The Sociotechnical Turn in AI Governance

Siddharth Mehrotra
University of Amsterdam & TU Delft
Amsterdam, The Netherlands
s.mehrotra@uva.nl

Eva De Winkel
TU Delft
Delft, The Netherlands
e.dewinkel@tudelft.nl)

Jacqueline Kernahan
TU Delft
Delft, The Netherlands
j.a.kernahan@tudelft.nl

Íñigo de Troya
TU Delft
Delft, The Netherlands
i.m.d.r.detroya@tudelft.nl

Sem Nouws
TU Delft
Delft, The Netherlands
s.j.j.nouws@tudelft.nl

Maurus Enbergs
TU Delft
Delft, The Netherlands
m.e.enbergs-1@tudelft.nl

Jin Huang
University of Amsterdam & TU Delft
Amsterdam, The Netherlands
j.huang2@uva.nl

Roel Dobbe
TU Delft
Delft, The Netherlands
r.i.j.dobbe@tudelft.nl

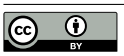
Abstract

The "Sociotechnical Turn" in artificial intelligence (AI) research and governance is a paradigm shift reflecting a growing recognition that AI systems must be understood within their broader social contexts. However, this recognition comes with a critical paradox: despite increasing references to "sociotechnical AI", interpretations have become fragmented, hampering coherent governance development. Therefore, through an ongoing narrative literature review, we are analyzing how researchers engage with these concepts, revealing distinct dimensions of sociotechnical thinking. In this paper, we are presenting our preliminary findings, focusing specifically on the conceptual dimensions of sociotechnical AI by examining how authors define, theorize, and engage with sociotechnical terminology. Our research offers significant value to stakeholders seeking genuine engagement with the complex interplay between technical systems and social contexts. By identifying the areas of convergence and divergence in sociotechnical AI research, we aim to offer theoretical clarity that can guide future AI governance approaches.

CCS Concepts

• **Computing methodologies** → **Artificial intelligence**; • **Human-centered computing** → **Human computer interaction (HCI)**; • **Social and professional topics** → **Government technology policy**.

ACM Reference Format:
Siddharth Mehrotra, Eva De Winkel, Jacqueline Kernahan, Íñigo de Troya, Sem Nouws, Maurus Enbergs, Jin Huang, and Roel Dobbe. 2025. The Sociotechnical Turn in AI Governance. In *First Workshop on Sociotechnical AI Governance: Opportunities and Challenges for HCI (STAIG '25), April 27, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/XX



STAIG '25, Yokohama, Japan

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/XX

1 Introduction

As artificial intelligence (AI) technology becomes increasingly integrated into society, the societal effects and the forces behind its creation are becoming evident to a broader audience, including technical experts, social scientists, policymakers, and the public. This growing awareness highlights the need for governance and regulation of AI systems [10, 31]. In response, various efforts have emerged to anticipate and address the implications of AI through governance strategies [7, 33]. Early initiatives introduced ethical principles and guidelines, as well as technical tools to address issues of fairness, accountability, and transparency [27]. However, as the limitations of both ethical and technical approaches have become more apparent, many scholars are now advocating for a sociotechnical approach to AI governance, often referred to as *Sociotechnical AI* [4, 9, 12].

We argue that AI research is experiencing what we term a "Sociotechnical Turn" — a significant paradigm shift characterized by the growing adoption of sociotechnical terminology and frameworks across academic, industry, and policy domains. This turn represents an evolution from earlier governance approaches that relied primarily on ethical principles or technical fixes to address issues of fairness, accountability, and transparency [27]. As the limitations of these siloed approaches became more apparent, sociotechnical perspectives emerged as a promising pathway to integrate technical considerations with social, organizational, and institutional contexts [4, 9, 12]. However, this sociotechnical turn is marked by a troubling paradox: while the use of the term "sociotechnical AI" proliferates, its interpretations and applications have become increasingly fragmented and diverse. Some scholars focus on the relationship between technology and actors [13], others examine the relationship between technology and values [16], while still others investigate the relationship between technology and institutions [32]. Additional perspectives use a sociotechnical lens to highlight the entanglements between technology, actors, and institutions, exploring how AI systems are shaped by and, in turn, shape organizational and institutional structures [17]. This conceptual plurality, though intellectually rich, creates significant challenges for developing coherent and effective governance frameworks.

Our research addresses this challenge through a comprehensive narrative literature review that examines how researchers and practitioners engage with sociotechnical AI terminology. Unlike systematic reviews that follow rigid structures to address narrowly focused questions, our narrative approach enables a broader interpretive synthesis that captures the nuanced ways sociotechnical concepts are deployed across disciplines. Following Baumeister's methodological framework [2], we prioritize conceptual interpretation and comprehensive overview to illuminate patterns of engagement with sociotechnical thinking.

