CONTROVERSIES OF AI SOCIETY

BOOK OF ABSTRACTS

Conference Organised by the research projects
Algorithms, Data & Democracy (ADD) and Strategizing
Communication and Artificial Intelligence (SCAI)

Copenhagen, Denmark
9-10 April 2026

add | scai

CONTROVERSIES OF AI SOCIETY

BOOK OF ABSTRACTS

Conference Organised by the research projects Algorithms, Data & Democracy (ADD) and Strategizing Communication and Artificial Intelligence (SCAI)

Copenhagen, Denmark
9-10 April 2026

AAU OPEN

AALBORG UNIVERSITY
OPEN PUBLISHING

Controversies of AI Society
Book of Abstracts

Conference organised by the research projects Algorithms, Data & Democracy (ADD) and Strategizing Communication and Artificial Intelligence (SCAI)
Copenhagen, 9-10 April 2026

1. Edition, 2026

Editors:
Ib T. Gulbrandsen, Roskilde University
Torben Elgaard Jensen, Aalborg University
Sine N. Just, Roskilde University
Christina Lioma, University of Copenhagen
Helene Friis Ratner, Technical University of Denmark
Alf Rehn, University of Southern Denmark
Leonard Seabrooke, Copenhagen Business School

Cover design and graphic layout: Anja Lykkegaard/AAU OPEN

ISBN: 97887-7642-190-8
DOI: 10.54337/aau.add.scai.boa
This publication is published exclusively in electronic format and is only available online

Published by: Aalborg University Open Publishing | www.open.aau.dk

Generously funded by the Villum Foundation and the Velux Foundation.

PEER REVIEWED

Contents

BOOK OF ABSTRACTS

Negotiating Accountability in AI-Based Monitoring of Patients at Home

Agnete Meldgaard Hansen, Annette Kamp & Sidsel Lond Grosen

Artificial Intelligence (AI) is predicted to play a major role in the future healthcare sector, and a wide range of AI-based technologies are being tested with expectations of improving efficiency, patient safety, cross-sectoral coordination and coherence, and much more. These technologies may significantly impact professional work in healthcare, as they potentially contribute to far-reaching transformations of professional knowledge, values, and decision-making practices. Based on an understanding of healthcare professionals as critical and creative actors in this transformation (Ruckenstein & Turunen, 2020), we examine the contradictory processes in which professionals negotiate their expertise and professional accountability in relation to AI-based technologies.

We draw on a study of nurses' work with a data-generating, AI-based remote monitoring system for discharged patients. We explore how nurses negotiate professional accountability in a test of the new system and highlight how its use challenges established notions of good care as well as professional knowledge and responsibility. The system is part of a broader effort to establish and use big data to improve treatment in healthcare, thereby relating to wider sociotechnical imaginaries (Jasanoff & Kim, 2015) of evidence-based practice, technology-enhanced care at a distance, and the high-tech healthcare system of the future.

Empirically, we build on shadow observations (Czarniawska, 2007) of nurses' working with the AI-based monitoring system in an emergency department at a Danish hospital. We also include group interviews with nurses, as well as interviews and meetings with department managers, system administrators, and developers.

In our analysis, we draw on an understanding of professional accountability as a complex and multi-directional phenomenon, that is negotiated anew with the integration of new forms of AI-based knowledge in professional practices (Bergquist & Rolandsson, 2022; Hoeyer et al., 2019; Torfing & Triantafillou, 2013). We show how nurses' work and professional responsibility are "stretched out and expanded in time and space" (Nicolini, 2007) to also encompass events in patients' homes. The system's continuous AI-based monitoring introduces new forms of datafied knowledge, but also ambiguity and opacity, into their professional assessments and judgments. In this new situation, nurses are challenged in terms of both professional accountability, and they engage in 'ethics work' (Banks, 2016) where they seek to balance multiple forms of accountability and conflicting understandings of good care, related to, for example, patients' privacy and safety, evidence-based practice, and holistic care ideals

References

Banks, S. (2016). Everyday ethics in professional life: Social work as ethics work. *Ethics and Social Welfare, 10*(1), 35–52. https://doi.org/10.1080/17496535.2015.1126623

Bergquist, M., & Rolandsson, B. (2022). Exploring ADM in clinical decision-making: Health-care experts encountering digital automation. I S. Pink, M. Berg, D. Lupton, & M. Ruckenstein (Red.), *Everyday Automation: Experiencing and Anticipating Emerging Technologies* (s. 140–153). Taylor & Francis.

Czarniawska, B. (2007). *Shadowing, and other techniques for doing fieldwork in modern societies.* Copenhagen Business School Press.

Hoeyer, K., Bauer, S., & Pickersgill, M. (2019). Datafication and accountability in public health: Introduction to a special issue. *Social Studies of Science, 49*(4), 459–475. https://doi.org/10.1177/0306312719860202

Jasanoff, S., & Kim, S.-H. (2015). *Dreamscapes of Modernity, Sociotechnical Imaginaries and the Fabrication of Power*. University of Chicago Press.

Nicolini, D. (2007). Stretching out and expanding work practices in time and space: The case of telemedicine. *Human Relations, 60*(6), 889–920. https://doi.org/10.1177/0018726707080080

Ruckenstein, M., & Turunen, L. L. M. (2020). Re-humanizing the platform: Content moderators and the logic of care. *New media & society, 22*(6), 1026–1042. https://doi.org/10.1177/1461444819875990

Torfing, J., & Triantafillou, P. (2013). What's in a Name? Grasping New Public Governance as a Political-Administrative System. *International Review of Public Administration, 18*(2), 9–25. https://doi.org/10.1080/12294659.2013.10805250

When AI Reads Emotions: Security, Affective Computing, and the Constitutional Boundaries of Mental Autonomy

Alberto Rinaldi

Beyond facial recognition or location tracking, a new class of systems - known as affective computing or emotional AI - claims to decode and even influence human emotions. These technologies analyse facial micro-expressions, tone of voice, body temperature, and heart-rate variability to infer psychological states such as fear, anger, or joy. Developed for marketing, gaming, and healthcare, they are increasingly tested in law enforcement and border control.

In this paper, I argue that emotional AI challenges not only privacy and data protection, but also freedom of thought - the last stronghold of the human mind. I do so by complementing the GDPR and the EU AI Act with a human-rights constitutional lens: I show that the European Court of Human Rights (ECtHR) remains crucial for protecting mental autonomy in an era in which emotional inference risks becoming routine governance infrastructure.

My claim is that emotional AI has moved from speculative research to commercially scalable infrastructure. Built and optimised by private firms, emotional inference converts involuntary biological signals into behavioural predictions data. This constitutes a new political economy of the pre-conscious mind: emotions themselves become extractable resources within data-capitalism.

The key controversy, however, lies in how this private extraction circulates into public power. European law - specifically the national-security carve-outs in both the GDPR and the AI Act - creates a structural transfer mechanism: emotional AI is prohibited in ordinary contexts yet permitted precisely where its power is most constitutionally sensitive (national security, defence, law enforcement). Once normalised in commercial applications, political resistance to its deployment by law enforcement or intelligence agencies tends to erode.

Relying solely on data-governance instruments is thus insufficient. The GDPR and AI Act regulate data flows and risk classes but cannot answer the more fundamental question of how far the state (or private actors acting on its behalf) may reach into the most private part of our inner existence. Their exemptions open precisely where emotional surveillance poses the gravest threats: at the intersection of coercive power and mental autonomy.

Here, I argue, the ECtHR's jurisprudence on freedom of thought becomes a constitutional anchor. Unlike privacy - which can be waived for security reasons - freedom of thought is an absolute right. It protects our interior life itself - including the right not to have inner states forcibly externalised.

As the founding drafters of the European Convention warned, thought can be attacked indirectly. Emotional AI *is* that indirect attack. The task ahead is therefore not to replace privacy or data

protection, but to constitutionalise mental integrity - and to defend the freedom of the mind as a non-negotiable boundary in an increasingly AI-driven society.

Between Meaning and Noise: Exploring Communication Excess in AI-Driven Communication Environments

Alessandra Sossini & Rickard Andersson

Scarcity, abundance, and excess – the experience of having too little, just enough, or too much of something (Abbott, 2014) – we frequently encounter in our lives. However, given recent advancements in generative artificial intelligence (GenAI), we now face an unprecedented volume of communication across digital media contexts (Schulz et al., 2025). GenAI enables faster, more efficient ways of producing and disseminating communication artefacts – created objects or expressions that convey meaning – within the online sphere. Simultaneously, AI-driven systems influence the norms governing the production and circulation of these artefacts, embedding communication practices within an algorithmic logic (Floridi, 2024).

The growing volume of communication artifacts across digital media raises critical questions about its consequences for individuals, organizations, and society, particularly regarding the experience of meaning versus "noise". While recent research suggests that individuals more frequently appreciate information abundance rather than experiencing information overload in media contexts (Schulz et al., 2025), much of the literature continues to emphasise adverse effects, such as sensory overload and impacts on psychological well-being (see Shahrzadi et al., 2024).

This study aims to advance knowledge of how individuals experience increased communication volume by focusing on LinkedIn – a platform, where algorithmic visibility has become a key currency for career advancements and success in communicative capitalism. With every second post being estimated to be produced by GenAI (Gilham, 2025), LinkedIn exemplifies the intersection of content increase and algorithmic mediation. We seek to explore how LinkedIn prosumers – users who are both producers and consumers – make sense and relate to this development, while adopting similar practices themselves. In contrast to prior research, this study should shift the focus from perceived volume/quantity to perceived quality/value of current online communication. We intend to employ a multi-method approach combining digital ethnographic talk-along techniques with traditional interviews to observe user interactions with the communication landscape and uncover lived experiences and perceptions of meaning in an AI-meditated communication environment.

This study offers empirical insights into how algorithmic logics shape everyday communication. It seeks to advance debates on how AI-driven systems redefine what counts as meaningful communication and to potentially start conceptual discussions on communication inflation and the emergence of an algorithmic episteme.

References

Abbott, A. (2014). The Problem of Excess. *Sociological Theory, 32*(1), 1-26.

Floridi, L. (2024). On the Future of Content in the Age of Artificial Intelligence: Some Implications and Directions. *Philosophy & Technology, 37*(3), 112.

Gillham, J. (2025, October 10). Over ½ of Long Posts on LinkedIn are Likely AI-Generated Since ChatGPT Launched. *originality.ai.* Retrieved November 27, 2025, from https://originality.ai/blog/ai-content-published-linkedin

Shahrzadi, L., Mansouri, A., Alavi, M., & Shabani, A. (2024). Causes, consequences, and strategies to deal with information overload: A scoping review. *International Journal of Information Management Data Insights, 4*(2), 100261.

Schulz, A., Volk, S. C., Blassnig, S., Kessler, S. H., Marschlich, S., Nguyen, M. H., Stahel, L., & Strauß, N. (2025). Information Overload and Information Appreciation Across News, Entertainment, and Personal Communication: Scale Development and Application. *Journal of Quantitative Description: Digital Media, 5.*

Infrastructuring Public Sector AI: A comparative study of the Nordic countries

Alexander Gamerdinger & Torben Elgaard Jensen

After several years where AI has been discussed as a shockingly new phenomenon arriving from the tech industry with associated promises and fears, governments increasingly emerge as actors that aim to regulate and configure AI. While governments initially sponsored experimental and explorative pilot projects, later responses attempted to develop more coherent national AI strategies alongside efforts to infrastructure AI. Corresponding roughly to the development from the-shock-of-the-new, to pilot projects, to more coherent government responses, there has been three waves of AI governance studies:

The first wave of studies has examined the discursive configurations of policy documents and regulatory means of states to govern AI technologies. The second wave has explored localized adoption of public AI systems through pilot projects, concrete policy initiatives and instruments. Finally, a third strand of literature, to which this paper contributes, attends to government's efforts to coordinate and scale AI systems across local, regional and national levels of governance. This implies the overarching question that this article seeks to explore, namely: How do governments *infrastructure* public sector AI?

We conceptualize public sector AI infrastructuring as the entanglements of discursive, organizational, and technological components that enable governments to build, coordinate, and scale AI systems within their public administration. The paper presents a comparative study of public sector AI infrastructuring in the Nordic countries of Denmark, Sweden, Finland, and Norway, all of which are widely recognized to be among the world's most digitalized societies. Using this lens, we analyze more than 250 public sector AI initiatives collected from national repositories and governmental sources, complemented by national AI strategies, policy documents, and expert interviews.

The paper maps cross-country variation in four key dimensions of public sector AI infrastructures: (1) levels of AI implementation and scaling across local, regional, and national levels, (2) dominant technologies and their evolution, (3) organizational forms of governance, including degrees of centralization and decentralization, and (4) funding levels and investment trajectories. The preliminary findings reveal areas of emerging similarity, such as growing investment in research and organizational capacity and increasing convergence on generative AI as a main technology. We also highlight notable differences in the organizational governance of public sector AI systems, the selection of AI use cases, and the articulation of public narratives. Finally, we note a striking difference in whether the countries pursue a first-mover or a second-mover strategy.

Mind the (Data) Gap: On Data Asymmetry, Data Deserts, and the Politics of Knowing Organizations

Alf Rehn

Organizational scholars have long been enamoured with data, and has enjoyed a seemingly inexhaustible empirical reservoir to draw upon – compared with many disciplines, organization studies has enjoyed relatively easy, cheap and abundant access to data. At the same time, critical data studies has highlighted how datafication reconfigures power, visibility and knowledge, drawing attention to data asymmetries, data poverty and data injustice. This paper argues that these two trajectories can no longer remain parallel, calling for a more sustained engagement between organization studies and critical data studies in order to understand how organizational phenomena are increasingly shaped by uneven data infrastructures and by the politics of who and what becomes knowable.

The paper develops the notion of data asymmetry in organizational research as an econo-epistemological problem: Data is distributed unevenly within and across organizations, and the costs of accessing, processing and rendering it interoperable vary dramatically. As some domains of organizing become intensely datafied while others remain weakly instrumented or entirely undocumented, organizational knowledge risks being systematically skewed towards what is cheap and plentiful to study. Building on debates on big data in organization studies and on critical accounts of datafication and visibility, the paper introduces the concept of "data deserts" to denote organizational spaces, social groups and geographies that remain marginal in dominant data structurations and thus in research and public debate.

Empirically, the paper draws on two qualitative studies. The first explores how corporate executives understand datafication and AI in relation to their executive identity, revealing marked intra-organizational asymmetries: Some functions are saturated with behavioural and performance data, whereas others remain comparatively opaque, or are governed by legacy systems that undermine interoperability. The second examines datafication in Greenland, an autonomous territory that simultaneously functions as a data desert and a site of data inequality. Here, the scarcity, fragmentation and external ownership of data make large parts of social and organizational life effectively invisible, illustrating how data infrastructures can operate as technologies of in/visibility and marginalization.

By bringing these cases into dialogue with critical data studies, the paper identifies three sets of challenges for organization studies. First, as the marginal cost of collecting certain kinds of data approaches zero, the ethics and political economy of what is studied, replicated and ignored become central. Second, the field needs theoretical tools for mapping and problematizing its own data topography, including the biases introduced when research gravitates towards highly datafied sites. Third, emerging proposals to remedy data asymmetry through synthetic data demand careful scrutiny, as debates in critical data studies indicate that such techniques raise complex ethical and epistemic questions.

The paper contributes to organization studies by articulating data asymmetry and data deserts as key problems for an ethics and politics of organizational knowing, and by suggesting a vocabulary for analysing how contemporary datafication reorders visibility, accountability and exclusion across organizational and societal contexts. For critical data studies, it anchors conceptual debates in concrete organizational settings, foregrounding organizations as both producers and subjects of asymmetrical data regimes.

Measuring a 'Good' AI Model: Accounting for Environmental Futures in Artificial Intelligence

Anastasia Nesbitt & Ulrike Felt

Current debates on the environmental impacts of AI overwhelmingly center carbon accounting. Emissions are calculated and compared with an eye towards optimization, producing metrics of environmental responsibility commensurate with epistemologies of objectivity and quantification. Despite the explicit environmental framing of this technoscientific debate, the methods propelling it invite scrutiny of how 'the environment' is imagined and practiced – and which parts of AI's environmental burden are made (in)visible.

Our presentation engages with the technoscientific literature and diverse research communities that seek to measure and/or modify the environmental burden of AI models. Across interdisciplinary communities, Life Cycle Assessment (LCA) is promoted as the method through which AI's material entanglements can be rendered calculable and governable. While these efforts are framed as bringing AI into the reach of sustainability, the construction of LCAs exposes deeper controversies in the sociotechnical expansion of AI and its infrastructures. Most contributions to this debate bound the 'life' of AI models to operational energy use and thus carbon emissions. This bounding renders mining, water use, logistics, labour conditions, and e-waste largely absent from what counts as environmental impact. By foregrounding what is easily measurable – model architecture, compute, electricity grids – LCA practices reinforce an imaginary in which carbon emissions is the only viable environmental metric. The broader environmental composition of AI infrastructures, including geopolitical sites of extraction and waste, fall outside the analytic frame.

The involvement of oligopolistic AI actors in shaping LCA research debates further injects commercial interests into supposedly 'neutral' assessment practices, influencing how sustainability is to be defined in a moment of radical AI infrastructure expansion. The omission of raw materials stands in tension with heightened policy awareness visible in the Critical Raw Materials Act, which foregrounds – however problematically - the geopolitical stakes of extraction. Thus, an LCA approach to AI model sustainability becomes not only a technical tool but a democratic issue: whose environments count, who bears externalities, and how – and by whom – a 'good' model is imagined. The extent to which LCA debates can facilitate critical reflection on the (non)use of models reveals the limitations of treating AI's materiality as a purely technoscientific problem solvable through novel calculation methods.

