DATA & SOCIETY

POLICY BRIEF

Explainer: A Sociotechnical Approach to AI Policy

Brian J. Chen
Jacob Metcalf

Introduction

Over the last year, as lawmakers and regulators worldwide assembled governance frameworks for AI, astute observers will have noticed the increasing prevalence of the term "sociotechnical" in policy documents and discussions.[1] The term may be relatively unfamiliar in law and governance, but it has historical applications that are well understood across a range of contexts. **Generally, a sociotechnical perspective means viewing society and technology together as one coherent system. In other words, it is not possible to understand the "social" without the "technical," nor the "technical" without the "social." Explaining the outcomes of any technology requires focusing on the messier "middle ground" between these two realms.**

Despite its increasing adoption in AI governance and industry circles, or perhaps because of it, the term "sociotechnical" may be among the most misunderstood in AI policy. Drawing on established literature and historical and present-day examples, this brief explains what a sociotechnical perspective is and why it matters in policy.

In their efforts to mitigate AI's negative effects on society — including its harm to workers, marginalized groups, the environment, and even the risk of catastrophic injury to humanity — many regulators and technologists are overly focused on technical solutions: finetuning algorithms for fair outputs, documenting the engineering process, correcting the training dataset, and finding vulnerabilities. These actions are all necessary for safe AI, but such a limited lens implies a narrow scope of responsibility that begins and ends with the technical system.

A sociotechnical approach recognizes that a technology's real-world safety and performance is always a product of technical design and broader societal forces, including organizational bureaucracy, human labor, social conventions, and power. By attending to that bigger picture, policymakers can advance solutions that more comprehensively address AI's sociotechnical complexities.

---

1 *See, e.g., Artificial Intelligence Risk Management Framework (AI RMF 1.0)* 1, National Institute of Standards and Technology (Jan. 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf ("AI systems are inherently socio-technical in nature[.]"); *Introducing the AI Safety Institute*, GOV.UK (Jan. 17, 2024), https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute (calling for "the sociotechnical infrastructure needed to understand the risks of advanced AI and enable its governance"); Seth Lazar and Alondra Nelson, *AI safety on whose terms*?, Science (July 13, 2023), https://www.science.org/doi/10.1126/science.adi8982 ("Only a sociotechnical approach can truly limit current and potential dangers of advanced AI.").

Background: What is a Sociotechnical Perspective?

A sociotechnical perspective draws a wide circle around both the technical and social components of a system. This perspective recognizes that the performance, effectiveness, and downstream consequences of technologies derive neither from technical design nor social dynamics in the abstract, but from the real-world interplay between the two. Although it may seem like an academic difference, this approach is distinct from the truism that technology *influences, impacts, or affects* society (and vice versa), as if the two are distinct entities that occasionally interact under specific circumstances. Rather, a sociotechnical approach asserts that technology and society are inextricable.

A sociotechnical analysis does not preclude a technical analysis. Indeed, it should be seen as a complement to technical analysis, which is necessary to ask and answer certain questions about technologies: does the code work, are the mechanics efficient, does it operate reliably inside of laboratory conditions, does it meet the technical specifications of the customer or regulator? A sociotechnical analysis asks an additional set of questions that are critical to understanding a system's real-world performance and consequences. Sociotechnical analysis is particularly relevant to deployment, where the systems actually meet people outside of the research and development contexts.[2]

A sociotechnical perspective is certainly broader than a technical perspective, but it is broader in a specific way: the "social" is not appended to the "technical" as an afterthought, but is integral for any explanation of the who, what, why, how, and when of a system's outcomes. A sociotechnical perspective is most interested in what is "in the middle" between the social and the technical, as an explanatory mechanism for a system's success or failure.

The concept of viewing technologies as "sociotechnical systems" is not new, even if it is newly prominent in policy and regulatory parlance. The term itself emerged from a group of British sociologists at the conclusion of their work for the British military during World War II. The sociologists formed an independent research and consulting institute, the Tavistock Institute for Human Relations.[3] One of their earliest clients, the owner of a large coal mine, was struggling to implement new large-scale industrial mining machinery. On paper, this new technology, which could shave enormous faces of coal off an open mine wall, should have resulted in substantial increases of output per worker. Yet widespread dissatisfaction among the workforce, and resulting labor slowdowns, were erasing efficiency gains.