This paper presents preliminary findings from our ongoing review, focusing specifically on the conceptual dimensions of sociotechnical AI. By examining how authors define, theorize, and engage with sociotechnical terminology, we provide critical insights into the current state of understanding in this rapidly evolving field. Our analysis reveals distinct modes of engagement with the sociotechnical paradigm that have significant implications for AI governance challenges, including anticipating high-priority risks, identifying appropriate governance focus areas and participants, designing effective interventions and tools, and evaluating their effectiveness in real-world contexts.

Through this work, we aim to contribute to a more coherent understanding of sociotechnical AI governance by clarifying its conceptual foundations, identifying areas of convergence and divergence, and highlighting both opportunities and challenges for meaningful implementation. The following sections detail our methodological approach and present our preliminary results, establishing a foundation for more integrated sociotechnical approaches to AI governance.

2 History of sociotechnical

"Sociotechnical" was coined by British sociologists from the Tavistock Institute for Human Relations, particularly Eric Trist and Fred Emery. Researchers at Tavistock were hired by owners of a coal mine to explain/repair how coal miners were rejecting new mining technologies through work stoppages and absenteeism despite the obviously improved working conditions and efficiency gains. The answer was that coal miners and their communities valued the social structure and personal esteem afforded by small-team mining practices, and had organized their society around it.

Tavistock researchers identified a need for an analytic concept between the social and the technical conceptual categories that were available to them: 'socio-technical systems theory'. In order to explain and repair the situation at the modernizing coal mines, they needed a middle category on which to operate.

"Sociotechnical" was subsequently adapted by multiple fields and theorists because of its analytic usefulness. In particular: sociotechnical analysis is used to explain the adoption and integration (or not), and the consequences of, a technology. In current times, we believe that the interpretation of sociotechnical AI has become fragmented [12]. What characterizes as sociotechnical account of AI should focus on accounting to power [11], empiricism [19], impacted communities by harms & hazards caused by AI [21], infrastructural aspects of AI system [8] and a more HCI focus approach. This characterization can help us in the critical reflection on how to best frame the ongoing developments around the concept of sociotechnical in the context of AI governance.

3 Methodology

To investigate how authors engage with the terminology of sociotechnical AI, we conducted a narrative literature review following established methodological guidelines for interpretive synthesis [14]. This approach was selected to capture the nuanced ways in which the concept of sociotechnical AI is understood and operationalized across disciplines, allowing us to identify patterns, convergences, and divergences in conceptualization.

The literature search was conducted in October 2024 across three major academic databases: Scopus, IEEE Xplore, and the ACM Digital Library. These databases were selected to ensure comprehensive coverage of both technical and social science literature on AI. We developed search strings tailored to each database's syntax requirements while maintaining consistent search parameters. Our search string is available in Appendix A.

Our search string conducted on October 15, 2024 resulted in 1,525 articles. No date restrictions were applied to capture the evolution of the sociotechnical concept in AI literature over time.
Articles are being screened in a two-stage process. Initially, titles and abstracts were reviewed to determine relevance according to predetermined inclusion and exclusion criteria. For articles passing this initial screening, full texts are being examined to confirm eligibility and assess their engagement with sociotechnical AI concepts. Our inclusion and exclusion criteria are included in Appendix B.
Our analysis focuses on three primary dimensions of how authors engage with sociotechnical AI:
a) Conceptual dimension: How do authors conceptualize sociotechnical AI and/or engage with existing sociotechnical concepts and theories?
b) Scope dimension: What elements do authors include in their sociotechnical analysis (artifacts, social factors, actors, institutions, norms, values, culture), and the boundary levels they establish (individual, organizational, societal)?
c) Methodological dimension: How do authors operationalize sociotechnical concepts in their research design, including methods of data collection and analysis, and the instrumental use of sociotechnical terminology?

For each included article, we are developing analytical memos addressing these dimensions, which will then be iteratively refined through team discussions to identify patterns and themes. The research team is dividing the complete set of articles into batches of reviewing 100 articles in two weeks. All the researchers are using the Rayyan web app [25] to organize their decisions. When there are discrepancies between their decisions, all the researchers involve the senior researchers in discussing them.

4 Preliminary Results: Initial Insights

In this section, we will present our current interpretations of the conceptual framing of sociotechnical AI in the literature, highlighting key patterns and variations in how authors approach this terminology.