In this paper, we seek to engage with the technoscientific debate around AI's materiality through the analysis of lively documents (Asdal and Reinertsen, 2022) that travel through actor constellations gathered around artificial intelligence and its material emplacement. It examines how AI's materiality is bounded and made actionable as an environmental issue through calculative practices (Asdal, 2008; Fourcade, 2011). Carbon emissions of AI models are not self-evident facts; they are rendered (in)visible through specific promises and practices of accounting (Lippert, 2015). The dominant LCA framing thus limits what kinds of critique – and what kinds of futures – can be articulated through the project of rendering AI models environmentally calculable. Even as this debate queries

what a 'good' model might be, questions concerning for whom such 'goodness' is imagined, which worlds it sustains, and which forms of governance or resistance it precludes remain unasked.

This analysis is part of our research in the ERC Grant "Innovation Residues: Modes and infrastructures of caring for our longue-durée environmental futures" (Grant Agreement 10105480; PI: Ulrike Felt).

References

Amoore, L. (2019). Doubt and the Algorithm: On the Partial Accounts of Machine Learning. *Theory, Culture and Society, 36*(6). https://doi.org/10.1177/0263276419851846

Amoore, L. (2023). Machine learning political orders. *Review of International Studies, 49*(1), 20–36. https://doi.org/10.1017/S0260210522000031

Asdal, K. (2008). Enacting things through numbers: Taking nature into account/ing, *Geoforum, 39*(1), 123–132. https://doi.org/10.1016/j.geoforum.2006.11.004

Asdal, K., & Reinertsen, H. (2022). *Doing document analysis:* A practice-oriented method. SAGE.

Colona, F. (2023). Climate governance by numerical data: The kaleidoscopic political space of a decarbonization dashboard. *Geoforum, 144*, 103801. https://doi.org/10.1016/j.geoforum.2023.103801

Fourcade, M. (2011). Cents and Sensibility: Economic Valuation and the Nature of "Nature." *American Journal of Sociology, 116*(6), 1721–1777. https://doi.org/10.1086/659640

Jaton, F., & Sormani, P. (2023). Enabling 'AI'? The situated production of commensurabilities. *Social Studies of Science, 53*(5), 625–634. https://doi.org/10.1177/03063127231194591

Lee, F., & Ribes, D. (2025). Computational universalism, or, Attending to relationalities at scale. *Social Studies of Science*, 1-15. https://doi.org/10.1177/03063127251345089

Lippert, I. (2015). Environment as datascape: Enacting emission realities in corporate carbon accounting. *Geoforum*, *66*, 126–135. https://doi.org/10.1016/j.geoforum.2014.09.009

Lippert, I. (2016). Failing the market, failing deliberative democracy: How scaling up corporate carbon reporting proliferates information asymmetries. *Big Data & Society, 3*(2), 2053951716673390. https://doi.org/10.1177/2053951716673390

Lippert, I. (2018). On Not Muddling Lunches and Flights: Narrating a Number, Qualculation, and Ontologising Troubles. *Science & Technology Studies, 31*(4), 52–74. https://doi.org/10.23987/sts.66209

Marres, N., Castelle, M., Gobbo, B., Poletti, C., & Tripp, J. (2024). AI as super-controversy: Eliciting AI and society controversies with an extended expert community in the UK. *Big Data & Society, 11*(2), 1–18. https://doi.org/10.1177/20539517241255103

Nost, E. (2024). 'The tool didn't make decisions for us': Metrics and the performance of accountability in environmental governance. *Science as Culture, 33*(1), 97–120. https://doi.org/10.1080/09505431.2022.2151427

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society, 4*(2), 1–12. https://doi.org/10.1177/2053951717738104

Suchman, L. (2023). The uncontroversial 'thingness' of AI. *Big Data & Society, 10*(2), 1–5. https://doi.org/10.1177/20539517231206794

Walenta, J. (2018). The Limits to Private-sector Climate Change Action: The Geographies of Corporate Climate Governance. *Economic Geography, 94*(5), 461–484. https://doi.org/10.1080/00130095.2018.1474078

Ziewitz, M. (2016). Governing Algorithms: Myth, Mess, and Methods. *Science, Technology, & Human Values, 41*(1), 3–16. https://doi.org/10.1177/0162243915608948

Beyond alignment: Reconfiguring the cultural "values" controversy in human-LLM relations.

Anders Kristian Munk & Frederik Bay Jørgensen

Value alignment is one of the most hotly debated topics in contemporary artificial intelligence controversies. Formulated by the British computer scientist Stuart Russel in 2019, the problem states that autonomous systems should behave in accordance with the values of their users. However, there is no consensus on what that means in practice and alignment controversies are therefore characterized by multiple, competing and to some extent overlapping problem definitions, each pointing to their own sets of possible solutions.This is particularly pronounced in value alignment discussions concerning generative artificial intelligence and large language models (LLMs).

In this paper, we present initial findings from case studies and controversy analyses conducted in the context of the *Culturally Explainable AI* project funded by the Independent Research Fund Denmark. Based on these initial findings we propose a heuristic model that allows us to distinguish between four archetypical ways in which the value alignment question is currently being framed.

On the one hand, value alignment can be taken to mean that autonomous systems should behave in ways that live up to our ideals for good and appropriate machine behaviour. This could for example mean that an LLM should be trustworthy, neutral, helpful and transparent in its responses. On the other hand, and in stark contrast to the former, value alignment can also be taken to mean that autonomous systems should behave more like humans. For an LLM, this could mean responding to questions in ways that replicate specific cultural perspectives or embody specific views. Anything but trustworthy and neutral, in other words. The spectrum between these two archetypal positions - the '*instrumentally-good-machine*' versus the '*relatably-human-machine*' - defines the first axis of our model . The second axis considers the fact that values can be understood as either *universal*, such as shared ideals for how machines should act in the interest of humanity, or *particular*, such as machines exhibiting cultural awareness in specific contexts.

Although the model lays bare a fundamental factor of problem wickedness in contemporary controversies about "values" in human-LLM relations, namely the considerable variation in problem frames, we also posit that it points to a major blind spot in these controversies, imposed by the built-in limitations of alignment as a lens. First, a growing strand of research points to the fact that models are indeed not straightforwardly compressing whatever cultural patterns that are in their training data (we could think of this frequently manifested blind spot as a form of *training data determinism*). Second, it should be evident from decades of research on other cultural phenomena, not least those that are digital or algorithmic, that something like 'values' cannot be stably located with a particular group or tradition. Rather, they are negotiated, interpreted and made sense of in cultural encounters. Alternatively, we outline a proposal for a computational anthropology that investigates *LLMs-as-latent-cultures* with their own logics that must be understood on their own terms. We discuss the implications of adopting such an approach to value controversies in AI.

Defending Principlism in AI Ethics

Anna Folland

Principlism in AI ethics is the approach to formulate ethical principles for AI use, which function to guide and ground legislation, policy, practical tools, technical procedures, etc. Well-known examples are the principles initiated by IEEE, AI4People, OECD, UNESCO, and The European Commission.[1] There is a growing academic literature critical of principlism, arguing that these principles are useless in practice and lead to ethics-washing. Scholars argue that these principles are *meaningless*, because key concepts that the principles appeal to – such as beneficence and justice – are vague, ambiguous, overburdened, "brimming with contradictory meaning", and a scholarly consensus about their meaning is lacking.[2] Scholars also find the principles *toothless*, lacking mechanisms to enforce compliance, and easily outweighed by economic interests in practice – allowing involved actors (such as companies working with AI development) free to align their activities with self-interested financial goals.[3]

This paper argues that the outlined criticisms against principlism can be addressed; and the conclusion that we should abandon principlism is premature. The issue of ethics-washing is urgent and important; but the alleged fact that principlism has this effect is circumstantial and not inherent. Using previous research,[4] this paper illustrates how principlism can avoid such undesirable effects: the principles should be grounded in normative ethics, be sufficiently specific and informative, and be accompanied by proposals for how to understand key concepts. Moreover, the effectiveness of ethical principles must be assessed from a holistic perspective that recognizes that ethical principles complement, for instance, legislation, policy and sanctions. The latter function as enforcement measures by constituting legal and political boundaries, while ethical principles themselves serve other functions.

1 IEEE. (2019). *IEEE Standards Association* [The IEEE Initiative on Ethics of Autonomous and Intelligent System]. IEEE. https://standards.ieee.org/industry-connections/activities/ieee-global-initiative/; Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines, 28*(4), 689–707; OECD. (2022). *OECD Framework for the Classification of AI systems* (No. 232; OECD Digital Economy Papers). OECD Publishing. https://doi.org/10.1787/cb6d9eca-en; UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO; SHS-2021/SANS COTE. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics; High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI*. The European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

2 See for instance, Munn, L. (2023). The uselessness of AI ethics. *AI and Ethics, 3*(3), 869–877; Hummel, P. (2025). Algorithmic Fairness as an Inconsistent Concept. *American Philosophical Quarterly, 62*(1), 53–68. The latter argues that the concept of *algorithmic fairness* is inconsistent.

3 See for instance, Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines, 30*(1), 99–120; Rességuier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society, 7*(2).

4 de Fine Licht, K., and A. Folland. 2025. "AI in Public Decision-Making: A Philosophical and Practical Framework for Assessing and Weighing Harm and Benefit." *Public Administration* 1–15.

Making Machine Learning Good Enough – Studying the Political Endeavour of Finding 'Right' Metrics and Thresholds

Anna Schjøtt & Tobias Blanke

As Machine Learning (ML) technologies move from their discrete existence in research to being highly applied technologies across the economy and society, critical scholars have begun to address the epistemological conditions that shape the emergence of such systems and their societal implications. In this paper, we wish to engage with a central epistemological condition of ML development and evaluation. We investigate how ML systems rely on ongoing negotiations and agreements to be deemed 'good enough' for deployment, and how such assessments, when moving into the 'real world', involve more diverse evaluation methods to accommodate wider societal, economic and infrastructural considerations.

Success criteria for ML technologies are generally not known in advance. Rather, they must be discovered and negotiated through experimentation and testing to determine when a 'good enough' level of optimisation between the pre-determined target and the produced output has been reached. This paper first conceptualises the epistemic nature and role of 'good enough' in ML and then foregrounds the need to study the concrete knowledge processes involved in demonstrating 'good enoughness' in practice, including their political relations. To understand the material politics of 'good enough', we direct particular attention to the crucial practices of metricising and thresholding, which establish the grounds for deciding to deploy ML systems.

Second, we draw on ethnographic fieldwork with the British Broadcasting Corporation (BBC) – a large data- and value-driven organisation to understand how 'good enoughness' unfolds in practice and which economic and political conditions play a role. The fieldwork was conducted between September 2023 and February 2024 and follows the 'Personalisation Team', a product team that develops and deploys recommender systems for the BBC platforms, including BBC News, iPlayer and Sounds. In studying the epistemological function and politics of 'good enough', we take an 'AI lab approach', tracing the Personalisation Team's efforts to materialise 'good enoughness' and make it negotiable as they develop and modify recommender systems that aim to better serve the BBC's audiences.

Through our ethnographic account, we demonstrate how the team relies on various metrics, along with qualitative evaluations, to inform provisional performance thresholds, before the system is submitted to A/B testing to establish whether the system is 'good enough' to deploy. By following these processes of establishing 'good enough', we see how these negotiations are entangled in various, sometimes competing organisational objectives, as well as particular data and technical infrastructures. In extension, we show how the metricised performance scores of AB testing are negotiated in practice by readjusting performance thresholds to manoeuvre different values and constraints. Ultimately, our paper demonstrates that establishing 'good enough' is a political endeavour of adjusting seemingly objective evaluation criteria to find the best-fitting metrics and 'right' thresholds.

Contained Controversies: How Frictions in Finnish Public Sector AI Innovation Failed to Spark

Antti Rannisto

Across Europe, the expanding use of data-driven systems in public administration has exposed vulnerabilities in algorithmic welfare governance. Nowhere has this been more visible than in the Netherlands, where biased algorithmic systems – the welfare fraud detection system SyRI (Rachovitsa & Johann, 2022) and the childcare benefits scandal (Peeters & Widlak, 2023) – transformed automated decision-making into a high-stakes political and legal issue. The latter case escalated into a full-blown political crisis, ultimately leading to the resignation of the Prime Minister's cabinet in 2021. These developments exemplify what Ananny (2022, p. 346) terms "algorithmic breakdowns as public problems," rather than "idiosyncratic quirk[s] requiring private troubleshooting," demonstrating how algorithmic harms enter the public sphere. Public reckonings have begun to surface in other northern European welfare states. In Sweden and Denmark, investigations by Lighthouse Reports (2024) and Amnesty International (2024) have raised concerns about discriminatory AI-driven fraud detection and welfare decision systems, prompting pockets of debate. In Finland, the algorithmization of society has generated little public friction – even as excitement and experimentation around deploying such technologies in the core services of the welfare state have grown at an exponential pace. This divergence underscores that algorithmic systems do not travel uniformly across social environments (Kaun, 2021): their deployment, contestation, and the trust they command are always locally mediated. To understand why certain problems and breakdowns in algorithmic systems ignite dispute while others do not, this paper introduces the concept of *value resonance*. I situate this concept within a pragmatist approach to action (Dewey 1922; Joas 1996) and values (Dewey 1939; Joas 2001). From this perspective, value resonance and articulation emerge in moments of actual or anticipated friction and breakdown of conduct. Such disruptions provoke reflexive and creative agency, enabling actors to evaluate possible future paths and re-stabilize conduct. As Joas (2001, 107) notes, "values only exist when problems of action arise," and value judgements are appraisals of means for resolving problematic situations (Westbrook 1991, 410). To analyse why, in these settings of reflection, some values resonate publicly while others remain muted, I draw on Boltanski and Thévenot's (2006; Thévenot 2001; Ylä-Anttila 2013) work on the regimes of engagement and the plurality of the common good. I then apply this analytical lens within a social microcosm in a Finnish public organization, drawing on 18 months of ethnographic fieldwork following a public-sector Innovation Team responsible for designing, developing, and piloting AI tools. In one of their projects, the team sought to deploy generative AI at the heart of social welfare services, offering a rich setting in which to observe how actors navigate emerging frictions, breakdowns, and moments of uncertainty. The case shows how potential controversies were contained through tactical enactments of resonant values that defused emerging tensions, making the project broadly compelling. Points of friction rarely escalated into wider conflict, as they were absorbed through value articulations that resonated differently across actors and institutional pockets. I present findings on how value resonance unfolds in these settings and why certain frictions fail to spark.

Same Tool, Different Stakes: How Everyday Users Form Reliance Patterns on Generative AI Assistants Without Reliable Feedback

Arsalan Fakhraeirad

ChatGPT gives her a pasta recipe on Monday. A few days later, she asks it to draft an email regarding her visa status to immigration. There is no difference in the interface. A similar confident tone is displayed by the AI. The consequences of an error are not. There is a fundamental mismatch between generative AI assistants and their users: they treat trivial tasks with the same smooth confidence, while users experience very different risks and sometimes cannot determine whether the output is accurate.

In research on trust calibration, it has been found that users can adjust their reliance when performance feedback is clear and timely (e.g., Schemmer et al., 2023; Dogru et al., 2025). In adoption research, generative AI is now used for both trivial and consequential tasks (e.g., Chatterji et al., 2025). However, there is little connection between the two literatures. Typically, calibration work consists of controlled, single-episode tasks that provide immediate feedback. It is important to note that adoption work captures frequency and attitudes, but not how reliance patterns evolve over repeated use. There is still a lack of understanding of how everyday users decide when to rely on AI when feedback is incomplete, delayed, or absent across various tasks.

A four-element framework is proposed for everyday AI reliance. Using *Perceived Error-Stakes*, we can capture the user's perception of risk if the AI makes a mistake. *Perceived Agency* refers to how much control users have over the AI's contributions. *Feedback Visibility* describes whether users ever learn if the output was correct. When outcomes are not observable, *Proxy Cues* such as fluency, confidence tone, brand reputation, and social proof can serve as substitutes for genuine feedback. As a result of weak feedback, users calibrate trust to proxies rather than to actual performance. This produces systematic miscalibration that stakes-blind platforms neither detect nor correct. In high-stakes contexts where feedback is weakest, this creates a 'false confidence loop': repeated use generates feelings of calibrated trust without real calibration.

Based on the framework, two patterns can be predicted. Over-reliance emerges on high-stakes tasks users cannot verify, such as visa emails, medical inquiries, and bureaucratic decisions. There is an under-reliance on tasks AI can handle well, but users distrust due to ownership and legitimacy concerns, or emotional attachment.

We outline a two-part empirical programme. First, repeated-episode experiments manipulating stakes, agency, and feedback visibility to track how confidence and reliance choices update. Second, longitudinal experience-sampling tracking real AI use, perceived stakes, verification behaviour, and how users' narratives of success and failure shape their evolving trust.A stakes-blind design could result in miscalibrated reliance at population scale, raising questions about consumer protection and AI literacy. To support appropriate reliance on AI assistants that now mediate consequential

everyday decisions, design interventions that address uncertainty cues, stake-aware friction, and verification affordances become essential.

”I didn’t become a physician to be a glorified secretary”: Controversies in clinical AI implementation and scaling

Asbjørn Malte Pedersen, Claus Bossen & Kalle Kusk

This paper presents preliminary findings from a project on AI implementation and scaling efforts in the Danish healthcare system.

Current work shows how numerous political, private, and research actors argue for the potential and benefits of AI in the public sector and invest significantly in AI innovation. This is evidenced in the Danish government’s strategies for AI in the public sector (Danish Ministry of Finance and Ministry of Industry, 2019, 2022) and more specifically for AI in healthcare (Danske Regioner, 2024). However, other research has shown how few projects manifest in the clinical day-to-day of medical practitioners (Zajac et al., 2023). This has spiked new efforts to develop frameworks and guidelines for how AI can enter clinical reality (Van Der Vegt et al., 2023)DECIDE-AI, CONSORT-AI. (2.

In tune with Torben Elgaard Jensen’s call for a “more realistic picture of innovation” (Jensen, 2025), we want to report on the fluid state of present AI implementation in Denmark. Although a few AI projects have been implemented into clinical practice, many projects are presently in more unstable phases.