2 Science and technology studies scholars would insist that even inside the lab any technical system is also a sociotechnical system. Their claim is correct, but given our policy and regulatory focus here we contend that there is a relevant distinction between pre- and post-deployment questions.

3 Eric L. Trist, *The Evolution of Socio-Technical Systems*, Ontario Quality of Working Life Centre (1981), *available at* https://www.lmmiller.com/blog/wp-content/uploads/2013/06/The-Evolution-of-Socio-Technical-Systems-Trist.pdf.

The Tavistock researchers set out to observe and interview the workers about why this was happening. They found that the traditional model of small-scale coal mining, despite being a dangerous and unsentimental form of drudgery, was integral to the miners' social cohesion and self-identity. Under that model, the miners were divided into small and stable teams, which were independently responsible for mining coal from small-scale tunnels. They were paid by weight, which provided immediate feedback, professional pride, and a source of motivating competition with other teams. *In their experience, the social and the technical were inseparable:* the physical layout of the mine and the form of the mining technologies were integral to how the coal miners understood and ordered their social world — even outside the mine.

When the mine owner imposed the new large-scale machines, requiring many workers to operate them all at once, this particular model of social cohesion ceased. The machinery demanded a new metric for productivity that aggregated and anonymized the workers, deskilling them from the familiar small-team roles and undermining the social structures through which the workers understood themselves. The Tavistock researchers determined that restoring efficiency of the mine would require understanding the machinery and social structures of the miners as a single unit: a "**socio-technical system**." As social theorists in a very pragmatic environment, the leverage point they found was not in manipulating the social or the technical alone, but in the *middle ground* between those domains, at the interaction between humans and machines. The Tavistock researchers coined "socio-technical" because it enabled them, and their clients, to understand the social and the technical aspects of the industrial system as fundamentally intertwined, and thereby repair some of the harmful outcomes of the adoption of the new technology.

The analytic value of "socio-technical" systems theory led to its widespread adoption across multiple disciplines. Management and organizational theorists adopted "socio-technical systems theory"[4] to explain, much like the Tavistock researchers, how the organizational structures of a workplace need to mesh with technical and software systems in order to find their full value and better shape the design of information systems.[5] Human-computer interaction (HCI) researchers adopted the terminology to identify how successful software or web platforms mesh with, and adapt to, the organic needs of users and communities who modify or mold them according to their own desires.[6] "Socio-technical" systems theory in this vein has been applied to a wide range of contexts including cybersecurity,[7] manufacturing,[8] informa-

---

4 Often abbreviated as "STS."

5 Enid Mumford, *A Socio-Technical Approach to Systems Design*, 5 Requirements Engineering 125 (2000), https://doi.org/10.1007/PL00010345; Enid Mumford, *The Story of Socio-Technical Design: Reflections on Its Successes, Failures and Potential*, 16 Information Systems Journal 317 (2006), https://doi.org/10.1111/j.1365-2575.2006.00221.x.

6 Brian Whitworth, *A Brief Introduction to Sociotechnical Systems*, Encyclopedia of Information Science and Technology 394 (2d ed.) (2009), https://www.igi-global.com/chapter/brief-introductionsociotechnical-systems/13604.

7 Masike Malatji, Sune Von Solms, and Annlizé Marnewick, *Socio-Technical Systems Cybersecurity Framework*, 27 Information & Computer Security 233 (2019), https://doi.org/10.1108/ICS-03-2018-0031.

8 Marlon Soliman and Tarcisio Abreu Saurin, *Lean Production in Complex Socio-Technical Systems: A Systematic Literature Review*, 45 Journal of Manufacturing Systems 135 (Oct. 2017), https://doi.org/10.1016/j.jmsy.2017.09.002.

tion and communication technologies,[9] the electrical grid,[10] and distributed knowledge systems like Wikipedia.[11] The resulting studies consider not just the role of individual people and pieces of technology, but regulations and laws, the physical environment, ecosystems, software, hardware, networks, and data structures.