4.1 Conceptual Dimension: Modes of Engagement

4.1.1 Instrumental use of sociotechnical AI. The instrumental approach to sociotechnical AI represents a utilitarian adoption of sociotechnical language and concepts, often without substantive engagement with theoretical underpinnings. This approach is characterized by sociotechnical terminology primarily used to enhance the marketability, legitimacy, or compliance status of technical solutions. In this paradigm, sociotechnical considerations serve as a form of "technical debt insurance" – a minimum viable acknowledgement of social factors that might otherwise impede technological implementation. The sociotechnical lens is wielded selectively, focusing on aspects that can be readily operationalized without fundamentally challenging technical paradigms. For instance, bias mitigation might be addressed through technical fixes without examining the underlying social structures that generate biased data in the first place [34].

This instrumental use typically manifests in several ways:

(1) First, through the "sociotechnical checklist" phenomenon, which often append social considerations as an afterthought rather than integrating them throughout the development process. These checklists often reduce complex sociotechnical entanglements to discrete, manageable items that can be "solved" without disrupting technical workflows [26].
(2) Second, in the "ethical white washing" of AI systems[1], where sociotechnical language is deployed rhetorically to signal ethical awareness while substantive engagement remains shallow. This is frequently observed in corporate AI ethics principles that acknowledge sociotechnical concerns without meaningful mechanisms for implementation.
(3) Third, in regulatory compliance strategies that adopt sociotechnical frameworks instrumentally to satisfy emerging AI governance requirements. The European Union's AI Act and similar regulatory frameworks have inadvertently incentivized this form of superficial sociotechnical engagement [5].

The instrumental approach, while not inherently problematic, risks reducing sociotechnical thinking to a veneer that legitimizes rather than transforms existing technical paradigms. It reflects what might be called "sociotechnical naïveté" – a genuine but limited understanding of sociotechnical complexity that fails to recognize how deeply social and technical factors are already entangled [24].

4.1.2 Implementation-oriented approach to Sociotechnical AI. The implementation-oriented approach to sociotechnical AI represents a *rather* genuine attempt to operationalize sociotechnical thinking in ways that meaningfully reshape AI development and deployment. Unlike the instrumental approach, it moves beyond rhetoric to experimental implementation, and unlike the critical approach, it focuses on constructive intervention rather than critique alone [30]. This mode of engagement is also characterized by methodological innovation that attempts to bridge theoretical sociotechnical insights with practical development processes. Researchers and practitioners in this space recognize the limitations of both purely technical and purely social approaches, seeking instead to develop frameworks that account for their mutual constitution [6].

The implementation-oriented approach manifests in several emerging practices such as participatory AI design, contextual auditing, and sociotechnical documentation [20, 28]. However, this approach faces significant challenges, including the difficulty of translating abstract sociotechnical theories into practical methodologies, the tension between standardization and contextual sensitivity, and resistance from established organizational practices. Nevertheless, it represents a pragmatic middle ground that acknowledges the limitations of both uncritical technical solutionism and purely theoretical critique.

4.1.3 Critical approach to Sociotechnical AI. The critical approach to sociotechnical AI deploys sociotechnical frameworks primarily as analytical tools to interrogate, problematize, and challenge dominant technical paradigms in AI development. Drawing from disciplines such as Science and Technology Studies (STS), critical theory, and system safety engineering [18], this approach moves beyond identifying technical limitations to examining the social, political, and economic structures that shape technological development.

Critical sociotechnical scholarship is characterized by several distinct analytical moves:

(1) First, it problematizes the assumed separation between "social" and "technical" domains, arguing that this dichotomy itself reflects and reinforces particular power relations [3, 22].
(2) Second, it highlights epistemological and ontological blindspots in conventional AI research, questioning whose knowledge counts, whose realities are represented in data, and whose interests are served by particular technical configurations [1, 15, 29, 32, 35].
(3) Finally, it examines the broader socioeconomic and political contexts in which AI systems are developed and deployed, attending to issues of labor exploitation, resource extraction, and environmental impact that are typically excluded from narrower sociotechnical analyses [8, 9, 23].

The critical approach acts as a necessary counterbalance to both instrumental appropriation and experimentation with sociotechnical concepts. However, this scholarship also faces challenges such as the risk of theoretical hermeticism[2] that limits practical impact, potential disconnection from technical implementation details, and the difficulty of translating critique into actionable alternatives. At its most effective, the critical approach does not merely criticize but opens up new conceptual spaces for reimagining sociotechnical relations in AI.