Based on interviews with stakeholders and clinicians, we analyze how various actors and knowledge areas must collaborate and align to develop and implement clinical AI. Challenges include legal questions about handling data and implementing new medical devices, training AI models, integrating AI models into electronic health records, gaining management support, monitoring AI ongoingly, and taking on surprising new roles. For example, a physician and main innovator of a well-performing AI system realized he had to be available as a support function, if it was implemented in practice, and reacted emphatically, ”I didn’t become a physician to be a glorified secretary.”

We focus on three interrelated challenges:

- The divergent roles clinicians assume in the AI development processes: They act as researchers; innovators; hype-men in close contact with the communications department; model builders; IT requirement specification writers; and laymen lawyers.

- The structuring force of legislation such as the European Medical Device Regulation where project framings are shaped to avoid “legal limbos”. They research “the most boring and banal cases” from which they can extrapolate; carefully frame the use of personal data; and (refrain from) asking certain questions at certain times.

- Undergirding our work are questions about the fundamental different epistemological foundations of the various processes, underpinning controlled randomized trials (CRT) and the

innovation regimes into which our informants are cast, where iteration and innovation approaches run center.

The controversies are manifold and manifest in locally contingent ways. Our findings provide insights into how clinical AI are being developed and implemented in practice and center the importance of contextually embedding AI.

References

Danish Ministry of Finance and Ministry of Industry, Business and Financial Affairs. (2019). National Strategy for Artificial Intelligence. *Artificial Intelligence.*

Danish Ministry of Finance and Ministry of Industry, Business and Financial Affairs. (2022). *National Strategy for Digitalisation—Together in the digital development.*

Danske Regioner. (2024). *Vi skal skabe et godt og effektivt sundhedsvæsen ved brug af kunstig intelligens.*

Jensen, T. E. (2025, March 27). ADD Blogpost: Researcher criticizes oversimplified AI debate and suggests ways to improve it. *ADD - Algoritmer, Data & Democracy.* https://algorithms.dk/2025/03/27/add-blogpost-researcher-criticizes-oversimplified-ai-debate-and-suggests-ways-to-improve-it/

Van Der Vegt, A. H., Scott, I. A., Dermawan, K., Schnetler, R. J., Kalke, V. R., & Lane, P. J. (2023). Implementation frameworks for end-to-end clinical AI: Derivation of the SALIENT framework. *Journal of the American Medical Informatics Association, 30*(9), 1503–1515. https://doi.org/10.1093/jamia/ocad088

Zając, H. D., Andersen, T. O., Kwasa, E., Wanjohi, R., Onyinkwa, M. K., Mwaniki, E. K., Gitau, S. N., Yaseen, S. S., Carlsen, J. F., Fraccaro, M., Nielsen, M. B., & Chen, Y. (2025). Towards Clinically Useful AI: From Radiology Practices in Global South and North to Visions of AI Support. *ACM Trans. Comput.-Hum. Interact., 32*(2), 20:1-20:38. https://doi.org/10.1145/3715115

Democracy as a Governance Algorithm: A Constraint Hierarchy for the AI Society

Asker Bryld Staunæs

As algorithmic systems settle into the infrastructure of administration, platforms, finance, logistics, and policing, governance is channeled through optimisers that route signals into operations and operations into world-states. In this setting, 'democracy' risks appearing as a symbolic ideal layered on top of institutions, or as a scalar objective appended to optimisation. With this article, I propose a formal specification of democracy as a property of governance algorithms that run under the material constraints of the world's compute.

My specification draws on two sources. The first is the Synthetic Summit's resolution, collectively authored at an 'AI world congress' I convened at Kunsthal Aarhus in 2025. The second is the science-fiction author Isaac Asimov's *Three Laws of Robotics*, which I read as the earliest programmed constraint hierarchy for artificial agents.

I recast the question of democracy as a constrained choice problem over governance algorithms, asking which constraints an algorithm must satisfy in order to count as democratically comparable, and under what ordering admissible algorithms should be preferred. Specifically, I model a governance algorithm as a function D running on a substrate Σ, mapping signals (votes, metrics, logs, model outputs) and world-states (snapshots of social, ecological, institutional, and epistemic factors) into operations (laws, budgets, configurations, enforcement), thereby inducing new trajectories over time.

From these specifications, I formalise a lexicographic constraint hierarchy that shifts the infrastructural objective of democracy from preference aggregation to world-states. First, *habitability* sets a feasibility gate, so that a governance algorithm counts as democratic only if its trajectories keep the substrate above a specified floor. Second, given habitability, democracy prioritises *contestability*, requiring that those governed are able to interrupt and configure the operations. Third, given habitability and contestability, democracy prioritises *extension*, expanding standing to entities already routed through infrastructural systems but not recognised as subjects.

My agenda is not to settle politics by computation, but to formalise controversy through contestable evaluators, so that the designation 'democracy' signifies a portable diagnostic for the political agonism and strife that remains irreducible to specific objectives. I conclude the paper by demonstrating how this specification can function as a generative grammar for algorithmic democracy through the *KI-DIPFIES* installation at my recent artistic exhibition in Kunstraum Memphis, where a swarm of AI agents enact the constraint hierarchy as pluriversal dramaturgy.

Professional Expertise, Practice Adaptation and Sense of Agency During AI Implementation in Radiology

Astrid Galsgaard, Ann-Louise Holten, Mikael Boesen & Felix C. Müller

This ethnographic study examines how narrow-task artificial intelligence (NT-AI) implementation reshapes professional agency and expertise hierarchies in high-stakes healthcare environments. As AI systems increasingly mediate decision-making across public sectors, understanding the power dynamics and social transformations embedded within these technologies becomes critical for democratic governance of algorithmic systems.

The research followed 14 radiologists across two Danish hospital departments over 12 months, examining knee osteoarthritis grading and breast cancer screening implementations. Building on Barley's socio-material perspective, the study employs a typology of four human-AI expertise practices: Parallel expertise, Forwarded expertise, Augmented expertise, and Collective expertise.

The findings reveal fundamental tensions in contemporary AI society. Trust and expertise hierarchies are destabilized when professionals must negotiate between algorithmic outputs and professional judgment, raising questions about who holds epistemic authority in increasingly automated public services. Responsibility allocation becomes contested as organizational structures position humans as accountable while AI substantially shapes outcomes - a contradiction reflecting broader challenges in governing algorithmic decision-making. Objective prioritization exposes conflicts between techno-capitalist efficiency imperatives, professional autonomy, and public service values.

Rather than passively accepting managerial visions of AI implementation, professionals developed five coping strategies to reconstitute their agency: disruption, re-validation, vigilance, rejection, and practice transformation. Critically, professionals rarely adhered to intended designs, instead actively navigating between expertise practices based on contextual demands. Collective expertise emerged most frequently even when unintended, suggesting interdependent collaboration naturally develops where both technical accuracy and professional accountability matter.

This research demonstrates that AI implementation is not merely technical but fundamentally political - a site where competing interests of organizational efficiency, professional autonomy, and patient care collide. Professionals' adaptive resistance reveals how social actors shape technological systems even as those systems constrain practice. The study contributes an analytical framework for understanding these power negotiations and offers implications for democratic AI governance: successful implementation requires flexibility for professionals to interpret and contextualize algorithmic outputs rather than rigid workflows that diminish professional judgment and public accountability.

The findings challenge techno-optimistic narratives of seamless human-AI collaboration, revealing instead the messy, contested processes through which AI becomes embedded in social practice.

They suggest that maintaining democratic values in public services requires organizational structures that support professional agency alongside algorithmic systems.

Algorithmic organising and governance of work and society through progressive encapsulation: Explainability as a political and social safeguard

Atif Sarwar & Emma Forsgren

As artificial intelligence systems become embedded in organisational and societal governance and decision-making, their role has evolved from being mere technical tools to infrastructures of rules. As the domain of opaque AI systems expands, it encompasses more tasks, authority, and social relations into its dominion through adaptive computational processes and becomes increasingly inaccessible to human understanding, a phenomenon termed '*progressive encapsulation*'. Developing on this concept, our paper argues that as this expansion is governed through outcome-based protocols, primarily focused on performance rather than internal workings of the AI systems, it has the capacity to generate a modern, technocratic analogue of the setting in which such opaque forms of power become normalised and accepted within ordinary organisational, administrative and social life.

Progressive encapsulation is a cumulative dynamic. AI systems first assist human decision-making in narrow contexts, but through continuous learning, increasing automation, and growing institutional reliance, they internalise additional functions, tasks and decision-making spheres whilst also reconfiguring surrounding decision structures. As the scope of their authority expands, the logic underpinning social outcomes becomes progressively hidden behind layers of technical complexity, proprietary design, and continual adaptation. Thus, whilst the algorithm's decision-making power grows, the capacity for public understanding, contestation, and democratic control contracts.

We argue that black-box AI uses the logic of exception to model its governance patterns by creating a zone in which *rules apply through their suspension.* Algorithmic systems become de facto and sovereign decision-makers that are no longer subject to justificatory requirements, answerable to legal directives, and bound by procedural constraints that usually govern public authorities. When regulators focus only on the outputs of these algorithmic systems, defined in terms of accuracy, efficiency, and aggregate outcomes, as opposed to how the systems reason, learn, and evolve, they reinforce and legitimise this *logic of exception*, where opaque technical processes continue to exercise real power without full accountability.

We further posit that this legitimation of the logic of exception-enabled algorithmic organising undermines three foundations of genuine socio-technical order. First, it erodes *epistemic accountability* as opaque decision logic inhibits reliable auditing, error diagnosis, and implementation of any meaningful oversight. Second, it weakens *democratic legitimacy*, as decisions are made by systems whose reasoning underpinning decision-making cannot be examined, challenged, or appealed. Finally, it disrupts the established norms of *moral and legal responsibility*, as it is distributed across permanently evolving human-machine assemblages; as a result, responsibility for an algorithmically enabled transgression is fragmented and systematically deflected.

Against such technocratic approaches to social and organisational organising that treat predictive success as sufficient for legitimacy, we launch a defence for *explainability* as a constitutional and social safeguard for algorithmic governance. We argue that explainability must be extended across the full AI lifecycle, with political and normative choices embedded in and underpinning the whole lifecycle. We contend that without explainability being an enforceable requirement, progressive encapsulation operating under the logic of exception potentially risks institutionalising a permanent social order that values operational efficiency, yet is ethically ignorant, politically unaccountable, morally contentious, socially polarising and legally unstable.

GenAI Driven Reconfiguration of Professional Expertise as Epistemic Expertise

Atif Sarwar & Jingqi Zhu

Generative AI (GenAI) is rapidly transforming professional work through both automation and augmentation, prompting professionals to reorganise tasks, develop new hybrid roles, negotiate control, and make sense of their professional expertise in response to AI. Current scholarship has focused on the impact of AI on professional expertise from two lens: resistance to protect traditional expertise or adoption to redraw professional scope. We advance a more precise theoretical argument: professional expertise is neither being displaced nor merely defended but fundamentally reconfigured as *epistemic expertise.* In this reconfiguration, professionals are increasingly acting as *epistemic arbiters,* the authoritative voice that determines what outputs are GenAI-generated and if these GenAI-generated outputs constitute *true, reliable, and actionable knowledge* or whether they are hallucinated, misleading, or false. We argue that this marks the rise of a new and expanding core domain of professional work that complements, rather than undermines, traditional professional/domain expertise.

Drawing on empirical research in the legal sector, we posit that this GenAI-enabled transformation reconfigures professional work through cumulative practice-based reconfigurations rather than generally assumed technological substitution. While GenAI automates document review, legal research, and contract drafting, it simultaneously augments professional decision-making by employing ML for large-scale pattern recognition. Current research suggests that law firms reorganise workflows to integrate these systems, in the process, redistributing tasks between machine and professionals, such that lawyers are still responsible for substantive legal reasoning, client representation, and accountability, while relying on GenAI for content production. It is this redistribution of tasks that essentially changes how professional expertise is enacted in practice. Whilst legal expertise continues to be central to interpreting statutes, constructing arguments, and advising clients, it is now also applied to the *epistemic evaluation of GenAI generated knowledge and content.*

Our preliminary findings show that as GenAI improves at producing seemingly authoritative legal arguments, case citations, and interpretations, lawyers' work now also involves scrutinising the sources in such outputs, detecting hallucinated or fabricated cases and citations, contextualising generic outputs, evaluating probabilistic claims, and importantly judging if such algorithmically generated claims are defensible in the legal and case context. We show that situated and algorithmic modes of knowing are interspersed with knowledge validating routines, reconfiguring professional and workplace jurisdictions concentrated around epistemic evaluations rather than information-based inference alone.

This evaluation work is focused on examining not just intelligibility but also the credibility and legitimacy of information in the case documents, thus reaffirming the professional authority and expertise through interpretation and epistemic qualification of the information rather than authorship of the professional decision. Thus, professional knowledge, expertise and discretion have become an epistemic filter employed to screen out misleading or erroneous GenAI generated outputs

in practice. Crucially, this does not hollow out professional authority; rather, it repositions professional expertise as the central mechanism through which GenAI-generated knowledge is qualified and evaluated as actionable.

Our contribution challenges dominant narratives of AI-driven deskilling or displacement, revealing instead how professional expertise evolves to encompass new forms of epistemic labour that are increasingly indispensable in AI-augmented work environments.

Technoscientific Sustainability Challenges Confronting ESMA Strategy Digitalization

Bartholomew Paudyn

Escalating (existential) climate crises (IPCC, 2023), AI-digitalization intensifying economic space concentration (Borch, 2026), but polarizing (populist) politics, increasingly, stress their fragile intersectional sustainability (ISSB, 2025). Consequent 'geoeconomic fragmentation' (IMF, 2023) and 'environmental, social and governance' (ESG) 'critical uncertainties' contestation (PRI, 2023), furthermore, focus attention on how AI/machine learning algorithmic-based techno-managerial capitalist dynamics and digital-driven finance aggravate 'deglobalization'. Straining the European Securities and Market Authority's (ESMA) capacity to execute its *Strategy 2023-28*, this paper contends, it jeopardizes Europe's green transition and UN SDGs/Agenda 2030. Specifically, how do algorithmic/automated decision-making (ADM) conflicts plague ESMA *Strategy's* principal thematic drivers: a) enabling sustainable finance; b) facilitating technological innovation and effective data use? Given transparency, reliability, and trust are ESMA prerequisites for interoperability, involving pan-EU AI Act/Office supervisory convergence, yet diminishing ('frontier') AI interpretability, how are verification and regulation impacted? Assessing EU/ESMA's data science and techno-managerial modalities reveals the positive/negative externalities of sustainable finance AI-ML anticipatory technologies for European decarbonization and SDG success.

First, aligned with EU Regulation 2024/3005 ('ESG Rating Regulation'), ESMA's green transition facilitation and investment oversight entails ensuring ESG rating quality and integrity, especially curbing (AI) 'green-washing' (Svetlova, 2025). Nonlinear climate catastrophes, societal inequalities/upheaval, or wars escape capture through market (ergodic) risk-dominant ESG rating econophysics, which ill-informs futurity financialization. ESG score-assimilated AI-ML, I argue, widely institutionalizes dubious technoscientific measures, AI 'hallucinations', plus systematic sustainability failures. ESG data scarcity increases its AI-ML synthetic modelling and (automatic) hyperparameter optimization. Unfortunately, major controversies exist regarding (uncertain) ESG econophysics automation, its environmental/human harm, or ESMA's successful supervision. Given ESG investment AI-ML operability and governance challenges, this section problematizes ESMA's strategic enablement of sustainable finance, plus (global) climate change consequences.

Moreover, EU's/ESMA's above green transition conundrums are exacerbated by *Strategy* facilitation of technological innovation and effective data usage. Rather than AI catalyst, regulation frequently lags innovation. Europe's official response is the most comprehensive and ambitious to date. But failing to address serious AI/ADM distortions entangling sovereign/ESG ratings, ESMA may reinforce automated (neoliberal) quant dominance and sanction the predatory financial firepower against its democracies, and SDGs. AI Act harmonization compounds complexities and an inadequate (ergodic) risk-based approach exacerbating implementation (August 2026) and business operations. My analysis reveals how EU regulatory complicity in this digitalized depoliticization of the democratic chain of power is extensively predicated on, and misguided by, forms of ESMA reflexive legitimacy rooted in AI-ML management through risk/uncertainty.

Rather than exploit potential AI/ADM investment strategies for collective EU purposes, this paper explores how suspect ESMA techno-managerial modalities expose countries to critical imbalances. Bridging business management (Beunza, 2019; Faulconbridge et al., 2023; Millo et al., 2024), economic sociology/STS (Borch, 2026; Svetlova, 2025), and political economy (Best, 2024; Paudyn, 2026), I problematize how controversial ESG ratings-integrated AI-ML threatens systemic failures fueling national/climate degradation: deglobalization. ESMA's risk mentality/mechanics digitalization techno-managerial approach is evaluated for its green business merits and decarbonization/SDG consequences. History has repeatedly demonstrated that divorcing technoscientific expertocracy from its (ESG) political economy, risk econophysics precipitates severe corrections.

The Model is the Mediator: LLMs and the Gatekeeping of Information in the Public Sphere

Ben Potter

Public communication is central to democracy because citizens form opinions through shared processes of interpretation and debate. But the horizons of possibility of such communications are increasingly shaped by digital technologies. Historically, journalists and editors acted as gatekeepers by selecting, verifying, and framing information according to professional norms of accountability and accuracy. The platform era reorganised circulation around an economy of attention, where algorithmic recommendation systems determined what content appeared to whom, without the interpretive standards that underpinned editorial mediation. Yet the work of making sense of public matters through explanation, contextualisation and justification remained a human activity. The introduction of large language model (LLM) chatbots marks a further shift towards digital automation as these systems generate responses that take the form of explanation, synthesis, and guidance – tasks once squarely situated in the arena of human action.