Researchers from the fields of science and technology studies[12] and the sociology of science adopted "sociotechnical" (sans hyphen[13]) a little later, and it has since become a central methodological conceit of those fields. Dutch philosopher-engineer-sociologist Wiebe Bijker and collaborators revitalized the term in order to emphasize the social construction of technologies and scientific knowledge.[14,15] Bijker's study of the history of bicycling, investigating why some of the early designs of bicycles survived while others did not, focused on the myriad social dynamics (especially gender) around this new form of transportation, and the push-and-pull between evolving social expectations and technical experimentation.[16] Later, science policy scholar Sheila Jasanoff proposed "sociotechnical imaginaries" to describe how nations coalesce around visions of good technological futures, use policy to stabilize and support those visions, and use science and technology to project their distinct forms of power.[17] For example, the major forces in the global AI economy (the US, EU, and China) have distinct visions of how AI will contribute to desirable technological futures, and how government policy can push that vision forward while limiting what they conceive of as harmful

---

9 Roberta Lamb, Steve Sawyer, and Rob Kling, *A Social Informatics Perspective on Socio-Technical Networks*, AMCIS 2000 Proceedings (2000), https://aisel.aisnet.org/amcis2000/1/; Steve Sawyer and Mohammad Hossein Jarrahi, *The Sociotechnical Perspective*, *in* Information Systems and Information Technology (Topi Hekki and Allen Tucker eds., 2015).

10 Pedro H.J. Nardelli and Florian Kühnlenz, *Why Smart Appliances May Result in a Stupid Grid: Examining the Layers of the Sociotechnical Systems*, 4 IEEE Systems, Man, and Cybernetics Magazine 21 (2018), https://doi.org/10.1109/MSMC.2018.2811709.

11 Sabine Niederer and José van Dijck, *Wisdom of the Crowd or Technicity of Content? Wikipedia as a Sociotechnical System*, 12 New Media & Society 1368 (2010), https://doi.org/10.1177/1461444810365297; Mir Saeed Damadi and Alan Davoust, *Fairness in Socio-Technical Systems: A Case Study of Wikipedia*, Collaboration Technologies and Social Computing 84 (Hideyuki Takada, et al., eds., 2023), https://doi.org/10.1007/978-3-031-42141-9_6.

12 Also often abbreviated as "STS," confusingly for anyone working in this space.

13 The use of a hyphen (socio-technical vs sociotechnical) is not particularly important, but it does indicate which genealogy of the term is being deployed. One will see the use of a hyphen more often in the business and management studies contexts. Because Data & Society researchers are mostly trained in the science and technology studies traditions, we prefer to use the version without a hyphen.

14 Wiebe E. Bijker, Thomas Parke Hughes, and Trevor Pinch, eds., *The Social Construction of Technological Systems, Anniversary Edition: New Directions in the Sociology and History of Technology* (2012); Wiebe E. Bijker and John Law, eds., *Shaping Technology/Building Society: Studies in Sociotechnical Change* (1992); Trevor J. Pinch and Wiebe E. Bijker, *The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other*, 14 Social Studies of Science 399 (1984).

15 The "social construction" of science and technology does not mean that knowledge is made up according to the whims of scientists and engineers. Rather, it means that social practices (e.g., universities, labor conditions, peer review, national funding priorities, et cetera) both enable and set the parameters of science and innovation.

16 For example, gender norms played a surprisingly significant role in the rapid changes in bicycle design — women suddenly had much greater freedom of travel and some designers sought to limit that through technical and moral means, whereas other designers sought to accommodate and/or regulate new forms of daredevil behavior among young men.

17 Sheila Jasanoff, *Future Imperfect: Science, Technology, and the Imaginations of Modernity*, *in* Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power 1 (Aug. 2015).

outcomes.[18] These "imaginaries" are by nature contested, evolving, and not entirely coherent, but nonetheless help explain how different nations have distinct and recognizable approaches to technological development.

The Inconsistent Meanings of "Sociotechnical" Today

Despite the well-established history of sociotechnical theory in academic research, the term has seen widely varying meanings among AI policymakers and industry.

The term "sociotechnical" is sometimes left entirely undefined, such as in official government policy,[19] or defined loosely as "the interaction between technology and society" (an accurate, if banal, description). Research labs at leading AI companies frequently invoke "sociotechnical" to describe some of their work.[20] Responsible AI teams have adopted the term to describe algorithmic harms and the safety practices intended to protect against them.[21] Other times, it is used as a foil to "techno-solutionism," the belief that social problems have simple technical solutions. Our point is not to belabor pedantically that some uses are right and others wrong; what matters is that varying, imprecise language is obscuring the specific value of sociotechnical perspectives in AI policy.