5 Way Forward

The sociotechnical turn in AI research represents perhaps the most significant conceptual shift in contemporary technological discourse, yet our analysis reveals a troubling landscape where this terminology has become simultaneously ubiquitous and hollow. Our ongoing research aims to address these challenges by systematically analyzing the diverse conceptualizations of sociotechnical

[1] https://www.linkedin.com/pulse/safety-washing-ai-summit-roel-dobbe-gy4oe

[2] https://blogs.uoregon.edu/rel399f14drreis/hermeticism/

AI to identify best practices, methodological innovations, and conceptual clarity that can inform more robust governance approaches. Through this comprehensive review, we intend to develop an actionable research agenda that bridges the gap between critical theoretical insights and practical implementation strategies. By synthesizing approaches that meaningfully operationalize sociotechnical thinking alongside critical perspectives that interrogate power dynamics and institutional structures, we aim to establish more coherent frameworks for sociotechnical AI governance.

References

[1] Chelsea Barabas, Colin Doyle, JB Rubinovitz, and Karthik Dinakar. 2020. Studying up: reorienting the study of algorithmic fairness around issues of power. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 167–176.

[2] Roy F Baumeister and Mark R Leary. 1997. Writing narrative literature reviews. *Review of general psychology* 1, 3 (1997), 311–320.

[3] Jonas Aaron Carstens and Dennis Friess. 2024. AI Within Online Discussions: Rational, Civil, Privileged? Ethical Considerations on the Interference of AI in Online Discourse. *Minds and Machines* 34, 2 (2024), 10.

[4] Brian Chen and Jacob Metcalf. 2024. Explainer: A Sociotechnical Approach to AI Policy. https://datasociety.net/library/a-sociotechnical-approach-to-ai-policy/

[5] Brian J Chen and Jacob Metcalf. 2024. Explainer: A sociotechnical approach to AI policy. *Data & Society* (2024).

[6] Yiqun T Chen, Angela DR Smith, Katharina Reinecke, and Alexandra To. 2023. Why, when, and from whom: considerations for collecting and reporting race and ethnicity data in HCI. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[7] Allan Dafoe. 2018. AI governance: a research agenda. *Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK* 1442 (2018), 1443.

[8] Eva de Winkel, Zofia Lukszo, Mark Neerincx, and Roel Dobbe. 2025. Adapting to limited grid capacity: Perceptions of injustice emerging from grid congestion in the Netherlands. *Energy Research & Social Science* 122 (2025), 103962.

[9] Sarah Dean, Thomas Krendl Gilbert, Nathan Lambert, and Tom Zick. 2021. Axes for Sociotechnical Inquiry in AI Research. *IEEE Transactions on Technology and Society* 2, 2 (2021), 62–70. https://doi.org/10.1109/TTS.2021.3074097

[10] J Delfos, AMG Zuiderwijk, S van Cranenburgh, CG Chorus, and RIJ Dobbe. 2024. Integral system safety for machine learning in the public sector: An empirical account. *Government Information Quarterly* 41, 3 (2024), 101963.

[11] Roel Dobbe, Thomas Krendl Gilbert, and Yonatan Mintz. 2021. Hard choices in artificial intelligence. *Artificial Intelligence* 300 (2021), 103555.

[12] Roel Dobbe and Anouk Wolters. 2024. Toward Sociotechnical AI: Mapping Vulnerabilities for Machine Learning in Context. *Minds and Machines* 34, 12 (2024). https://doi.org/10.1007/s11023-024-09668-y

[13] Upol Ehsan, Koustuv Saha, Munmun De Choudhury, and Mark O. Riedl. 2023. Charting the Sociotechnical Gap in Explainable AI: A Framework to Address the Gap in XAI. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 34 (April 2023), 32 pages. https://doi.org/10.1145/3579467

[14] Rossella Ferrari. 2015. Writing narrative style literature reviews. *Medical writing* 24, 4 (2015), 230–235.

[15] Deborah G Johnson and Mario Verdicchio. 2024. The sociotechnical entanglement of AI and values. *AI & SOCIETY* (2024), 1–10.

[16] Deborah G. Johnson and Mario Verdicchio. 2025. The sociotechnical entanglement of AI and values. *AI & Society* 40 (2025), 67–76. https://doi.org/10.1007/s00146-023-01852-5

[17] Olya Kudina and Ibo van de Poel. 2024. A sociotechnical system perspective on AI. *Minds and Machines* 34, 3 (2024), 21. https://doi.org/10.1007/s11023-024-09680-2

[18] Nancy G Leveson. 2016. *Engineering a safer world: Systems thinking applied to safety*. The MIT Press.