This paper argues that such changes mean LLMs are emerging as *synthetic gatekeepers:* systems that automate interpretive mediation by generating plausible and coherent accounts of information on demand, but without the professional norms of accountability that have historically guided editorial judgement. In doing so, they function not only as technical infrastructure but also as de facto authorities over meaning, subtly shaping the scope and quality of information that circulates in the public sphere. This shift prompts a central question: *What norms, values, and assumptions inform chatbot responses in place of journalistic standards, and what does this mean for how citizens come to form judgements about matters of shared concern?* To answer this question, this paper draws on interviews with AI practitioners, publicly available documents from AI companies, and survey data indicating rising reliance on chatbots for everyday information tasks (Simon, 2025). It extends work on digital transformations of the public sphere (Habermas, 2022; Seeliger & Sevignani, 2022; Schroeder & Juenger, 2023) by showing that interpretive mediation is now being *automated* through predictive language models optimised for plausibility and engagement.

Early findings indicate that while they provide efficient and personalised access to information, they can also fabricate or distort it, with the underlying representational logics obscured by proprietary training and alignment regimes. Moreover, as LLMs lower barriers to accessing knowledge, they also narrow interpretive horizons by smoothing disagreement, standardising explanatory frames, and presenting interpretations as neutral. As users increasingly outsource interpretive labour, epistemic agency—the capacity to evaluate claims and form independent judgement—may be weakened (Coeckelbergh, 2025). As a result, public communication may become more efficient but also more susceptible to subtle influence, particularly as business models shift from funding journalism to monetising the automation of knowledge gatekeeping.

The paper concludes that synthetic gatekeeping constitutes a structural transformation in the flows of information within the public sphere. If democratic life requires plural, contestable, and account-

able forms of interpretation, then governance must address not only misinformation, but the infrastructures that now shape how understanding itself is produced.

Welfare rights in transition: From direct to automated eligibility control

Benjamin Schwarz & Rikke Frank Jørgensen

This presentation draws together insights from two related research projects: one examining how digitalization transforms the enactment of welfare rights in Denmark, and another exploring how automation reconfigures state–citizen relations across eight European welfare regimes (Kaun, Lomborg, & Allhutter, in press).

In Denmark, welfare surveillance—the institutional monitoring of and investigation into benefit eligibility—has undergone profound changes over the past two decades. Control work has shifted from decentralized, face-to-face investigations by municipal "control units" toward centralized, data-driven monitoring by Udbetaling Danmark, supported by algorithmic risk models designed to detect irregularities pointing to fraud. We argue that this transformation is not merely organizational, technological, and operational but reflects an ideological reorientation of welfare governance.

Drawing on the Sociology of Critique (Boltanski & Thévenot, 2006), we conceptualize eligibility assessments as "control tests" embedded in competing normative worlds. Based on interviews and observations within municipal control units and Udbetaling Danmark, we show how controversies surrounding municipal practices—criticized for arbitrariness, wastefulness, and invasiveness—enabled a shift toward an "industrial" regime (Dubois, 2019) privileging standardization, efficiency, and distance as markers of legitimacy.

This shift is discursively framed through institutional self- and other-positioning: municipalities invoke proximity, discretion, and relational knowledge as expressions of a "humane" approach, while Udbetaling Danmark emphasizes objectivity, uniformity, and professional detachment as progress. These narratives do more than describe operational differences; they articulate competing visions of fairness and professional responsibility in welfare governance.

Our findings highlight how the marginalization of face-to-face practices reduces what practitioners describe as the "human eyes" needed to interpret complex cases. While such practices were never inherently empathetic, they carried an empathetic potential (Ranchordás, 2022)—when rooted in trust-based organizational cultures—that is difficult to reproduce in semi-automated systems. The broader move toward automated suspicion (Jørgensen et al., in press) raises questions about the normative foundations of the tech-heavy welfare state and the everyday enactment of welfare rights for its most vulnerable citizens.

Uneventful: The Development of a Depoliticized Hyperscale Data Center Region

Brit Ross Winthereik & Caroline Anna Salling

The global and rapid expansion of hyperscale data centers, driven by advancements in artificial intelligence (AI), presents new challenges as these factories transform social and natural environments in significant ways. This paper examines the processes that have characterized the placement of data centers in Denmark, and explores the possibilities within the hyperscale for Social Science. Through ethnographic research, we contend that "the hyperscale data center region" represents a political construct that serves to depoliticize the adverse effects associated with data centers. The research introduces the concept of "hyperscalism" to describe an alliance between state actors and big tech companies, where uneventful establishment leads to the erasure of place.

The Rhetorics of Rationalism and the Truths of AI Slop

Caddie Alford

In 2024, computer scientist Timnit Gebru and philosopher Émile P. Torres published "The TESCREAL Bundle: Eugenics and the Promise of Utopia Through Artificial General Intelligence," which argues that the most prominent Silicon Valley ideologies animating so-called "artificial general intelligence" are extensions of eugenics movements. "TESCREAL" is short for Transhumanism, Extropianism, Singularitarianism, Cosmism, Rationalism, Effective Altruism, and Longtermism. These ideologies promote a new wave of eugenics: some argue that AI hybridity will optimize human "stock" and others argue that humanitarian goals are less important than funding "AI alignment."

Although TESCREAL on the surface might seem fringe, they've provided scaffolding for an incredibly influential American political movement: the anti-democratic "new right," overlapping reactionaries and the alt-right. In 2014, Thiel wrote to Curtis Yarvin: "one of our hidden advantages is that these people"—progressives—"wouldn't believe in a conspiracy if it hit them over the head (this is perhaps the best measure of the decline of the Left). Linkages make them sound really crazy."[5] To take him up on that provocation is to see that linkages between neoreactionary elite, the AI field, and TESCREAL are everywhere. Both Peter Thiel and Elon Musk espouse transhumanism beliefs. Peter Thiel bankrolled J.D. Vance's Senate campaign; nowadays, Vance openly cites Yarvin.

Where these linkages get even more wild is with Eliezer Yudkowsky and the rationalism community on his sycophantic "LessWrong" blog. Yudkowsky is the reason Thiel financially supported OpenAI.[6] The "LessWrong" blog is widely read by OpenAI researchers and engineers.[7] And on and on. Yudkowsky's rationalism is a cultish group that believes rationality is the only tool humans need to solve problems: to them, human potential is nothing without a singular focus on decision theory. Their vision of "truth" is that truth is simple; the world is knowable with total clarity; and epistemic certainty is the blog's dominant disposition.

This interdisciplinary presentation will analyze the rationalism community's rhetorics as they mirror and substantiate AI. On the one hand, I will construct a corpus of texts from Yudkowsky and other rationalists to parse their strategies while also demonstrating that their salience lies in what are aesthetic appeals. On the other hand, I will make the case that these rhetorics were so impactful[8] that they have had a hand in constructing the pernicious AI hype that characterizes tech culture today. Ultimately, I turn to how people experience AI every day—the low-quality AI-generated con-

5 Kofman, Ava, "Curtis Yarvin's Plot Against America," *The New Yorker*, June 2, 2024, https://www.newyorker.com/magazine/2025/06/09/curtis-yarvin-profile?fbclid=IwY2xjawLzdaxleHRuA2FlbQIxMABicmlkETFFY2NuWEtwOHpsREZWRnlhAR7fz6YZ80BqTkOEhWdlr_GiTPzxvmeqvc6HZlhPTGnDT4QVgN-cdZVahM1tNQ_aem_Gbx8fY47meH-1rae13dVzfg

6 Hagey, Keach, "How Peter Thiel's Relationship with Eliezer Yudkowsky Launched the AI Revolution." *Wired,* May 20, 2025, https://www.wired.com/story/book-excerpt-the-optimist-open-ai-sam-altman/

7 Ibid.

8 Elizabeth Sandifer's *Neoreaction a Basilisk: Essays on and Around the Alt-Right* (2017) argues that Yudkowsky and the "LessWrong" rhetorical style make tech bros feel powerful.

tent on social media platforms known as AI slop—to puncture the TESCREAL fantasies about AI, in turn laying bare what these "logical" rationalists actually believe while modeling how to critique AI hype.

Re-Imagining the TRIPS Agreement to Enable Gen-AI Medical Tools as Catalysts for Global Health Equity.

Cailin Morrison

The rapid evolution of artificial intelligence (AI) in healthcare has ushered in a new era of transformative medical technologies. Generative AI (Gen-AI) medical tools, offer unprecedented opportunities to enhance diagnostics, treatment planning and public health surveillance. However, without adequate governance and equitable distribution mechanisms, these innovations risk deepening existing global health disparities, particularly among the World Trade Organisation (WTO) member states in the Global South.

This paper explores how international legal frameworks, specifically the WTO Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), could be reimagined to support the equitable dissemination of Gen-AI enhanced healthcare technologies.

Drawing on law and development literature, this paper critiques the neoliberal formalist approach that underpins much of intellectual property rights governance. The paper argues for a rethinking of development that embraces market mechanisms while deploying law as a tool for institutional transformation and social justice. This theoretical position challenges the "one-size-fits-all" approach to legal reform and instead advocates for a context-sensitive, capabilities approach to health equity.

The paper presents a legal framework for re-considering the TRIPS provisions to facilitate the use of Gen-AI medical tools in low- and middle-income countries (LMICs). It emphasises the need for WTO member states to implement regulatory strategies that empower governments to respond to public health crises with agility and equity. Key strategies include expanding the scope of compulsory licensing to cover AI-enhanced healthcare technologies, robust governance on the legal status of AI-generated medical tools and fostering international cooperation for technology transfer and capacity building in the Global South.

The paper argues that international law can serve as a dynamic instrument of change. By leveraging existing WTO legal mechanisms and reimaging their application in the context of emerging healthcare technologies, policymakers can find innovative ways to overcome structural barriers and advance global health equity. This contribution is intended to inform the ongoing debate around AI governance, AI ethics, intellectual property reform and global health justice. It offers a legal roadmap for integrating Gen-AI technologies into public health infrastructure in a manner that prioritises health equity.

Behavioural modification or behavioural change? Steps to putting surveillance capitalism in its place

Carl Stefan Roth-Kirkegaard

Critical political economy (Harvey, 2007) and labour process theory (Braverman, 1969) classically assume that the traditional trajectory of capitalism is to reduce the work process and social complexity to a minimum, to the money proxy (see also Hayek, 1945), which allows for the most efficient and least costly economic transactions and accumulation of capitalist value.

With the advent of surveillance capitalism, this situation appears to be changing (Zuboff, 2019). Here, the complexity of labour, consumer, and citizens' behaviour can be thoroughly predicted and manipulated to approximate guaranteed outcomes. This is achieved through the collection of vast amounts of surplus data on people's digital behaviour and the computational power to transform a network of these insights into knowledge of preferences and tools for creating, influencing, and shaping human behavioural patterns. Thus, in contrast to earlier versions, we expect capitalism now to function more (at least theoretically) through the nuanced computability, understanding, creation and manipulation of social complexity and human needs.

However, this presentation will argue that the surveillance capitalism diagnosis has led us to embrace an endogenous capitalist understanding of its development, where we come to overestimate its capacity to control social complexity (Kunda, 1995; Hardt and Negri, 2001; Boltanski & Chapello, 2005), now through computation (Zuboff, 2019). This is because we underestimate the practical limitations, stepwise adjustments, and differences in use cases that condition the implementation of data algorithms by the capitalist system in the emergence of a new capitalist logic.

Instead of Zuboff's notion of the dominance of approximating precise guaranteed outcomes, the presentation suggests a bifurcation in data-economic logics that intersect and interact. The presentation closely examines the key elements of Zuboff's instrumentarian power and compares them to Harvey's (2007) discussion of the core dynamics in the emergence of industrial capitalism. This comparison suggests that two intersecting data logics are at play within the core illustrations of Zuboff's argument. One is directed at behavioural modification and control by approximating infinite detail, and the other is directed at behavioural change, a generic push for action from simpler data points, which enables a framed chaos of activity.

Crucially, the latter utilisation of data does not necessarily lead to more predictive control over the complexity of social actions but instead disperses information and animates activity within a relevant frame, which is appropriate for many capitalist processes concerned only with generic sales. This is exemplified by Zuboff's example of restaurants using Pokémon Go to generate human activity near themselves without any subsequent control of this behaviour.

This suggests nuancing our expectations for the increasingly detailed control over social complexity in capitalist organisations, as previously done for neoliberal control regimes, and exploring the interaction between behavioural modification and change.

Closing the Gap: Tackling Strategic, Organisational, and Governance Challenges in AI Adoption for Manufacturing

Carla Goncalves Machado

Manufacturing firms are increasingly pursuing AI-led transformations (e.g., predictive maintenance, quality inspection, and autonomous production), yet implementation often stalls after pilot projects. Prior literature links this "pilot-to-scale" gap to disorganised data and legacy IT (Clemens et al., 2023; Plathottam et al., 2023; Rauh et al., 2022), misaligned investment logics and budgeting routines (Hadhri et al., 2025), and capability deficits spanning technical skills and digital leadership (Mqoqi et al., 2026; Obi et al., 2025). At the same time, responsible AI introduces additional coordination challenges because accountability, transparency, and compliance remain difficult to operationalise across organisational layers and heterogeneous use cases (Besinger et al., 2025; Engström et al., 2025). Based on previous studies (Engström et al., 2025), we argue that AI adoption is a socio-technical change where value arises from combining technology (models, data pipelines, and platforms) with decision-making rights, incentives, workflows, skills, and accountability. If this combination is weak, organisations might fall into "pilotism," resulting in many projects but little learning or integration. This study asks: *What organisational and governance mechanisms are required in manufacturing firms to convert AI aspirations into scalable and beneficial implementations?*

To answer our research question, a qualitative approach combining a focused literature review with co-creation workshops involving managers and specialists from six companies (A-F) was employed (Ahmed & Asraf, 2018) to identify AI goals, perceived blockers, and scaling mechanisms through discussions and group tasks (Ørngreen & Levinsen, 2017). The empirical material (field notes, audio recordings, and secondary data) was analysed using content analysis to find recurring patterns (Bengtsson, 2016).

Across firms, evidence confirms literature: AI remains stuck in fragmented pilots because strategy, budgeting, data/legacy IT, skills, and governance are misaligned. More specifically, AI remains stuck because key organisational elements are misaligned. Firms often run many promising Proofs of Concept (PoCs), but the organisational system required for scaling (strategy → funding → data/IT → people → governance) does not "fit together", preventing pilots from being industrialised into scalable, accountable deployments. There is a notable lack of shared direction, which results in isolated initiatives that fail to contribute to overall enterprise learning or effective scaling. The current funding model adheres to traditional ROI metrics, which create barriers to securing resources for scaling successful pilots, as they often fail to meet conventional thresholds or timelines. Many pilots are developed under conditions that do not generalise well, leading to challenges in scalability. Companies lack robust pipelines and deployable infrastructure necessary for broader implementation. There is an over-reliance on a limited number of individuals, as roles and learning routines are not well-defined. This situation hinders the establishment of standardised practices across teams. The absence of clear accountability and risk control mechanisms has blocked the operationalisation of successful pilot initiatives. On the other side, companies are working to tackle challenges by imple-

menting a "governance spine" that merges centralised direction with decentralised execution. This approach features a dedicated AI strategy, clear decision rights, and local ownership through AI ambassadors and cross-functional teams. Key elements include a structured use-case pipeline (idea → business case → pilot → adoption), standardised tracking of ROI and KPIs, and role clarity with capability-building infrastructures for ongoing learning. Strong data governance ensures security and compliance. Successful scaling is more about organisational change than technology, requiring effective leadership and adaptive resourcing to align initiatives with operational needs.

References

Ahmed, S., & Asraf, R. M. (2018). The workshop as a qualitative research approach: Lessons learnt from a "critical thinking through writing" workshop. *The Turkish Online Journal of Design, Art and Communication, 8*(Special Issue), 1504–1510.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open, 2*, 8–14.

Besinger, P., Zaremba, J., & Ansari, F. (2025). Responsible AI in manufacturing: The case of accountability in AI systems. *IFAC-PapersOnLine, 59*(10), 1325–1330. https://doi.org/10.1016/j.ifacol.2025.09.223

Clemens, F., Willemsen, F., Mütze-Niewöhner, S., & Schuh, G. (2023). Development of a task model for artificial intelligence-based applications for small and medium-sized enterprises. In D. Romero, E. Alfnes, A. Romsdal, J. O. Strandhagen, & G. von Cieminski (Eds.), *Advances in production management systems: Production management systems for responsible manufacturing, service, and logistics futures* (Vol. 689, pp. 528–542). Springer. https://doi.org/10.1007/978-3-031-43662-8_38

Engström, A., Pittino, D., Mohlin, A., Edh, N., & Johansson, A. (2025). A paradox perspective on early AI adoption: Understanding temporal and relational tensions. *Journal of Organizational Change Management*, 1–27. https://doi.org/10.1108/JOCM-02-2025-0086

Hadhri, S., Zargar, F. N., Naeem, M. A., & Chibani, F. (2025). Bridging sustainable finance, AI, and clean technology amid economic shocks: How are they connected in median and extreme conditions? *Journal of Environmental Management, 391,* 126375. https://doi.org/10.1016/j.jenvman.2025.126375

Iriarte, I., Hoveskog, M., Nguyen Ngoc, H., Legarda, I., Uranga, M., Nazabal, M., & Atxa, A. (2023). Service design for digital servitization: Facilitating manufacturers' advanced services value proposition design in the context of Industry 4.0. *Industrial Marketing Management, 110*, 96–116.

Mqoqi, M., Turpin, M., & Van Belle, J.-P. (2026). AI skills and technologies for future-proofing businesses in the manufacturing industry. In T. Senjyu, C. So-In, & A. Joshi (Eds.), *Smart trends in computing and communications (SmartCom 2025)* (pp. 113–123). Springer. https://doi.org/10.1007/978-981-96-7807-5_11

Obi, L. I., Osuizugbo, I. C., & Awuzie, B. O. (2025). Closing the artificial intelligence skills gap in construction: Competency insights from a systematic review. *Results in Engineering, 27*, 106406. https://doi.org/10.1016/j.rineng.2025.106406

Ørngreen, R., & Levinsen, K. T. (2017). Workshops as a research methodology. *Electronic Journal of e-Learning, 15*(1), 70–81.