Viewed more cynically, the term has arguably been used by some technologists to wave away difficult social complexities that, in their view, belong outside engineers' responsibilities (an interpretation that is counter to the basic idea of sociotechnical systems). Further, because it is used in such imprecise ways, "sociotechnical" governance has found itself an easy target in the AI "culture war," which sees any concern about discrimination and marginalization as contrary to AI's innovations.[22]

---

18 *See also* Amba Kak and Sarah Myers West, eds., *AI Nationalism(s): Global Industrial Policy Approaches to AI*, AI Now Institute (March 2024), https://ainowinstitute.org/ai-nationalisms.

19 *Blueprint for an AI Bill of Rights*, Office of Science and Technology Policy (Oct. 2022), https://www.whitehouse.gov/ostp/ai-bill-of-rights/. Understandably, the Blueprint's authors likely decided that the document's purpose was to advance sociotechnical principles, not a sociotechnical definition.

20 Microsoft Social Media Collective, *Sociotechnical Systems*, Microsoft Research (undated), https://www.microsoft.com/en-us/research/theme/sociotechnical-systems/; Jan Leike & Ilya Sutskever, *Introducing Superalignment*, OpenAI (blog) (July 5, 2023), https://openai.com/blog/introducing-superalignment; Jamila Smith-Loud, *Responsible AI at Google Research: The Impact Lab*, Google Research (blog) (Mar. 16, 2023), http://research.google/blog/responsible-ai-at-google-research-the-impact-lab/.

21 *See, e.g.*, Renee Shelby, et al., *Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction*, Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society 723 (Aug. 2023), https://doi.org/10.1145/3600211.3604673; Laura Weidinger, et al., *Sociotechnical Safety Evaluation of Generative AI Systems*, arXiv (Oct. 31, 2023), http://arxiv.org/abs/2310.11986; Bertie Vidgen, et al., *Introducing v0.5 of the AI Safety Benchmark from MLCommons*, arXiv (April 18, 2024), http://arxiv.org/abs/2404.12241; Usman Anwar, et al., *Foundational Challenges in Assuring Alignment and Safety of Large Language Models*, arXiv (April 15, 2024), http://arxiv.org/abs/2404.09932.

22 *See* Will Knight, *The AI Culture Wars Are Just Getting Started*, WIRED (Feb. 29, 2024), https://www.wired.com/story/fast-forward-ai-culture-wars-just-getting-started/; Ben Schreckinger, *Inside the AI Culture*

But framing AI as a sociotechnical system is not so much about a particular set of issues as it is an empirical method to observe how technologies operate in a broader context. It is an historically useful way to view technology and society as a single, coherent unit in order to better understand a technology's outcomes. It is not a synonym for "messy" or a bucket for everything that is "too hard to do." Nor is it a placeholder for "ethical" or "good" (or, for what it's worth, for "woke").

Most importantly for policy: by seeing the "social" and "technical" together, a sociotechnical perspective expands the scope of responsibility for the consequences of technologies. It undermines the notion that developers are only responsible for their products up to the point of deployment and no further. It calls into question the belief that lawmakers' role is to "protect innovation" from regulatory burdens, which effectively disclaims the responsibility to govern technologies. Treating technology and society as one coherent unit means that developers, deployers, and regulators are accountable not just for the mechanical workings of technical machines, but for how those machines integrate with, reshape, and sometimes harm social systems.

Three Sociotechnical Questions

For specific policy recommendations reflecting a sociotechnical approach, check out our companion resource: "Answering Three Sociotechnical Questions."

The following section offers an initial set of questions to govern AI systems in a sociotechnical manner. This is not a comprehensive policy toolkit addressing all the ways one should think about technology issues. Instead, it outlines several starting points to begin to apply a sociotechnical framework.

1. WHAT ARE THE SYSTEMS AROUND THE TECHNOLOGY?

With any new technology, there is a tendency to focus on the machine itself, divorced from any particular social configuration. With AI, the marketing incantation of "intelligence," evoking a self-functioning machine capable of solving general problems, acutely marginalizes the role of outside social factors in AI's deployment.[23] The consequence has been an inordinate focus on machine learning techniques and technical specs. Standards for AI governance,

---

War, Politico (May 15, 2023), https://www.politico.com/newsletters/digital-future-daily/2023/05/15/inside-the-ai-culture-war-00096963.