[19] Siddharth Mehrotra, Catholijn M Jonker, and Myrthe L Tielman. 2021. More similar values, more trust?-the effect of value similarity on trust in human-agent interaction. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 777–783.

[20] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*. 220–229.

[21] Aparna Moitra, Dennis Wagenaar, Manveer Kalirai, Syed Ishtiaque Ahmed, and Robert Soden. 2022. AI and disaster risk: a practitioner perspective. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–20.

[22] Merel Noorman and Tsjalling Swierstra. 2023. Democratizing AI from a sociotechnical perspective. *Minds and Machines* 33, 4 (2023), 563–586.

[23] Sem Nouws and Roel Dobbe. 2024. The Rule of Law for Artificial Intelligence in Public Administration: A System Safety Perspective. In *Digital Governance: Confronting the Challenges Posed by Artificial Intelligence*. Springer, 183–208.

[24] Vicki L. O'Day, Daniel G. Bobrow, and Mark Shirley. 1996. The social-technical design circle. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work* (Boston, Massachusetts, USA) *(CSCW '96)*. Association for Computing Machinery, New York, NY, USA, 160–169. https://doi.org/10.1145/240080.240246

[25] Mourad Ouzzani, Hossam Hammady, Zbys Fedorowicz, and Ahmed Elmagarmid. 2016. Rayyan—a web and mobile app for systematic reviews. *Systematic reviews* 5 (2016), 1–10.

[26] Ayomide Owoyemi, Joanne Osuchukwu, Megan E Salwei, and Andrew Boyd. 2025. Checklist Approach to Developing and Implementing AI in Clinical Settings: Instrument Development Study. *JMIRx Med* 6 (20 Feb 2025), e65565. https://doi.org/10.2196/65565

[27] Erich Prem. 2023. From ethical AI frameworks to tools: a review of approaches. *AI and Ethics* 3, 3 (2023), 699–716.

[28] Mahima Pushkarna, Andrew Zaldivar, and Oddur Kjartansson. 2022. Data cards: Purposeful and transparent dataset documentation for responsible ai. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1776–1826.

[29] Tapani Rinta-Kahila, Ida Someh, Nicole Gillespie, Marta Indulska, and Shirley Gregor. 2022. Algorithmic decision-making and system destructiveness: A case of automatic debt recovery. *European Journal of Information Systems* 31, 3 (2022), 313–338.

[30] Harini Suresh, Emily Tseng, Meg Young, Mary Gray, Emma Pierson, and Karen Levy. 2024. Participation in the age of foundation models. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1609–1621.

[31] Araz Taeihagh. 2021. Governance of artificial intelligence. *Policy and society* 40, 2 (2021), 137–157.

[32] Ibo Van de Poel. 2020. Embedding values in artificial intelligence (AI) systems. *Minds and machines* 30, 3 (2020), 385–409.

[33] Yoshija Walter. 2024. Managing the race to the moon: Global policy and governance in artificial intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences. *Discover Artificial Intelligence* 4, 1 (2024), 14.

[34] Sue Whetton and Andrew Georgiou. 2010. Conceptual challenges for advancing the socio-technical underpinnings of health informatics. *The open medical informatics journal* 4 (2010), 221.

[35] Maranke Wieringa. 2023. "Hey SyRI, tell me about algorithmic accountability": Lessons from a landmark case. *Data & Policy* 5 (2023), e2.

A Search String

(1) **Scopus:** (TITLE-ABS-KEY (sociotechnical OR socio-technical) AND TITLE-ABS-KEY (ai OR "artificial intelligence"))
(2) **IEEE Xplore:** (("All Metadata":sociotechnical OR "All Metadata":socio-technical) AND ("All Metadata":AI OR "All Metadata":"artificial intelligence"))
(3) **ACM Digital Library:** [[Abstract: sociotechnical] OR [Abstract: socio-technical]] AND [[Abstract: ai] OR [Abstract: "artificial intelligence"]]

B Selection Criteria

Our inclusion criteria is as follows:

(1) English language peer-reviewed publications
(2) Explicit engagement with both sociotechnical concepts and artificial intelligence
(3) Sociotechnical terminology appears in the article's purpose, methods, or stated contributions
(4) AI is a central topic or component of the sociotechnical system being discussed

Our exclusion criteria is follows:

(1) Publications where sociotechnical terminology appears only as a reference without substantive engagement
(2) Articles where sociotechnical concepts are used only as descriptive adjectives without conceptual development
(3) Studies where AI is peripheral rather than central to the discussion

(4) Non-peer-reviewed literature including conference abstracts, editorials, and book reviews