embodied health movements, visual and performative ways of understanding illness, and global health. She is author of DES Daughters: Embodied Knowledge and the Transformation of Women’s Health Politics (Temple, 2009); guest editor with Alan Radley of a special issue of Health, “Another Way of Knowing: Art, Disease, and Illness Experience” (2011); and coeditor with Anne Figert of Reimagining (Bio)Medicalization, Pharmaceuticals and Genetics: Old Critiques and New Engagements (Routledge, 2015). Recent articles include “Placing Care: Embodying Architecture in Hospital Care for Immigrant/Refugee Patients” (Sociology of Health & Illness, 2018) and “Interpreter Assemblages: Caring for Immigrant and Refugee Patients in US Hospitals” (Social Science & Medicine, 2019).

Taylor Cruz is an assistant professor of Sociology at California State University, Fullerton. She studies the social, political, and ethical dimensions of big data and health tech. Her current project examines the evolving political landscape of health care following the adoption of electronic health records within large integrated delivery systems. By combining key informant interviews with ethnographic fieldwork, her research seeks to answer the following: How have electronic health records and data-driven technologies transformed clinical care, system administration, and public policy making within the health care arena? She has also published on the politics of measuring social difference within population-based data systems and the impact of stigma on access to care for underserved populations. She is the recipient of numerous honors and awards, including the Diana Forsythe Award for Social Studies of Science, Technology, and Health and the National Science Foundation Graduate Research Fellowship.

Steve G. Hoffman is an assistant professor of sociology at the University of Toronto. His scholarship and teaching orbit around sociological theory; the cultural politics of science, knowledge, and technology (e.g., AI); and the practical use of “other realities.” Recent publications can be found in Theory and Society, Science, Technology, & Human Values, Sociological Forum, Cultural Sociology, and Social Studies of Science. He is currently the principal investigator of a multisited study, funded by a Social Sciences and Humanities Research Counsel Insight Grant through 2023, on the production and use of simulation techniques among disaster managers in the greater Toronto area. Despite nostalgia for the avocado and tangerine trees of his southern California youth, he has largely imbibed what it means to be a dual citizen of the Great Lakes region of Turtle Island.

Safiya Umoja Noble is an associate professor at the University of California, Los Angeles, in the Departments of Information Studies and African American Studies, and is codirector of the UCLA Center for Critical Internet Inquiry. She is the author of Algorithms of Oppression: How Search Engines Reinforce Racism and coeditor of The Intersectional Internet: Race, Sex, Culture and Class Online and Emotions, Technology & Design. She holds a PhD and an MS from the Information School at the University of Illinois at Urbana-Champaign and a BA in sociology from Fresno State.

Benjamin Shestakofsky is an assistant professor of sociology at the University of Pennsylvania. His research centers on how digital technologies are affecting work and employment, organizations, and economic exchange. He is currently developing a book manuscript that examines how venture capital shapes experiences of work in a tech startup. His article on software automation and the future of work, published in in Work and Occupations, was awarded the 2019 W. Richard Scott Award for Distinguished Scholarship by the American Sociological Association’s Section on Organizations, Occupations, and Work.