23 *See* Matteo Pasquinelli, *The Eye of the Master: A Social History of Artificial Intelligence* 9 (2023).

for example, tend to closely index technical considerations and push aside an accounting of social factors.[24,25]

The problem with a strictly technical approach is that many of AI's benefits and harms will come from the broader system around the technology, i.e. from the interaction between the technical and the social. An AI system can be perfectly functional yet fail at the point of use due to the way people use it or the unintended ways it interacts with institutions and processes. The foremost question, from a sociotechnical perspective, is not how sophisticated is the technology's design or how advanced are its capabilities, but rather what are the technical and societal dynamics that *together* make the system operate as it does, and produce the outcomes that it does.

To answer that question, a sociotechnical approach encourages people to consider the broader system at play. Policymakers thinking about AI should view the model in question — whether a generative AI system, an algorithm to determine public benefits eligibility, or predictive policing used by law enforcement — as part of a sociotechnical system, and bring the "external" or "non-technical" components of that system into policy consideration.[26]

For example, there are many claims about AI's ability to better detect medical conditions, thus improving patient outcomes and quality of care.[27] Because machine learning technologies can quickly and efficiently recognize patterns in large data sets, much discussion around AI's ability to diagnose cancer credits the "technical advances of the time[.]"[28]

But research on the effective integration of new technologies shows that "AI interventions must always be thought of as sociotechnical systems, in which social context, relationships, and power dynamics are central, not an afterthought."[29] AI's innovations in medicine, for example, will depend on all these system forces working with the machine learning tool. Even

---

24 During the European Union AI Act negotiations, a joint committee proposal in the European Parliament to require that AI standards "also take into account risks to fundamental rights, the environment, and society as a whole and other democratic and sociotechnical aspects of the AI system" was rejected. The final text focuses on "technical standards." *REPORT on the Proposal for a Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts,* European Parliament Committee on the Internal Market and Consumer Protection & Committee on Civil Liberties, Justice and Home Affairs (May 2, 2023), https://www.europarl.europa.eu/doceo/document/A-9-2023-0188_EN.html.

25 *See* Alicia Solow-Niederman, *Can AI Standards Have Politics*?, 71 UCLA L. Rev. Disc. 2 (2023).

26 *See* Andrew D. Selbst, et al., *Fairness and Abstraction in Sociotechnical Systems* 60, Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT '19) (Jan. 2019) https://dl.acm.org/doi/10.1145/3287560.3287598.

27 *See, e.g.*, Nadia Jaber, *Can Artificial Intelligence Help See Cancer in New, and Better, Ways*?, National Cancer Institute (Mar. 22, 2022), https://www.cancer.gov/news-events/cancer-currents-blog/2022/artificial-intelligence-cancer-imaging.

28 Bo Zhang, et al., *Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach*, 16 Journal of Multidisciplinary Healthcare 1779, 1787 (June 2023), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10312208/.

29 Madeleine Clare Elish and Elizabeth Anne Watkins, *Repairing Innovation: A Study of Integrating AI in Clinical Care*, Data & Society Research Institute (Sept. 2020), https://datasociety.net/library/repairing-innovation/.

the most precise diagnostic model will fail if it cannot be integrated into existing systems of patient care.

CASE STUDY: Studying a Duke University Hospital program called SepsisWatch, which deploys machine learning to help clinicians identify and treat sepsis (a difficult-to-diagnose and often deadly condition with a variety of complex indicators), researchers found that the system's effectiveness was due not only to the machine learning tools but also to the nurses, doctors, and other staff whose adaptations to the technology "repaired" the AI's disruptions.[30] The AI screening system altered workflows, changed hierarchies, broke unspoken rules, and shifted social and professional relationships. In seeing SepsisWatch as a sociotechnical system, the researchers observed the human labor needed to repair procedural breakdowns and create new forms of interaction and communication. Most notably, the designers of the system had not considered the fact that nurses and medical technicians would be responsible for the majority of the labor of interacting with the program and interpreting its behavior and outputs for patients. The designers had modeled the scientific authority of physicians' diagnostic skills but did not account for the supporting labor that is critical for actually delivering medical care in a hospital setting. For the technical innovation to function properly, it would require training and integrating other staff members, not just physicians.[31]

A sociotechnical approach understands that a technology's technical innovations do not occur in a vacuum. The system requires human labor, reconfigurations of practices, and continual real-world adaptations to "repair" its disruptions. In this case, those practices — the "middle ground" between the machine learning tool (the "technical") and the staff now tasked with using it (the "social") — made the difference.