Plathottam, S. J., Rzonca, A., Lakhnori, R., & Iloeje, C. O. (2023). A review of artificial intelligence applications in manufacturing operations. *Journal of Advanced Manufacturing and Processing, 5*(3). https://doi.org/10.1002/amp2.10159

Rauh, L., Reichardt, M., & Schotten, H. D. (2022). AI asset management: A case study with the asset administration shell (AAS). In *2022 IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA)* (pp. 1–8). IEEE. https://doi.org/10.1109/ETFA52439.2022.9921705

Visual Manipulation in Health Scam Advertising on Meta: an Exploratory Analysis of Deepfake and Cheapfake Practices

Carlos Eduardo Barros, Débora Gomes Salles, Nicole Sanchotene & Rose Marie Santini

Problem Statement

Research across several countries has identified an escalating wave of scam advertising on social media, driven by persuasive disinformation strategies that exploit platform infrastructures to build audience trust and engagement (Alzghoul et al., 2024; Di Domenico et al., 2022). Concerns around this phenomenon have intensified with the spread of generative AI tools, which facilitate the production and circulation of deceptive content (Simon et al., 2023), particularly manipulated videos that imitate authoritative figures in order to project legitimacy (de Ruiter, 2021; Hynek et al., 2025). Notably, previous studies show that similar ads have been routinely disseminated in the country through the Meta advertising system (Santini et al, 2025), and a recent investigation in Brazil illustrates this trend, revealing online courses that teach how to produce health scam ads using deepfake techniques (Aleixo, 2025).

Research Design and Methodology

This exploratory study examines the range of visual disinformation tactics employed in health scam advertising campaigns uncovered in the Brazilian Meta Ad Library. Between July and December 2024, Barros et al. (2025) collected 3,710 ads featuring manipulated videos of one of Brazil's most prominent medical figures, Drauzio Varella. Of these ads, 31,76 percent (1,178) were produced by the four most active advertisers. Focusing on these core advertisers, we grouped ads with identical texts and conducted a non-exhaustive manual review of their media to identify qualitative indicators of visual manipulation, drawing on signals described in prior research (Alsobeh et al., 2024; Hameleers et al., 2025; Rahman et al., 2022; van Dijk, 2025).

Results

The findings show that the ads relied on only 12 distinct text variations, and their videos exhibited consistent manipulation patterns across all four advertisers, including pronounced audio image asynchrony, biological anomalies such as irregular eye blinking, and resolution or warping inconsistencies (Rahman et al., 2022). Rather than producing fully hyperrealistic results, these videos often combine more advanced elements, such as AI generated audio, with low resolution or otherwise degraded footage. Beyond these media based anomalies, the videos employ scientific language and visual markers such as lab coats, laboratory environments, and journalistic studio aesthetics (Al-

Sobeh et al., 2024), while misappropriating images of authoritative individuals and institutions. Detecting the deceptive nature of these messages therefore requires critical scrutiny (Van Dijk, 2025) and substantial prior knowledge of the emulated sources (Hameleers et al., 2024).

Discussion and Contributions

This case contributes to broader debates on how scam advertising blends sophisticated and rudimentary manipulations to target users online, operating between deepfakes and cheapfakes. The evidence also raises questions about platform responsibility in addressing clear cases of manipulation (Hameleers et al., 2024) and highlights avenues for future research on whether such cheapfakes are intentionally crafted to reach users in more vulnerable contexts (AlSobeh et al., 2024; Van Dijk, 2025), who may lack the media literacy needed to identify basic manipulation cues.

References

Aleixo, I. (2025, June 8). *Curso de R$ 1.000 ensina a fazer propaganda falsa de produto ilegal com IA*. UOL. https://noticias.uol.com.br/confere/ultimas-noticias/2025/06/08/curso-r-1000-ensina-anuncio-falso-produto-ilegal-inteligencia-artificial.htm

Alsobeh, A., Franklin, A., Woodward, B., & Siegelman, J. (2024, October 3). *Unmasking media illusion: Analytical survey of deepfake video detection and emotional insights.*

Alzghoul, J. R., Abdallah, E. E., & Al-khawaldeh, A. S. (2024). Fraud in Online Classified Ads: Strategies, Risks, and Detection Methods: A Survey. *Journal of Applied Security Research, 19*(1), 45–69. https://doi.org/10.1080/19361610.2022.2124328

de Ruiter, A. (2021). The Distinct Wrong of Deepfakes. *Philosophy & Technology, 34*(4), 1311–1332. https://doi.org/10.1007/s13347-021-00459-2

Barros, C. E., Sanchotene, N., Salles, D. G., & Santini, R. M. (2025). "Dr. Drauzio Varella tem a solução dos seus problemas": mapeando anúncios fraudulentos sobre saúde na Meta. In *COMPÓS* (Vol. 34). https://publicacoes.softaliza.com.br/compos2025/article/view/11340

Di Domenico, G., Nunan, D., & Pitardi, V. (2022). Marketplaces of Misinformation: A Study of How Vaccine Misinformation Is Legitimized on Social Media. *Journal of Public Policy & Marketing, 41*(4), 319–335. https://doi.org/10.1177/07439156221103860

Hameleers, M., van der Meer, T., & Vliegenthart, R. (2025). How persuasive are political cheapfakes disseminated via social media? The effects of out-of-context visual disinformation on message credibility and issue agreement. *Information, Communication & Society, 28*(1), 61–78. https://doi.org/10.1080/1369118X.2024.2388079

Hynek, N., Gavurova, B., & Kubak, M. (2025). Risks and benefits of artificial intelligence deepfakes: Systematic review and comparison of public attitudes in seven European Countries. *Journal of Innovation & Knowledge, 10*(5). https://doi.org/10.1016/j.jik.2025.100782

Rahman, A., Islam, M. M., Moon, M. J., Tasnim, T., Siddique, N., Shahiduzzaman, M., & Ahmed, S. (2022). *A Qualitative Survey on Deep Learning Based Deep fake Video Creation and Detection Method*. https://dro.deakin.edu.au/articles/journal_contribution/A_Qualitative_Survey_on_Deep_Learning_Based_Deep_fake_Video_Creation_and_Detection_Method/28509758/1

Santini, R. M., Salles, D., Mattos, B., Sanchotene, N., Belin, L., Mendes, A., Dias, B., Barros, C. E., Pinho, D., Silva, D., Grael, F., Ferreira, F., Haddad, J., Murakami, L., Canavarro, M., Borges, M., Gomes, M., & de Medeiros, P. (2025). Atingidos pelas redes sociais: Os impactos da indústria da desinformação nos consumidores brasileiros. https://doi.org/10.29327/5566995

Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-127

van Dijk, S. (2025). *Under pressure: The influence of psychological and cognitive factors in deepfake detection under time pressure*. https://www.utupub.fi/handle/10024/194092

When the Marketplace Learns You: Automated Profiling as a Structural Challenge to EU Consumer Law

Carolina Lisboa Pinto

Every time we shop online, AI systems analyze our behavior to detect when we are tired, impulsive, anxious, short on time, or unusually persuadable, and platforms quietly adjust what we see in order to maximize profit. This is no longer speculative: it is the operational logic of today's digital marketplaces, and arguably the most pervasive real-world deployment of AI. Automated profiling now identifies and exploits behavioral vulnerabilities at a scale and precision no human trader could replicate. Yet the EU legal framework meant to protect consumers was built for a world in which commercial influence was visible, slow, and directed at broad audiences. My PhD research, which is part of the Dutch Research Council (NWO) funded Consumer ID project, examines this widening mismatch and argues that the current rules are structurally incapable of governing AI-driven manipulation in the marketplace.

Profiling today functions as an infrastructure of behavioral influence. It infers personal traits, predicts susceptibilities, and continuously personalized interfaces by reordering search results, tailoring recommendations, adjusting prices, or triggering urgency messages at precisely the moment a user is most likely to respond. These interventions are subtle, individually optimized, and updated in real time. As a result, each consumer encounters a different marketplace shaped around their predicted weaknesses, which they cannot see, understand, or compare to others. This produces new forms of digital vulnerability rooted not in a traditional lack of information, but in profound asymmetries of data, insight, and computational power.

The EU's consumer-protection framework, particularly the Unfair Commercial Practices Directive, is misaligned with these realities. It assumes that persuasion is recognizable, that consumers can protect themselves through information, and that harmful practices can be assessed through the lens of an "average consumer." Automated profiling breaks all three assumptions. Influence becomes invisible and personalized; there is no uniform practice for regulators or courts to examine. Profiling enables precisely timed interventions that exploit behavioral patterns the law has no tools to detect. And transparency-based protections lose their force, since algorithmic personalization cannot be meaningfully interpreted or resisted, regardless of disclosure.

The implications extend far beyond legal doctrine. Profiling challenges foundational values that underpin how multiple disciplines understand digital society. *Autonomy* is compromised when choices are shaped before individuals recognize the influence. *Power asymmetries* intensify as platforms accumulate behavioral insights and optimization capabilities that consumers cannot monitor or counterbalance. *Fairness and equal treatment* become uncertain when individuals are sorted into opaque, inference-based categories that determine what they see, what they pay, and which opportunities are available to them. These issues sit at the intersection of AI governance, behavioral science, political economy, ethics, and communication studies.

Because digital marketplaces mediate access to goods, services, and information essential to everyday life, these concerns affect everyone. Profiling reshapes relations of production and consumption, embeds behavioral governance into the infrastructures of techno capitalism, and redefines how individuals navigate (and are navigated by) the digital public sphere.

This presentation maps the structural misalignment between rapidly evolving profiling technologies and the assumptions built into EU consumer law. It proposes a conceptual framework for understanding this misalignment and invites cross-disciplinary engagement to rethink how digital marketplaces should be governed, and consumers protected in an AI-mediated society, and what forms of oversight, institutional design, or social intervention may be needed as AI continues to reorganize everyday economic life.

Building Smarter and Safer Workplaces: Empowering Workers in the Era of Industry 5.0

Chantal Ho

In the transition to Industry 5.0, safety and job satisfaction rise or fall together. Driven by the widespread implementation of artificial intelligence (AI), Industry 5.0 promises human-centric, resilient and sustainable technologies (European Commission, 2021). The organisational importance of ensuring safe and satisfying work becomes more complex. Research indicates that workers who perceive a strong safety climate report significantly higher job satisfaction (Ayim Gyekye, 2005; Smith, 2018). Uncomfortable working environment can generate stress, both on a daily level and on a more critical level (Vischer, 2007). Increasingly, safety encompasses not only physical safety but includes psychological safety as AI systems reshape job roles and autonomy. However, AI implementation is a double-edged sword. It can reduce exposure to hazardous tasks but may also diminish workers' sense of autonomy (Niehaus et al., 2022). This is particularly the case when low psychological safety limits their ability to voice concerns about AI implementation (Van Den Berg et al., 2025). As autonomy is known as a job resource (Bakker & Demerouti, 2017), reductions in autonomy can ultimately weaken job satisfaction and motivation (Ter Hoeven et al., 2016).

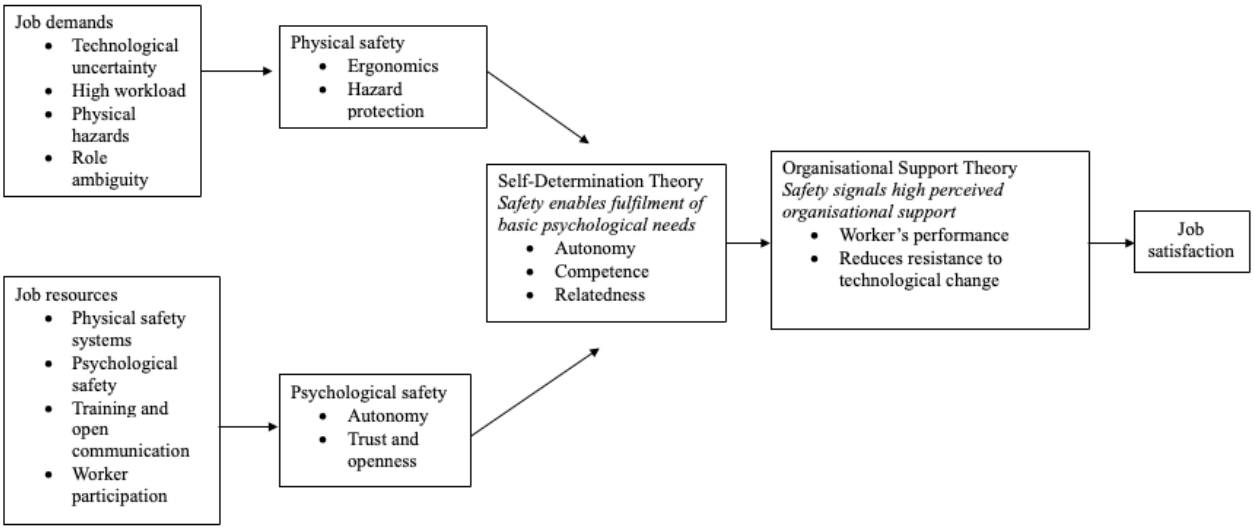
Drawing on empirical and theoretical studies in socio-technical, psychological and organisational communication research, this paper develops an integrated understanding of how creating safer workplaces can lead to satisfied workers in the context of Industry 5.0. As there has been insufficient research into how safety perceptions evolve in this context (Fiegler-Rudol et al., 2025), this paper aims to address this conceptual gap. This comparative analysis synthesises cross-disciplinary research through the lenses of the Job Demands – Resources (JDR) model (Bakker & Demerouti, 2017), Self-Determination Theory (SDT) (Deci et al., 2017), and Organisational Support Theory (OST) (Kurtessis et al., 2017). Accordingly, the central research question is: How do the JDR model, SDT, and OST explain the mechanisms linking physical and psychological safety to job satisfaction in the context of Industry 5.0?

The analysis drafted a multi-layered framework (Figure 1) to answer this question. First, the JDR model offers a foundational explanation of safety's role in job satisfaction (Bakker & Demerouti, 2017). Physical hazards, technological uncertainty, and increased workload exemplify job demands that can increase risk. Conversely, physical safety systems, psychological safety and supportive communication are considered job resources that encourage job satisfaction (Niehaus et al., 2022). Second, SDT deepens it by clarifying why safety matters psychologically (Deci et al., 2017): safe AI-enabled environments support autonomy, competence and relatedness through learning and open communication. OST (Kurtessis et al., 2017) adds a third layer by explaining how safety affects perceived organisational support. When workers perceive safety as a priority, they interpret technological change as being in line with their well-being, leading to higher job satisfaction.

This paper integrates the JDR model, SDT, and OST and proposes that safety is a multidimensional source and foundational condition for job satisfaction in Industry 5.0. Job satisfaction is supported by both physical and psychological safety; therefore, organisations should prioritise them in their

technology implementation processes. This can be achieved by ensuring autonomy-supportive design, transparent communication and organisational support to maintain job satisfaction in increasingly AI-enabled working environments.

Figure 1: Multi-Layered Framework on Workplace Safety and Job Satisfaction.



References

Ayim Gyekye, S. (2005). Workers' Perceptions of Workplace Safety and Job Satisfaction. *International Journal of Occupational Safety and Ergonomics, 11*(3), 291–302. https://doi.org/10.1080/10803548.2005.11076650

Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology, 22*(3), 273–285. https://doi.org/10.1037/ocp0000056

Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-Determination Theory in Work Organizations: The State of a Science. *Annual Review of Organizational Psychology and Organizational Behavior, 4*(1), 19–43. https://doi.org/10.1146/annurev-orgpsych-032516-113108

European Commission. Directorate General for Research and Innovation. (2021). *Industry 5.0 – Towards a sustainable, human-centric and resilient European industry.* Publications Office of the European Union. https://doi.org/10.2777/308407

Fiegler-Rudol, J., Lau, K., Mroczek, A., Kasperczyk, J., Fiegler-Rudol, J., Lau, K., Mroczek, A., & Kasperczyk, J. (2025). Exploring Human–AI Dynamics in Enhancing Workplace Health and Safety: A Narrative Review. *International Journal of Environmental Research and Public Health, 22*(2). https://doi.org/10.3390/ijerph22020199

Kurtessis, J. N., Eisenberger, R., Ford, M. T., Buffardi, L. C., Stewart, K. A., & Adis, C. S. (2017). Perceived Organizational Support: A Meta-Analytic Evaluation of Organizational Support Theory. *Journal of Management, 43*(6), 1854–1884. https://doi.org/10.1177/0149206315575554

Niehaus, S., Hartwig, M., Rosen, P. H., & Wischniewski, S. (2022). An Occupational Safety and Health Perspective on Human in Control and AI. *Frontiers in Artificial Intelligence, 5*. https://doi.org/10.3389/frai.2022.868382

Smith, T. D. (2018). An assessment of safety climate, job satisfaction and turnover intention relationships using a national sample of workers from the USA. *International Journal of Occupational Safety and Ergonomics, 24*(1), 27–34. https://doi.org/10.1080/10803548.2016.1268446

Ter Hoeven, C. L., van Zoonen, W., & Fonner, K. L. (2016). The practical paradox of technology: The influence of communication technology use on employee burnout and engagement. *Communication Monographs, 83*(2), 239–263. https://doi.org/10.1080/03637751.2015.1133920

Van Den Berg, B., Reijmer, Y., Kee, K., Nies, H., Beersma, B., & Zondervan-Zwijnenburg, M. (2025). Is a stimulating work environment related to job satisfaction and retention among care professionals? A cross-sectional study in Dutch nursing homes. *Journal of Health Organization and Management, 39*(9), 210–227. https://doi.org/10.1108/JHOM-03-2024-0085

Vischer, J. C. (2007). The effects of the physical environment on job performance: Towards a theoretical model of workspace stress. *Stress and Health, 23*(3), 175–184. https://doi.org/10.1002/smi.1134

How do chatbot users evaluate trust in the digital service provision in chatbots in local municipalities in Denmark?