Like the old parable about trees and forests, it is all too common for those working in tech policy to focus so much on the technology that they can't make out the broader system. But to effectively govern emerging technologies like AI, policymakers will need to consider the social dynamics surrounding the technology, explicitly drawing them into the policymaking process.

2. WHAT IS THE PROBLEM THIS TECHNOLOGY CAN ACTUALLY SOLVE?

Because a sociotechnical lens views any technology in light of larger social structures, it rejects the assumption that social problems can be easily solved by new tech.[32] Digital technologies offered as a "solution" to social problems often fail to meet their promised purpose, and they sometimes end up worsening the problem they set out to solve.[33] Sociotechnical scholarship demonstrates how technological fixes fail when their theory of change

30 *Id.*

31 *Id.* at 27.

32 *See, e.g.*, Max Oelschlaeger, *The Myth of the Technological Fix,* Southwest Journal of Philosophy (1979); Eugene M. Burns and Kenneth E. Studer, *Reflections on Alvin M. Weinberg: a case study on the social foundations of science policy*, Research Policy (Mar. 1975), https://www.sciencedirect.com/science/article/abs/pii/0048733375900098.

33 *See* Lily Irani, *Chasing Innovation: Making Entrepreneurial Citizens in Modern India* (2019); David Greene, *The Promise of Access* (2021).

— introducing or improving tech design to solve social problems — comes up against complex issues where technology is, at most, one factor among many (political, economic, and cultural).[34]

Tech vendors have claimed that AI will solve the day's troubles. AI has been positioned as a fix for crime, global warming, health care, war, access to public benefits, fraud in public benefits, and a lack of diverse representation in the media, among many other things. A common policy response to such AI applications is to call for the technical components of the AI system to be accurate and algorithmically "fair" (so to speak). But sound technical engineering may fail to improve, and may even exacerbate, the systemic conditions that produce algorithmic harms. A sociotechnical analysis encourages policymakers to consider the limits of solving longstanding structural problems with new technology.

CASE STUDY: The use of electronic monitoring in care work is one example of a techno-solutionist policy approach that has exacerbated the initial problem.[35] In 2016, federal lawmakers mandated that Medicaid-funded home health aides use an Electronic Visit Verification (EVV) app, which tracks workers' locations and requires them to log their activities.[36] The idea for the tech vendors was simple enough: by closely monitoring workers' activities and switching to a digital platform, the technology would ensure workers' time was not "wasted" on idleness or non-care activities, thus improving efficiency for cash-strapped agencies. Many federal policymakers, meanwhile, cared about reducing "fraud, waste, and abuse" in Medicaid-funded services.

But where the app was intended to improve the delivery of care, it has had the opposite effect. Lawmakers mandated that EVV be "minimally burdensome," but scant federal policy guidance resulted in obtrusive state-level policies and design choices. Seemingly designed for a centralized form of management, the technology did not align with the real world experiences of home health aides in self-directed programs. The app was designed to maximize workers' time on certain care activities, but much of home health aides' labor cannot be quantified and reduced to discrete tasks. The result has been "destabilizing," with workers feeling burdened by technology that makes little sense in their jobs and is often poorly designed and maintained, while also feeling surveilled and, in some cases, criminalized.[37]

A sociotechnical approach, at a minimum, would have considered the existing workflows of home health aides and the disruption of a new technology, as well as accounting for workers' *reaction and adaptation to the technology* as integral to the program's success. It might have affirmatively sought input from Medicaid beneficiaries on their at-home caregiving needs. It might have also considered the fraught politics of "fraud" accusations in a low-wage sector predominantly made up of women of color.[38] More to the point, policymakers adopting a sociotechnical approach might have considered whether surveillance

34 *See* Donald A. MacKenzie and Judy Wajcman, eds., *The Social Shaping of Technology* (2011).

35 Alexandra Mateescu, *Electronic Visit Verification: The Weight of Surveillance and the Fracturing of Care*, Data & Society Research Institute (Nov. 16, 2021), https://datasociety.net/library/electronic-visit-verification-the-weight-of-surveillance-and-the-fracturing-of-care/.

36 21st Century Cures Act. H.R. 34, 114th Congress (2016).

37 Mateescu, *supra* note 35.

38 *Id.* at 20.

apps were appropriate in the first place, given deep and enduring fissures in the US care infrastructure.

3. WHAT POWER INEQUALITIES ARE AT PLAY?

Seen through a sociotechnical lens, many technology problems are better understood as problems of power. Because it considers the social forces influencing a given technology's real-world performance, a sociotechnical analysis investigates the underlying power dynamics to better understand an issue and craft a solution.

Power, as "the property of broader social, economic, cultural, and political networks, institutions, and structures," manifests in many ways.[39] Power shows up as the concentration of capital, infrastructure, and R&D in a handful of market-dominant firms.[40] It shows up in the outsize influence of technologists with the resources to steer AI safety discourse and policy.[41] It shows up in the ability of automated systems to shape people's opportunities and outcomes — or, in other words, "to determine what type of people we can be."[42] And power, of course, is central to the staggering degree of surveillance needed for AI to function.[43]

CASE STUDY: Efforts to advance AI safety that fail to consider power relations may end up exacerbating inequalities. For example, many policymakers and technologists are calling attention to the datasets used to train AI, noting that biased or explicit data can result in harmful outputs ("garbage in, garbage out").[44] But one of the main mechanisms to mitigate AI's discriminatory or negative impacts — cleaning and debiasing datasets — has arguably intensified tech firms' reliance on underpaid, contracted-out labor.[45] To limit the production of toxic language in the AI chatbot GPT-3 (ChatGPT's predecessor), designers developed tools to recognize and remove such outputs. In order to produce the training data that enables the system to recognize inappropriate language, OpenAI contracted with an outsourcing firm in Kenya that

---

39 Faridun Sattarov, *Power and Technology: A Philosophical and Ethical Analysis* 20 (2019).

40 Amba Kak and Sarah Myers West, *AI Now 2023 Landscape: Confronting Tech Power,* AI Now Institute (Apr. 11, 2023), https://ainowinstitute.org/2023-landscape; *see also* Corinne Cath, *Introduction: eaten by the internet*, *in* Eaten by the Internet (Corinne Cath, ed., Oct. 10, 2023).

41 Shazeda Ahmed, et al., *Field-Building and the Epistemic Culture of AI Safety*, First Monday (April 14, 2024), https://doi.org/10.5210/fm.v29i4.13626; Timnit Gebru and Émile P. Torres, *The TESCREAL Bundle: Eugenics and the Promise of Utopia through Artificial General Intelligence*, First Monday (April 14, 2024), https://doi.org/10.5210/fm.v29i4.13636.

42 Jenna Burrell and Jacob Metcalf, *Introduction for the special issue of "Ideologies of AI and the consolidation of power": Naming power*, First Monday (April 14, 2024), https://firstmonday.org/ojs/index.php/fm/article/view/13643.

43 *See generally* Aiha Nguyen, *The Constant Boss*, Data & Society Research Institute (May 2021), https://datasociety.net/library/the-constant-boss/; Jennifer Lee, *Power and Technology: Who Gets to Make the Decisions?*, ACM Interactions (Jan.-Feb. 2021), https://dl.acm.org/doi/pdf/10.1145/3442420.

44 *See, e.g.*, IBM Data and AI Team, *Shedding Light on AI Bias with Real World Examples*, IBM (Oct. 16, 2023), https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/ ("Eliminating AI bias requires drilling down into datasets, machine learning algorithms and other elements of AI systems to identify sources of potential bias.").

45 *See* Agathe Balayn and Seda Gürses, *Beyond Debiasing: Regulating AI and its inequalities,* European Digital Rights (Sept. 2021), https://edri.org/wp-content/uploads/2021/09/EDRi_Beyond-Debiasing-Report_Online.pdf.

required workers to manually label the data.[46] Consequently, to avoid exposing users to potentially traumatizing content, the company instead exposed poorly paid workers in Kenya to content such as sexual abuse, violence, and racist hate speech — some of which included content illegal under US law. While the business decision to outsource this labor was ostensibly motivated by a desire to mitigate harms, this practice took advantage of legal and economic power imbalances to shift the harms to an "invisible" population.[47]

Removing troubling or harmful data — if mediated *solely* through technical interventions — can end up buttressing global inequalities. While policymakers should attend to the technical issues contributing to traumatizing or discriminatory outputs, they should simultaneously recognize that all parts of the AI lifecycle (e.g., physical and cloud infrastructures, R&D, training, deployment, and monitoring) are an exercise of power.