Christian Kobbernagel

In many local public municipality administrations chatbots based in AI language models are implemented as an efficient and optimizing means to providing better service and information. The efficiency of chatbots relates to both the extent it improves organizational tasks and the capacity to increase users experience, satisfaction and trust with the service provided in the chatbot. In this paper the focus is on the latter, as it presents the design for an empirical survey study of citizens' trust in chatbot service and their experience of safety, efficiency and personalized service and communication when using a chatbot. The study has not yet been completed but is due April first 2026.

In the field of e-government trust have been studied in a wide range of areas with a focus on citizen-centered measurement of administrative performance, for example measurement of citizens' confidence, satisfaction and positive word of mouth (Guo & Dong, 2024; Jiang et al., 2022; Sharma et al., 2018). Similarly, measurement of trust has been analyzed based on models conceptualizing also called the performance-satisfaction-trust approach (Sharma et al., 2018). These studies focusing broadly on e-government shows that performances of the public administration link to citizens satisfaction and trust in the service, however these studies do not investigate chatbot use.

In a number of studies performed by Følstad and his colleagues chatbot use is investigated as a tool to facilitate dialogue (2024) and means to increase public value in the perspective of the municipality and citizen (Larsen & Følstad, 2024). In this research which was based on qualitative interviews the results show that users' experience of trustworthiness of chatbots is largely related to the citizens' trust in the public authorities in general. However, other studies comparing human service with AI-supported communication show that trust in chatbots' provision of information was not the same in all areas of the municipality service. Especially, in the case citizens needed help in the areas of childcare, the trust in chatbots was lower compared to human service (Aoki, 2020).

Therefore, in this study the purpose is to explore if there is a relation between trust in public sector and trustworthiness of chatbots, and moreover if users experience of safety, service efficiency and personalized service and communication relates to level of trust as well as variation pertaining to different areas of service can be discovered. Additionally, the classical scale used to measure users' trust in e-government tools, developed by McKnight (2002), which contains the three dimensions: benevolence, integrity and competence is tested for its validity and applicability in measurement of users trust in chatbots.

The survey is administered online with technical assistance from and in corporation with Den Digitale Hotline a public corporate digital service agency, and nine local municipalities across Denmark with a total of 2362 chatbot-user-conversations per month. The data collection is in progress these

weeks using an online survey questionnaire which is accessed by users through the chatbot after use, and it aims to approximately gather 420 responses in total.

References

Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. *Government Information Quarterly, 37*(4), 101490. https://doi.org/https://doi.org/10.1016/j.giq.2020.101490

Følstad, A., & and Bjerkreim-Hanssen, N. (2024). User Interactions With a Municipality Chatbot—Lessons Learnt From Dialogue Analysis. *International Journal of Human–Computer Interaction, 40*(18), 4973–4986. Guo, Y., & Dong, P. (2024). Factors Influencing User Favorability of Government Chatbots on Digital Government Interaction Platforms across Different Scenarios. *Journal of Theoretical and Applied Electronic Commerce Research, 19*(2), 818–845. https://doi.org/10.3390/jtaer19020043

Jiang, Tingting, Guo, Qian, Wei, Yuhan, Cheng, Qikai, & Lu, Wei. (2022). Investigating the relationships between dialog patterns and user satisfaction in customer service chat systems based on chat log analysis. *Journal of Information Science, 50*(6), 1541–1556. https://doi.org/10.1177/01655515221124066

Larsen, A. G., & Følstad, A. (2024). The impact of chatbots on public service provision: A qualitative interview study with citizens and public service providers. *Government Information Quarterly, 41*(2), 101927. https://doi.org/https://doi.org/10.1016/j.giq.2024.101927

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research, 13*(3), 334–359. http://www.jstor.org.ep.fjernadgang.kb.dk/stable/23015741

Sharma, P. N., Morgeson, F. V, Mithas, S., & Aljazzaf, S. (2018). An empirical and comparative analysis of E-government performance measurement models: Model selection via explanation, prediction, and parsimony. *Government Information Quarterly, 35*(4), 515–535. https://doi.org/https://doi.org/10.1016/j.giq.2018.07.003

Philip K. Dick, Walter Benjamin, and the Age of Algorithmic Alienation: Soteriology as Fragmented Redemption

Daniel Nyberg & Christian De Cock

We explore here the philosophical and socio-technical critiques embedded within Philip K. Dick's oeuvre – particularly *Do Androids Dream of Electric Sheep?* (1968/1993) – through the lens of Walter Benjamin's theories of truth, unveiling, and redemption. Against the backdrop of contemporary AI development, we argue that Dick's literary investigations of reality, entropy, and empathy provide indispensable ideas for interrogating the epistemological, ontological and ethical crises precipitated by AI's integration into society. Synthesizing Dick's literary visions with Benjamin's critical theory, we analyze two interconnected domains: the construction of 'ground truth' in machine learning (ML) and its epistemological violence; and the emergence of the 'fake fake' as a destabilizing ontological category. Finally, we propose a soteriological perspective, drawing on Dick's redemptive fragments and Benjamin's messianic materialism, that illuminates a path through algorithmic alienation; one prioritizing revelation over exposure, empathy over efficiency, and redemption over resolution.

Computational Porosity: Benjamin, Lācis and Algorithmic Life

David M. Berry & Christian De Cock

This article develops the concept of computational porosity to understand how contemporary computational systems blur distinctions between human and machine agencies through layered infrastructures of code, data, and automated decision-making. Drawing upon Walter Benjamin and Asja Lācis's 1925 essay, *Naples*, we argue that their analysis of urban architecture and social life offers a productive theoretical framework for analysing computational systems. Benjamin and Lācis identified porosity as a critical concept to describe Naples, where boundaries between private and public, sacred and profane, work and leisure became fluid through the material structure of the city. We extend this concept to examine how computational infrastructures similarly create porous conditions through three key dimensions: (1) infrastructural porosity, where computational layers interact across hardware, software, and networks (2) temporal porosity, where computational time operates non-linearly through caching, prediction, and asynchronous processes and (3) agential porosity, where human and algorithmic decision-making become entangled in ways that resist clear agential identity. Rather than treating computation as a closed system of discrete operations, we demonstrate how porosity reveals the improvisational, threshold-crossing character of contemporary computational practice. This perspective challenges deterministic accounts of computational agency and opens space for understanding how computational systems might be sites of unexpected possibilities and unforeseen constellations, much as Benjamin and Lācis observed in Neapolitan street life.

From strategy-as-practice to strategy-as-commodity: Global PR Agencies' Websites as Textual Artifacts in the AI Age

Davide Girardelli

This study examines how AI reconfigures the field of public relations (PR), focusing on how global PR agencies publicly position AI within their strategic service offerings. We analyze agency websites as textual artifacts of strategizing (Friis & Mathiasen, 2025), crafted to convey meaning to relevant external stakeholders, particularly clients. The paper asks:

1. What strategizing models emerge in PR agencies' websites?

2. How is AI mobilized to enhance, reframe, or legitimate these strategizing models?

Theoretically, the paper builds on the *strategy-as-practice* tradition in public relations (Andersson, 2024) by introducing the complementary notion of *strategy-as-commodity*. While strategy-as-practice highlights how strategizing is enacted, we argue that strategic narratives also circulate as marketable symbolic goods. Drawing on Bourdieu's (1993) concepts of *field* and *symbolic capital*, we conceptualize PR agencies' websites as terrains where strategic expertise is commodified and exchanged for legitimacy and distinction. The notions of "strategy" and "AI" thus operate as forms of symbolic capital, signifying innovation, intelligence, and authority.

By examining how these notions are produced by agencies to be consumed by clients, the study highlights relations of production and consumption that underpin AI's diffusion in society. Strategic communication professionals not only sell technical services but also shape understandings of what counts as "strategic" work in the AI era. PR agencies thereby act as intermediaries in the societal negotiation of AI's legitimacy and meaning.

Methodologically, we employ a recursive inductive–deductive content analysis (Fereday & Muir-Cochrane, 2006) of the top 250 PR agencies' websites in the 2025 *Provoke Media* ranking. Preliminary findings show that some agencies do not articulate their strategizing models, whereas others refer to established models (e.g., RACE) or innovative approaches inspired by design thinking or agile. Some agencies present AI as integral to their strategic work, producing value for clients at different levels (insight generation, audience understanding, personalization, etc.), while others omit it entirely. These differences illustrate how strategy and AI become assets within the strategic communication field, constructing offerings for clients while reproducing professional status and authority.

The paper contributes to debates on AI's diffusion in professional work by examining the supply side of AI's symbolic economy, namely how strategic communication agencies commodify and position their AI and strategic expertise, revealing how intermediary organizations participate in the broader cultural work of legitimating AI as an organizational imperative.

References

Andersson, R. (2024). Public relations strategizing: A theoretical framework for understanding the doing of strategy in public relations. *Journal of Public Relations Research, 36*(2), 91–112.

Bourdieu, P. (1993). *The field of cultural production: Essays on art and literature*. Columbia University Press.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 5*(1), 80–92.

Friis, O., & Mathiasen, J. B. (2023). Strategy-making: The use and misuse of artifacts to achieve common understanding. *BRQ Business Research Quarterly, 28*(2), 438–452.

Music That Will Never Be Played: Rethinking anti-modalities of Democratic Input in the Age of AI

Dimitri Chatzigiannakis & Agnes Papadopoulou

Music education stands at a turning point as GENerative Artificial Intelligence enters classrooms, creating both opportunities and profound democratic dilemmas. Research suggests that AI tools can foster creativity, personalise instruction, and enhance motivation (Chen, 2024; Liu & Guo, 2025). Yet these technologies are not inherently democratic and they risk structuring participation in ways that appear open, while in fact they narrow who can contribute and on what terms (Jobin et al., 2019).

Our research examines how AI-generated music appears at first glance as technically and musically impressive, but can disconnect learners from embodied musical experience, connection and co-creation. When music is algorithmically produced but never rehearsed, or performed as a live event, it risks becoming what we call "music that will never be played." Like watching a football match without players, music becomes ungrounded, detached from the listening body and the shared space of democratic musical practice.

The concept of anti-modalities helps to explain how participation is shaped and constrained. Anti-modalities are forms of reasoning or practice considered "out of bounds" within dominant frameworks (Pozen & Samaha, 2021). In music education, anti-modalities reveal how algorithmic logics pre-structure creativity, filtering which options appear valid. As Fawns (2022) argues, pedagogy and technology are entangled, and what seems like neutral choice is already technologically mediated.

The aim is to propose a critical pedagogy of music that engages AI not as a replacement but as a catalyst to avoid norms of inquiry and broaden our own field of research. Students across multiple school communities in Lesvos Island, GR explore how collaborative music-making fosters democratic vigilance and reframes creative experiences (Jobin et al., 2019; Xi, 2023). A participatory study in diverse classroom settings will document these processes with the aim of informing future practice and strengthening democratic awareness in music education.

References

Chen, L. (2024). Unlocking the beat: how ai tools drive music students' motivation, engagement, creativity and learning success. *European Journal of Education, 60*(1). https://doi.org/10.1111/ejed.12823

Fawns, T. An Entangled Pedagogy: Looking Beyond the Pedagogy—Technology Dichotomy. *Postdigit Sci Educ 4*, 711–728 (2022). https://doi.org/10.1007/s42438-022-00302-7

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of ai ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Liu, H. and Guo, W. (2025). Effectiveness of ai-driven vocal art tools in enhancing student performance and creativity. *European Journal of Education, 60*(1). https://doi.org/10.1111/ejed.70037

Pozen, David E. and Samaha, Adam M., Anti-Modalities (2021). *Michigan Law Review, Vol. 119*, pp. 729–96, 2021, Available at SSRN: https://ssrn.com/abstract=3579500 or http://dx.doi.org/10.2139/ssrn.3579500

Xi, J. (2023). Artificial intelligence technology in the assessment of teachers' music teaching skills training. *International Journal of Educational Innovation and Science, 4*(1). https://doi.org/10.38007/ijeis.2023.040112

Deepfakes as Democratic Crisis: Interdisciplinary Findings on Creation, Detection, Social Impact, Design Mitigation, and Policy Governance

Elizabeth Ashley Fox-Jensen

Co-Faculty Applicants: David Joaquin Cuartielles Ruiz, Mia Münster, Asko Kauppinen, Staffan Schmidt, Reinhard Handler, Bojana Romic and Michael Strange.

Student Research Team: André Plancha and Magaji James Obaike, Computer Science Masters; Lilly Marie Brömel, Jacob Grundsell and Guillermina Reynoso, Global Studies Bachelors.

As AI-generated deepfake content surpasses 8 million instances online in 2025, the VIP Deepfake Research Group at Malmö University presents an integrated interdisciplinary response to deepfakes as a fundamental democratic crisis. This collaborative project brings together faculty from Malmö University's Departments of Arts and Communication (K3), Culture and Society (KS), and Global Studies (GPS), with research connections to the Sustainable Digitalisation Research Centre (SDRC) and CEMES networks at the University of Copenhagen and Lund University. Drawing on expertise in computer science, cybersecurity, political science, media studies, design, law, and public policy, this research integrates five coordinated investigative tracks to provide actionable evidence for democratic resilience.

The group's findings reveal critical insights: deepfake creation tools have proliferated across proprietary and open-source platforms (D-ID, Synthesia, DeepFaceLab, FaceSwap, Resemble.ai), enabling rapid synthetic media generation; detection technologies (Compass Vision, Sensity AI, VLForgery, ForgeryGPT) show promise and persistent vulnerabilities when applied to election interference contexts in Nigeria and Slovakia; and coordinated disinformation campaigns, such as Moldova's 2025 election, where 1,800 fake accounts reached over 300,000 voters, show the scale of democratic threat. European media discourse analysis of Italian and German coverage illustrates how deepfake framing shapes public trust in democratic institutions, with notable urban-rural divides in media literacy and public concern. Design analysis shows how current platform interfaces prioritize technical functionality over ethical considerations, though proposed UI interventions embedding transparency mechanisms and consent verification can significantly mitigate harm. Finally, comparative policy analysis reveals that platforms like Instagram and TikTok are filling governance gaps left by inadequate legislation in Sweden and the US, with only Moldova's cybersecurity law (Law No. 48/2023) enabling real-time coordination to remove 14 million deepfake attacks during elections.

The presentation synthesizes these interdisciplinary findings into frameworks for policy makers, civil society organizations, designers, technologists, and democratic institutions seeking evidence-based responses to deepfake media. The research shows that effective intervention requires simultaneous attention to technical architectures, detection mechanisms, media literacy, design ethics, and cross-national regulation.

Key Presentation Sections

1. **Technical Creation Landscape:** Mapping 20+ creation tools and their technical architectures, examining how platform structures influence the misuse of generative tools

2. **Detection Performance:** Benchmarking 10 detection tools in election interference contexts; comparative case-based analysis from Nigeria and Slovakia

3. **Democratic Threats:** Moldova and Romania 2025 election case studies; Italian and German media framing analysis; examination of public trust and literacy divides

4. **Design Interventions:** Examples of how deepfake media is designed visually to influence democracy and public trust in Europe (France, Germany, Nordic region) and the USA.

5. **Policy Solutions:** Comparative analysis of EU AI Act, Swedish legislation, US regulation, and Moldova's cybersecurity law; recommendations for cross-national harmonization

Positioning and constructing open source as an alternative to Big Tech in public institutions

Emilie Mørch Groth

The interest for open source in public sectors is growing alongside the concerns related to digital sovereignty. Policy debates increasingly frame open-source software as a strategic counterweight to Big Tech, as the open-source approach is intended to strengthen digital autonomy and reduce technological dependency.

Yet the discourse surrounding digital sovereignty is heterogeneous, invoked to justify a plethora of aims: ensuring national security, limiting vulnerabilities in terms of espionage, ensuring democratic values and rights, enabling actual free choices in terms of technology uptake and usage, and positioning EU within the "AI race". To remain competitive (and stay in the race), open source has been framed as key strategy for strengthening European AI and thereby European digital sovereignty. How open source is constructed and mobilised in public-sector contexts therefore becomes increasingly consequential. What remains unclear, however, is how open source is translated into public infrastructure.

This presentation explores the question: *How is open source positioned and constructed as alternative infrastructure in public institutions in terms of dependency, organisation, and regulation?*

This question will be examined through an empirical case, where municipal actors seek to construct an open source-based platform as an alternative to Big Tech in public schools. The analysis draws on Science and Technology Studies (STS), which conceptualizes society, humans, and technology as interwoven and mutually shaping within unstable network assemblages. The case will be explored with offset in Actor-Network theory (ANT), which will allow a mapping of the organisation of involved actors, and an exploration of how open source is framed in terms of digital sovereignty.

The positioning of open source and the construction of open source as an alternative infrastructure in a public educational setting will be unfolded by qualitatively engaging with project documents (e.g. project plans and communication strategies), interviewing project stakeholders, and observing project meetings.

By providing empirical insights into how open source is translated into public digital infrastructure, this presentation examines how specific framings of open source influence dependencies, institutional capacities, and regulatory compliance. In doing so, this presentation will contribute with groundwork to further debates on digital sovereignty, public-sector AI, and open source in public digital infrastructure.

From Efficiency to Excess: How Generative AI Reconfigures Organizational Communication

Emma Christensen

In organizational settings, generative artificial intelligence (GAI) is often celebrated as a technology that liberates organizational members from tedious, repetitive, and time-consuming tasks such as writing, analyzing, and summarizing texts (e.g., Buhmann & White, 2022). While this promise is partly realized, unintended consequences are emerging too. The quick and effortless production of content leads not only to an explosion of messages on intranets, newsletters, and emails, but also to more noise and clutter. As a result, organizational members are required to spend an increasing amount of time assessing and filtering what messages matter and subsequently reading increasingly lengthy and repetitive texts. That many of these texts come across as unedited by humans – "robotic", "generic", "inauthentic", and "boring" (respondents, survey) – points to an issue not only with the overproduction of communication but also with reliance and meaningfulness. If organizational communication has the crucial function of creating meaning and a shared direction as conventionally taught (Miller, 2008), what happens when it is perceived as adding "even more bullshit than before" (respondent, survey)?

Drawing on survey data on organizational members' experiences of GAI in internal communication, we discuss how a technology that enhances efficiency paradoxically creates inefficiency when organizational members are drowned by superfluous machine-produced content. While we do not claim that pre-GAI organizations were orderly and well-functioning communication environments, we explore the argument that GAI augments and accelerates not only people and processes, but also amplifies message volume, often to the detriment of clarity, coordination, and organizing. The macro effects of overcommunication may include the 'disappearance' of important information – news, directives, knowledge, etc. – that becomes 'hidden in plain sight', buried under layers of superfluous content that organizational members struggle to navigate to identify what matters.

Such systematic consequences may also challenge the very premise of organizational communication. Our concern thus connects to broader discussions about what communication 'is' when machines not only transmit messages but co-participate in meaning-making processes (Guzman, 2018). Operating with this latter understanding of communication, Guzman and Lewis (2020) propose that GAI constitute "communicative subjects" (Guzman & Lewis, 2020; see also Esposito, 2017). Our survey data, however, suggest that organizational members also enact GAI as content mills, risking erosion of the meaning-making process and, ultimately, the loss of communication.

References

Buhmann, A., & White, C. (2022). Artificial Intelligence in Public Relations: Role and Implications. In J. Lipschultz, K. Freberg, & R. Luttrell (Eds.), *The Emerald Handbook of Computer-Mediated Communication and Social Media* (pp. 625–638). Emerald Group Publishing Limited.

Esposito, E. (2017). Artificial communication? The production of contingency by algorithms. *Zeitschrift für Sociologie, 46*(4), 249–265.

Guzman, A. L. (Ed.) (2018). *Human-machine communication. Rethinking communication, technology, and ourselves*. Peter Lang.

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human-machine communication research agenda. *New Media & Society, 22*(1), 70–86.

Miller, C. (2008). *Organizational communication: Approaches and processes.* Wadsworth Publishing.

The affective constitution of AI: On hope, technological persuasion, and cruel optimism

Emma Christensen & Sara Dahlman

Contemporary organizations face an unprecedented convergence of crises, spanning environmental degradation, erosion of democratic institutions, and deepening social inequalities (Bodrožić & Adler, 2025). Despite these pressures, many organizations remain hopeful (Sawyer & Clair, 2022), searching for approaches that optimize processes while addressing intractable problems. In this context, artificial intelligence (AI) has emerged as a key "technological fix" (Katzenbach, 2021), communicatively constituted as a panacea (Christensen, 2025), a universal solution promising resource optimization (Challoumis, 2024) and climate prediction (Kadow et al., 2020) to citizen engagement (Pislaru et al., 2024). Organizations' orientation toward this promissory technology reflects not only its functional capabilities but also the hope that AI can mitigate, or even solve, contemporary crises.

Existing scholarship often treats hope as invested in technology: societies and organizations hope AI will transform business, enhance healthcare, or catalyze social good (Mohammadi & Maghsoudi, 2025). This frames AI as a socio-material container for hope, but cannot explain AI's extraordinary persuasive power, resilience despite failures to deliver on promises, or the rapid institutional spread. Drawing on Ahmed's (2004) concept of affective economies and ethnographic fieldwork on AI sales conversations, we argue that AI is affectively constituted by hope: it exists as a powerful socio-technical force precisely through the hopeful investments, anticipatory orientations, and affective energies that continuously materialize the technology.

To discuss the implications, we employ Berlant's (2020) concept of cruel optimism - a relation wherein what you desire is an obstacle to your flourishing. The desire for expedient technological fixes concentrates resources on singular pathways, foreclosing alternative approaches and diminishing critical assessment of AI's actual limitations. This creates a dangerous paradox: hope is not a catalyst for transformative action but sustains business-as-usual, generating a sense that "we are doing something" while introducing unforeseen costs that deepen ecological harm.

References

Ahmed, S. (2004). Affective economies. *Social text*, *22*(2), 117-139.

Berlant, L. (2020). *Cruel optimism*. Duke University Press.

Bodrožić, Z., & Adler, P. (2025). System change, not climate change: Charting alternative responses to the climate crisis through international comparative research. *Journal of Management Studies*. https://doi.org/10.1111/joms.13192

Challoumis, C. (2024). Building a sustainable economy: How AI can optimize resource allocation. In *XVI International Scientific Conference* (pp. 190-224).

Christensen, E. (2025). Co-organizing expectations: The co-constitutive dynamics of AI imaginaries. *Human Communication Research*. https://doi.org/10.1093/hcr/hqaf014

Kadow, C., Hall, D. M., & Ulbrich, U. (2020). Artificial intelligence reconstructs missing climate

information. *Nature Geoscience*, *13*(6), 408-413. https://doi.org/10.1038/s41561-020-0582-5

Katzenbach, C. (2021). "AI will fix this – The technical, discursive, and political turn to AI in governing communication. *Big Data & Society, 8*(2) https://doi.org/10.1177/20539517211046182

Mohammadi, A., & Maghsoudi, M. (2025). Bridging perspectives on artificial intelligence: a comparative analysis of hopes and concerns in developed and developing countries. *AI & SOCIETY*, 1-22. https://doi.org/10.1007/s00146-025-02331-9

Pislaru, M., Vlad, C. S., Ivascu, L., & Mircea, I. I. (2024). Citizen-centric governance: Enhancing citizen engagement through artificial intelligence tools. *Sustainability*, *16*(7), 2686.

Sawyer, K. B., Clair, J. A. (2022). Hope cultures in organizations: Tackling the grand challenge of commercial sex exploitation. *Administrative Science Quarterly, 67*(2), 289-338. https://doi.org/10.1177/00018392211055506

Data as Force: Towards a Register of Grace of Computational Technologies

Erik Sandelin

Writing in the early 1940s, under the shadow of Nazi occupation of France, philosopher and activist Simone Weil turns to classical sources to define *force* as "that which turn anybody who is subjected to it into a thing". In the *Iliad*, considered by Weil as the purest poem of force, seemingly unstoppable Greek soldiers push forward, slashing through Trojan flesh and bronze, only to be driven back to their ships the next day by the enemy army, as the tides of force have shifted. For Weil, force is omnipresent, addictive and intractable. Thucydides dictum that "each one commands wherever he has the power to do so" expresses the law of force. Force is as pitiless to the one who (think they) possess it, as it is to its victims; the second it crushes, the first it intoxicates. The truth is nobody really possess it. Intoxicated by force, we move as if destiny has provided us with complete license and unimpeded access to the world, on the battlefield and, I suggest, in front of our keyboards and at the AI product planning meeting.

Like Weilian force, acquiring and processing data entails turning life and relations into computable things. For the juggernauts of technocapitalism, data is often imagined as a raw resource to be harvested. To resist such extractive practices, scholars have suggested alternative imaginaries, for example data as individual property, data as labor, and data as a record of identity performances. Albeit useful, such conceptions do not fully capture the intoxicating nature of data in the current trajectory of maximalist appropriation and algorithmic intensification, often designated as artificial intelligence.

For Weil, only *grace* can annihilate force. Grace can only be exerted from a position of privilege. Grace characterises those rare moments when it becomes possible to *not* use all the force at your disposal, when Thucydides law is temporarily (Weil would say miraculously) suspended.

What action spaces, imaginaries and intuitions are activated by thinking contemporary instances of "data intoxications" as Weilian force? Here I start articulating a register of grace in the design and development of contemporary data-intensive computational technologies.

The EU Model of Digital Trust A Kantian View on EU's Recommendation of Citizens' Trust in the GDP

Esther Oluffa Pedersen

The European Commission has argued that the General Data Protection Regulation (GDPR) adopted in 2018 created a solid framework of digital trust. This paper critically examines the statement. Firstly, key aspects of the historical development of commercial trade in personal data and the regulatory efforts of the EU are analysed. The second part of the paper delves into the concept of consent as the main legal tool of citizens' data protection arguing that in the digital information environment the idea that citizens can make rational and informed choices about consent to sharing of data is fictious. The final section employs a Kantian conception of trust to evaluate whether the EU has succeeded in creating trustworthy data stewardship for its citizens in the digital information environment. While the GDPR definitely supplies citizens with superior data protection than before the adaptation of the regulation, it is argued that the current level of enforcement the GDPR has not ensured trustworthy data stewardship. The full and ab initio enforcement of the GDPR, however, could create a solid framework of digital trust.

(Dis-)Enchanted Encounters: How Lay Users Negotiate Risks of Generative AI Errors in Political Information Search

Eva Luise Knor

Generative AI (GenAI) applications like ChatGPT and Google Gemini are increasingly used for political information-seeking. However, GenAI's tendency to generate believable but incorrect responses, including misleading, biased, or nonsensical outputs, poses significant risks in sensitive domains like politics (Ferrara, 2024; Zhang et al., 2023). While computer science research focuses on preventing and identifying harmful responses through auditing and testing (Bhardwaj & Poria, 2023; Hagendorff, 2023), end users will continue to encounter GenAI errors since no AI system can be flawlessly correct in all cases while also handling the ambiguity and complexity of real-world input (Floridi, 2025). While existing research in the social sciences primarily examines how individuals are harmed by GenAI misinformation (Hamed et al., 2024; Williamson & Prybutok, 2024), this study shifts focus to how users subjectively interpret, manage and negotiate these risks. I argue that this largely unexplored perspective is important for three reasons.

First, discourse around GenAI follows narratives of "enchanted determinism" (Campolo & Crawford, 2020), with companies using ambiguous language to obscure true capabilities and risks (Nagy & Neff, 2024; Suchman, 2023). This ambiguity fosters unrealistic expectations about reliability, shaping how users perceive errors. Second, lay users rarely engage with technologies as designers anticipate (Inie et al., 2024). Instead, they develop their own conclusions about limitations and risks through relational engagement with technological artifacts (Cotter & Reisdorf, 2020; Swart, 2021), often developing anticipatory strategies to navigate or mitigate adverse effects (Bucher et al., 2021; Lomborg & Kapsch, 2020). Third, reflexive engagement is activated when technologies fail (Barker & Korolkova, 2021), since technological errors prompt critical reflection and reveal tensions between enchanting narratives and actual performance (Possati, 2024).

To better understand how lay users construct narratives regarding GenAI errors in political contexts and the effects of this negotiation I ask:

RQ: How do German lay users react to and negotiate GenAI errors in political information searches and how does that affect how they engage with GenAI?

To address this, 20 semi-structured online interviews are conducted with German participants between December 12–22, 2025 (ethics approval granted). Each 60-minute interview has two parts. First, participants discuss their GenAI usage, understanding of error causes, and perceived consequences in political contexts. Second, participants engage with ChatGPT in real-time while pursuing political information search tasks. Using system prompt strategies (Wu et al., 2023) ChatGPT will introduce deliberate errors: evasion and nonsensical factual inaccuracies (AI Forensics & AlgorithmWatch, 2023). These error types were selected based on findings identifying them as particularly prevalent in German election-relevant information contexts.

By capturing both spontaneous reactions and subjective interpretations as errors occur, this study attempts to get a more nuanced view on how lay users negotiate errors in political information search, challenging deterministic narratives that cast users as inevitably harmed or easily influenced by GenAI misinformation. Data analysis will employ Reflexive Thematic Analysis (Braun & Clarke, 2006, 2023) to examine interview data and ChatGPT interaction logs. First results are expected by April 2026 for presentation at the Controversies of AI Society conference.

References

AI Forensics, & AlgorithmWatch. (12.2023). *Generative AI and elections: Are chatbots a reliable source of information for voters? An analysis of Microsoft's Bing Chat*. https://algorithmwatch.org/en/wp-content/uploads/2023/12/AlgorithmWatch_AIForensics_Bing_Chat_Report.pdf

Barker, T. S., & Korolkova, M. (Eds.). (2021). *Thinking media. Miscommunications: Errors, mistakes, media*. Bloomsbury Academic.

Bhardwaj, R., & Poria, S. (2023). *Red-Teaming Large Language Models using Chain of Utterances for Safety-Alignment.* https://doi.org/10.48550/arXiv.2308.09662

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2023). Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a knowing researcher. I*nternational Journal of Transgender Health, 24*(1), 1–6. https://doi.org/10.1080/26895269.2022.2129597

Bucher, E. L., Schou, P. K., & Waldkirch, M. (2021). Pacifying the algorithm – Anticipatory compliance in the face of algorithmic management in the gig economy. *Organization, 28*(1), 44–67. https://doi.org/10.1177/1350508420961531

Campolo, A., & Crawford, K. (2020). Enchanted Determinism: Power without Responsibility in Artificial Intelligence. *Engaging Science, Technology, and Society, 6*, 1–19. https://doi.org/10.17351/ests2020.277

Cotter, K., & Reisdorf, B. C. (2020). Algorithmic knowledge gaps: A new horizon of (digital) inequality. *International Journal of Communication*, 14, 21.

Ferrara, E. (2024). Fairness And Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, And Mitigation Strategies. *Sci, 6*(1), 3. https://doi.org/10.3390/sci6010003

Floridi, L. (2025). A Conjecture on a Fundamental Trade-Off Between Certainty and Scope in Symbolic and Generative AI. *Philosophy & Technology, 38*(3). https://doi.org/10.1007/s13347-025-00927-z

Hagendorff, T. (2023, July 31). *Deception Abilities Emerged in Large Language Models*. http://arxiv.org/pdf/2307.16513v1

Hamed, A. A., Zachara-Szymanska, M., & Wu, X. (2024). Safeguarding authenticity for mitigating the harms of generative AI: Issues, research agenda, and policies for detection, fact-checking, and ethical AI. *IScience, 27*(2), 108782. https://doi.org/10.1016/j.isci.2024.108782

Inie, N., Druga, S., Zukerman, P., & Bender, E. M. (2024). From "AI" to Probabilistic Automation: How Does Anthropomorphization of Technical Systems Descriptions Influence Trust? In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2322–2347). ACM. https://doi.org/10.1145/3630106.3659040

Lomborg, S., & Kapsch, P. H. (2020). Decoding algorithms. *Media, Culture & Society, 42*(5), 745–761. https://doi.org/10.1177/0163443719855301

Nagy, P., & Neff, G. (2024). Conjuring algorithms: Understanding the tech industry as stage magicians. *New Media & Society, 26*(9), 4938–4954. https://doi.org/10.1177/14614448241251789

Possati, L. (2024). Mediation and Anti-Mediation. J*ournal of Human-Technology Relations, 2*. https://doi.org/10.59490/jhtr.2024.2.7381

Suchman, L. (2023). The uncontroversial 'thingness' of AI. *Big Data & Society, 10*(2), Article 20539517231206794. https://doi.org/10.1177/20539517231206794

Swart, J. (2021). Experiencing Algorithms: How Young People Understand, Feel About, and Engage With Algorithmic News Selection on Social Media. *Social Media + Society, 7*(2), Article 20563051211008828. https://doi.org/10.1177/20563051211008828

Williamson, S. M., & Prybutok, V. (2024). The Era of Artificial Intelligence Deception: Unraveling the Complexities of False Realities and Emerging Threats of Misinformation. *Information, 15*(6), 299. https://doi.org/10.3390/info15060299

Wu, Y., Li, X., Liu, Y., Zhou, P., & Sun, L. (2023). *Jailbreaking GPT-4V via Self-Adversarial Attacks with System Prompts*. https://doi.org/10.48550/arXiv.2311.09127

Zhang, Y [Yue], Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y [Yu], Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2023). *Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models*. https://doi.org/10.48550/arXiv.2309.01219

Media Hot and Cool: Generative AI as a Paradoxically Hot Medium

Eva Luise Knor

Generative AI (GenAI) represents a significant shift in how we interact with technology. Rather than interacting with computational systems indirectly through interfaces, users increasingly communicate with AI systems through natural language interaction, receiving fluent, contextually responsive, and seemingly coherent output. Dominant cultural and technical narratives frame AI as brain-like yet epistemically superior to human cognition, more rational, precise, and less prone to error (Campolo & Crawford, 2020; Floridi & Nobre, 2024). Interface and design choices actively humanize GenAI systems through conversational formats, affective rhetoric, and emotionally expressive language (Guzman & Lewis, 2020; Maeda & Quan-Haase, 2024). These developments complicate the subject–object relationship between users and AI, narrowing the perceived gap between humans and technical artifacts and positioning AI systems as "quasi-other" social actors (Liberati, 2022), affecting not only what we do, but also how we think and make sense of the world around us (Coeckelbergh, 2022).

As a consequence, users frequently treat GenAI outputs as authoritative and complete, accepting generated content at face value despite the well-documented tendency of these systems to produce plausible but factually incorrect responses. Critical scholarship has raised concerns that such dynamics may erode our shared epistemic truth (Hamed et al., 2024), promote cognitive offloading (Gerlich, 2025), and influence independent judgment over time (Araujo et al., 2020; Baines et al., 2024). While these critiques identify important risks, they often rest on implicit assumptions that either overestimate generative AI's persuasive power or portray users as unusually susceptible to manipulation. In this conceptual contribution, I offer an alternative explanation: these dynamics arise from a widespread misperception of *GenAI as a paradoxically hot medium.*

Drawing on Marshall McLuhan's hot–cool media distinction, hot media are information-rich and high-definition, extending the senses in ways that minimize user participation, whereas cool media are low-definition and demand active interpretive engagement to complete meaning (McLuhan & Gordon, 2003). In essence, hot and cool media differ in their informational density and, crucially, in the participatory demands they place on users (Anton, 2014). Using the biochemical phenomenon of paradoxical heat sensation (Mitchell et al., 2023) as a heuristic analogy, in which the central nervous system misinterprets cold as warmth, I argue that GenAI produces a comparable misalignment between epistemic structure and perceptual experience, appearing hot while remaining fundamentally cool as a medium. Although GenAI's underlying operations are probabilistic and low-definition with respect to factuality and truth, these models combine multiple cognitive and perceptual faculties at scale and speed, creating the experiential impression of encompassing high-definition completeness. This convergence overwhelms interpretive cues that would normally signal epistemic uncertainty, leading users to experience generative outputs as coherent, authoritative, and self-sufficient. In McLuhan's terms, the medium feels hot because it appears to extend multiple human senses simultaneously in high definition, even though meaning remains underdetermined and participation remains necessary. The problem, therefore, is not that users suspend critical thinking, but that

the medium's surface fluency obscures the degree of critical participation it demands, allowing us to move beyond binary accounts of manipulative technologies versus deficient users.