Many of the workforces that underpin the development of AI systems are drawn from marginalized, precarious populations that include prison inmates, refugees, and other displaced persons.[48] These workers typically lack voice or representation within the broader organizational structures of the tech industry, both with regards to working conditions and ethical considerations. Tech worker unions, such as the Alphabet Workers Union, have sought to break down artificial distinctions between high-paid, salaried "tech" and temporary/contract "non-tech" workers in order to create avenues for collective representation.

If many technology problems are about the power of technologists and well-capitalized firms, it's also true that power exists where people have the agency and ability to improve, resist, or demand change to technological systems.[49] By considering the wider constellation of power dynamics, policymakers can advance additional policy solutions, like enhancing worker voice, to govern emerging technologies such as AI. Policies like data governance and algorithmic "fairness" are critical to safe and trustworthy AI, but they should be understood as complementing, and not replacing, traditional mechanisms of power-building (worker organizing and the enforcement of labor and employment laws, to name a few).[50]

---

46 Karen Ho, *Cleaning Up ChatGPT Takes Heavy Toll on Human Workers*, Wall Street Journal (July 24, 2023), https://www.wsj.com/articles/chatgpt-openai-content-abusive-sexually-explicit-harassment-kenya-workers-on-human-workers-cf191483.

47 *See* Mary L. Gray and Siddharth Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* (May 2019).

48 Morgan Meaker, *These Prisoners Are Training AI*, WIRED (Sept. 11, 2023), https://www.wired.com/story/prisoners-training-ai-finland/; Phil Jones, *Refugees help power machine learning advances at Microsoft, Facebook, and Amazon*, Rest of World (Sept. 22, 2021), https://restofworld.org/2021/refugees-machine-learning-big-tech/.

49 *See* Abeba Birhane, et al., *Power to the People? Opportunities and Challenges for Participatory AI*, arXiv (Sept. 15, 2022), https://arxiv.org/abs/2209.07572; Mona Sloane, et al., *Participation is not a Design Fix for Machine Learning*, Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22) (Oct. 2022), https://arxiv.org/abs/2007.02423.

50 *See* Alexandra Mateescu, *Challenging Worker Datafication*, Data & Society Research Institute (Nov. 2023), https://datasociety.net/library/challenging-worker-datafication/ ("Data governance in many cases may only be a small facet of much bigger challenges around worker power and often-weak union presence in many sectors."); *see also* Josh Bivens and Ben Zipperer, *Unbalanced labor market power is what makes technology—including AI—threatening to workers*, Economic Policy Institute (Mar. 28, 2024), https://www.epi.org/publication/ai-unbalanced-labor-markets/.

Conclusion

Recognizing the sociotechnical complexities of AI and data-centric technologies surely does not make the work of policymakers any simpler. Nonetheless, a sociotechnical perspective is necessary to advance more impactful policy solutions. Technologies emerge not only from machine learning innovations but from labor practices, power configurations, and social relations. Policymakers' approach to observe and understand AI — and their tools to regulate it — must be just as expansive.

Acknowledgments

The authors would like to thank the following individuals for offering their insights to an early draft of this brief: Marc Aidinoff, danah boyd, Joshua Kroll, Seth Lazar, and Laura Weidinger. We'd also like to thank our Data & Society colleagues Sareeta Amrute, Alexandra Mateescu, Aiha Nguyen, Serena Oduro, and Ranjit Singh who offered valuable feedback and support for our project. Additionally, conversations about the meaning of "sociotechnical" with Emanuel Moss, Elizabeth A. Watkins, Andrew Smart, and Atoosa Kasirzadeh were essential to this work.

Data & Society is an independent nonprofit research institute that advances new frames for understanding the implications of data-centric and automated technology. We conduct research and build the field of actors to ensure that knowledge guides debate, decision-making, and technical choices.

www.datasociety.net

@datasociety

Layout by Hector Sandoval

May 2024