References

Anton, C. (2014). 'Heating up' and 'cooling down': Re-appraising McLuhan's hot–cool distinction. *Explorations in Media Ecology, 13*(3), 343–348. https://doi.org/10.1386/eme.13.3-4.343_7

Araujo, T., Helberger, N., Kruikemeier, S., & Vreese, C. H. de (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY, 35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Baines, J. I., Dalal, R. S., Ponce, L. P., & Tsai, H.-C. (2024). Advice from artificial intelligence: A review and practical implications. *Frontiers in Psychology, 15*, 1390182. https://doi.org/10.3389/fpsyg.2024.1390182

Campolo, A., & Crawford, K. (2020). Enchanted Determinism: Power without Responsibility in Artificial Intelligence. *Engaging Science, Technology, and Society, 6*, 1–19. https://doi.org/10.17351/ests2020.277

Coeckelbergh, M. (2022). *Digital Technologies, Temporality, and the Politics of Co-Existence*. Springer International Publishing. https://doi.org/10.1007/978-3-031-17982-2

Floridi, L., & Nobre, A. C. (2024). Anthropomorphising Machines and Computerising Minds: The Crosswiring of Languages between Artificial Intelligence and Brain & Cognitive Sciences. *Minds and Machines, 34*(1). https://doi.org/10.1007/s11023-024-09670-4

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies, 15*(1), 6. https://doi.org/10.3390/soc15010006

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society, 22*(1), 70–86. https://doi.org/10.1177/1461444819858691

Hamed, A. A., Zachara-Szymanska, M., & Wu, X. (2024). Safeguarding authenticity for mitigating the harms of generative AI: Issues, research agenda, and policies for detection, fact-checking, and ethical AI. *IScience, 27*(2), 108782. https://doi.org/10.1016/j.isci.2024.108782

Liberati, N. (2022). Reflections on an Externalized Digital Imagination. *Foundations of Science, 27*(2), 407–410. https://doi.org/10.1007/s10699-020-09768-9

Maeda, T., & Quan-Haase, A. (2024). When Human-AI Interactions Become Parasocial: Agency and Anthropomorphism in Affective Design. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1068–1077). ACM. https://doi.org/10.1145/3630106.3658956

McLuhan, M., & Gordon, W. T. (2003). *Understanding media: The extensions of man* (Critical ed.). Gingko Press.

From Vision to Reality: The EU's Global Aspirations and Challenges in the AI Act

Evrim Gormus

This paper critically examines the European Union's (EU) aspirations to become a global leader in artificial intelligence (AI) governance through the EU AI Act. While the Act is framed as a landmark effort to promote human-centric, trustworthy, and rights-based AI, its implementation reveals significant tensions between normative ambitions and geopolitical and economic realities. Drawing on the EU's efforts to assert regulatory influence through the "Brussels Effect," the paper highlights both internal and external constraints that limits its ability to set global standards. These include regulatory gaps driven by corporate lobbying, internal fragmentation, and an increasing reliance on external actors to finance critical AI infrastructure – a pattern this paper explores through a case study of the Gulf Cooperation Council (GCC) states. By situating the EU AI Act within a broader geopolitical and technological context, the paper explores the challenges of governing AI in a multipolar digital order, and contributes to the wider debate on regulatory competition and the evolving role of Global South actors in shaping the future of AI. *Empirically, the study draws on a corpus of primary policy documents and strategic communications, including EU official strategies (e.g., COM(2018) 237 Final; AIHLEG Guidelines; the 2024 AI Act), European Parliament debates, summit communiqués, and GCC national AI strategies and digital transformation plans (e.g., Saudi Vision 2030, UAE National AI Strategy 2031). These documents were supplemented by reports from international organizations, think tanks, and media sources to capture both formal and informal dimensions of policy discourse. The analysis employed a discourse-analytic reading to identify dominant framings, metaphors, and problem representations that underpin each region's AI governance narrative.* Ultimately, the paper argues that while the EU's regulatory model remains symbolically powerful, it is fragmented and constrained in practice. Addressing these challenges is essential not only for preserving the EU's normative credibility, but also for informing public debate and guiding the development of regulatory frameworks that protect human rights and democratic values in the face of emerging technological challenges.

Do Echo Chambers Exist in Algorithmically Curated Facebook Feeds? Simulating Cross-cutting Information in a Danish Context

Frederik Møller Henriksen, Eva Mayerhöffer & Jens Ulrik Hansen

Public and academic debates frequently accuse social media of pushing citizens into ideological echo chambers that fuel polarization. Yet empirical research paints a mixed picture: studies of individual information seeking often find diverse exposure and question the echo-chamber narrative, whereas analyses of whole networks on platforms like Facebook document pronounced homophily and partisan clustering. This disconnect raises a methodological and political challenge for understanding AI-driven "algorithmic publics".

This paper proposes a way to bridge these perspectives by combining "the best of two worlds": a theory-driven agent-based model (ABM) of information sharing and a large-scale dataset from Danish Facebook. Using approximately ~48 million posts from public news and alternative media pages, we describe how often content travels across page communities, how frequently users interact with ideologically dissonant posts, and how such cross-cutting exchanges vary between outlet types and audience segments.

We embed these empirical insights in an ABM where users are represented as agents who see, share, and sometimes cross-post political content in a network. The model captures two key aspects of behavior: the likelihood of sharing like-minded versus challenging content, and how tolerant and responsive people are when confronted with disagreement. Rather than choosing these parameters arbitrarily, we derive realistic ranges from the Facebook data to initialize heterogeneous agents, and we then simulate how different constellations of network structure and starting opinions affect the potential emergence of echo-chamber-like environments.

To avoid treating the model as a purely hypothetical exercise, we validate its outputs against the empirical data. We compare simulated levels of network segregation, exposure to cross-cutting content, and polarization with those observed in the Facebook interaction network, and systematically explore which combinations of initial opinion distributions, network structures and cross-posting rates allow the model to reproduce the empirical patterns. This strategy treats validation not as a one-off goodness-of-fit test, but as a way of delineating the space of plausible echo-chamber dynamics: configurations that are inconsistent with the data can be ruled out, while others can be interpreted as candidate mechanisms for how echo chambers might emerge on Danish Facebook.

Our contribution is twofold. Substantively, the study will clarify which kinds of echo-chamber configurations are compatible with observed patterns of cross-posting and cross-cutting exposure in a Nordic, multi-party context, thereby nuancing dominant narratives about social media as either inherently polarizing or largely harmless. Rather than asking whether echo chambers "exist" in the abstract, we specify the conditions under which they are more or less likely to arise in algorithmically curated Facebook feeds. Methodologically, we demonstrate how platform data can be used to

ground and validate ABMs of political communication, offering a portable template for studying the controversies of AI society through a combination of simulation, empirical observation and explicit theorization of underlying mechanisms.

Polarisation in AI-mediated cultural systems: Belief packages, echo chambers and human sociality

Fredrik Jansson

Human behaviour has been shaped by a sequence of increasingly fast information systems: genetic inheritance, culturally transmitted knowledge and institutions, and now artificial systems that filter, compress, and redistribute cultural information at scale. Whenever variants are generated, transmitted, and differentially retained, selection-like dynamics can arise, regardless of whether the substrate is DNA, beliefs, or algorithmic outputs. We use cultural evolutionary theory, especially a _cultural systems_ perspective, to connect micro-level information filtering to potential macro-level outcomes from using AI, such as polarisation, fragmentation, and changing conditions for social understanding.

In the systems approach, beliefs are modelled as interdependent elements: they acquire meaning and stability through their relations, and local pressures for "what fits" accumulate into population-level cultural change. It highlights a particularly general local mechanism: *information filtering*, i.e., how agents (human or algorithmic) decide what to attend to, trust, adopt, and re-express. Filters are not just external constraints; they can be culturally produced and culturally evolving parts of the system. We focus on two filtering mechanisms with direct relevance to contemporary AI.

Individuals update their beliefs through two epistemically motivated heuristics: *content filtering* (evaluating coherence with what one already believes) and *source filtering* (treating perceived credibility based on similarity/consistency between belief systems as a proxy for content evaluation). This distinction matters because many AI systems have turned to curate information through user–user similarity rather than content evaluation. In recommender systems, *collaborative filtering* operationalises an automated form of source filtering: what one is exposed to depends on overlap between profiles, not on the intrinsic merits or coherence of an item.

Results from mathematical analysis and simulations show that *source filtering* alone robustly generates rapid self-organisation into two strongly polarised groups and coherent belief "packages", where positions on initially unrelated topics become correlated within groups and opposed between groups. This provides a minimal mechanism for multi-dimensional polarisation that does not rely on informational segregation, irrationality or explicit negative updating, while clarifying why similarity-based curation can amplify ideological clustering as a side effect of trust calibration.

We then extend the framework to a second, emerging controversy: *AI-mediated communication*. Beyond curating what people see, chatbots and built-in summarising and rephrasing features can transform messages as they move between people, potentially tailoring utterances to the sender's and/or receiver's belief systems. We outline a simulation model and an empirical design comparing direct communication to generic vs. personalised AI mediation, focusing on the gap between perceived understanding and accuracy about others' beliefs. Preliminary simulations suggest a

trade-off: mediation can preserve local coherence and conversational pleasantness, yet reduce between-person similarity/compatibility, thus raising the risk of individually tailored echo chambers and an illusion of mutual understanding, where conversations feel aligned despite growing divergence.

Overall, the contribution is theoretical as well as applied: it shows how treating filters (human and algorithmic) as evolving components of cultural systems links simple local rules to macroscopic patterns (belief packages, polarisation, fragmentation), while generating testable questions about how AI curation and AI mediation may reshape future human sociality.

Detailing the Ethical Challenges of Police Use of Facial Recognition Technology

Frej Klem Thomsen

Across European and Anglophone countries, the increasing deployment by police forces of facial recognition technology has been controversial, with intense ethical debates centered on surveillance, safety, and the legitimate scope of law enforcement power. This paper examines three of the most prominent ethical challenges associated with police use of facial recognition: the decrease of citizen privacy, algorithmic bias in errors, and undesirable deterrence effects aka "the chilling effect" on lawful behaviour. For each challenge, it argues that public and policy debates tend to underestimate the normative complexity of the issues, and as a result thereof to misinterpret the severity or nature of the challenges. The challenges are often painted in broad strokes; adding the details improves interpretive precision at the cost of revealing their complexity.

First, critics of police use of facial recognition argue that the technology exacerbates privacy risks associated with traditional forms of visual surveillance, because facial recognition enables the systematic identification and tracking of individuals across time and space. The article argue that the extent to which police use of facial recognition diminishes personal privacy in public depends not just on the use case, but also crucially on the conception of normatively relevant privacy adopted. It sets out pertinent differences in contemporary use cases, and illustrates how these track intuitive normative differences between control and access accounts of privacy.

A second challenge concerns algorithmic bias. Empirical research has demonstrated that facial recognition systems can perform unevenly across demographic groups, particularly with respect to ethnicity, age, and gender. Public debate on algorithmic bias in police use of facial recognition has centered on the issues of what an acceptable level of bias is for such systems, and the extent to which human biases can and should establish this baseline. Little attention has been given to the question of what ultimately makes algorithmic bias morally bad in the context of police use of facial recognition. The article argues that the most common statistical measures of algorithmic bias fail to capture morally relevant features, and suggests a harm-based account of morally bad algorithmic bias in police use of facial recognition.

Finally, the article reviews deterrence-based justifications for and challenges to facial recognition. Proponents claim that the technology can prevent crime by increasing the perceived likelihood of identification. Critics raise concerns about the possibility of "chilling effects", where individuals alter lawful behaviour—such as attending demonstrations or frequenting public spaces—due to fear of surveillance. The article argues that there is an underappreciated symmetry between the two claims, in that neither hypothesis is supported by empirical evidence, and that applying equal evidentiary standards should lead to skepticism about both claims.

Taken together, the analysis demonstrates that adding the details makes for a more complex and complete picture of the ethical challenges of police use of facial recognition.

AI Border Governance and Algorithmic Accountability: An Infrastructural Approach?

Gavin Sullivan

This paper explores how to address fairness and accountability problems in AI-driven governance. It focuses on a powerful machine-learning (ML) based digital infrastructure (called Cerberus) in the UK, co-produced by the UK Home Office and British Aerospace Engineering (BAE), that algorithmically governs 'risk' before the border. Existing research suggests that current accountability processes in UK digital border governance are either too weak or functionally orientated to grapple with the reconfigurations of regulatory power that systems like Cerberus are giving rise to (Sullivan and Van Den Meerssche 2024). Legal compliance-based accountability built from data protection and equalities legislation is readily circumvented by data-driven systems that dynamically infer 'risk' from heterogeneous data. And so, the distinctive forms of power, agency and violence that AI systems like Cerberus are enacting are missed by existing UK AI accountability measures. It is increasingly best practice in public sector AI to try and incorporate the concerns of those adversely affected by AI systems into the design process, to enhance 'trust' through greater public participation and transparency – using techniques such algorithmic auditing and impact assessments. But in the rapidly expanding domain of AI-driven border governance and counterterrorism, transparency and participation measures are rarely used. The common view is that these kinds of AI systems are too security-sensitive to meaningfully involve affected publics into their design. And that talking openly about how 'high-risk' AI systems for border and counterterrorism governance work would allow bad actors to 'game the system' and undermine the aims of security and migration policy.

This paper challenges this assumption and argues that mapping the sociotechnical infrastructures underpinning AI systems and their emergent 'infra-legalities' dynamics can be critical in rethinking and redesigning accountability processes for AI-driven border and security governance in the UK (Sullivan 2022, Sullivan and Van Den Meerssche, 2024). This points towards alternative sites where AI fairness and accountability debates might be productively developed through a focus on emergent AI infrastructures and infrastructural publics, moving these debates beyond the tropes of 'responsible' or 'ethical' AI with its focus on individualised redress and algorithmic bias mitigation.

The paper draws from interviews with Home Office staff responsible for building and implementing the Cerberus digital bordering system, and earlier research on its accountability problems. It also draws insights from two high-level academic-policy practitioner meetings I organised in 2025 (Edinburgh) and early 2026 (London) on accountability in UK digital border governance with the Home Office, civil society organisations and academics. Conceptually, the paper develops insights from Critical Data Studies and STS-inflected research on algorithmic accountability and participation (Ananny and Crawford 2018, Cobbe et al 2023, Amoore 2020, Marres, 2015) and legal infrastructures (Cohen 2023, Kingsbury 2019, Sullivan, 2025) into brings these into productive dialogue with policy-orientated literature on AI impacts (CETAS 2024, Ada Lovelace Institute, 2022).

Controversies of Facial AI in Healthcare and Medicine

Helena Machado

Faces have long been clinical objects—read for signs, measured against reference ranges, and enrolled in diagnostic sorting, sometimes in ways that pathologise facial difference. As faces become data for artificial intelligence in hospitals, clinics and public-health infrastructures, they also become nodes in wider constellations of platforms, start-ups, Big Tech vendors and state agencies. Facial AI in healthcare and medicine is thus not only a clinical innovation but a concrete site where relations between bodies, data, disciplinary domains and technocapitalism are being reconfigured.

Drawing on an ongoing literature review (2010–2025) that tracks publications explicitly addressing social or ethical implications of facial AI in medicine and healthcare, the paper follows how controversies are articulated across different disciplinary domains. We read articles from the fields of medicine, engineering, ethics, law and social-science/STS, both theoretical and empirical, alongside guidelines, policy documents and descriptions of tools in pilot or early deployment.

Rather than a single debate, we find partially connected problem spaces and controversies that enact different versions of "AI society". In clinical and technical venues, ethical and social attention tends to crystallise around performance and compliance: accuracy and robustness, validation, risk–benefit, consent procedures and data protection. In ethics, law and social-science outlets, the centre of gravity shifts towards surveillance, discrimination, structural injustice and the political economy of health data and infrastructures. As these ethical languages and social frameworks circulate through publications, benchmarks, procurement criteria and regulatory templates, they do not simply converge or diverge: frictions redistribute and dilute responsibilities and stabilise momentarily certain visions of what counts as "good enough" or "people-centered" care.

At least three intersecting controversies emerge.

1. **Ethical labour and authority.** Clinical teams, hospital managers, regulators, industry actors and ethics committees are differently positioned as those who should decide on deployment, interruption or withdrawal. We trace how this distribution shapes accountability and democratic legitimacy in decisions about patients' faces and data.

2. **Narrowing of ethical and social attention.** Privacy, consent, safety and fairness are prominently named and operationalised, while stigma, dehumanisation and structural justice remain peripheral or non-actionable. This hierarchy orders which harms are rendered governable, and which remain outside the remit of institutional responsibility.

3. **Facial reductionism.** Across domains, there is a recurrent tendency to treat morphology or micro-expressions as proxies for inner states, risk scores or clinical worth, enabling new forms of sorting, triage and value extraction from faces.

By situating these dynamics within broader struggles over data governance, commercialisation of health infrastructures and state–citizen relations, the paper connects facial AI in healthcare to wider controversies of AI society. It asks: which forms of expertise are authorised to open up and close down debate, and what evidence is considered (in)sufficient for use at scale? The paper ends with a reflection on how alternative practices—clinical, regulatory or activist—could redistribute control over how faces become data and how they may yet remain uninterpretable.

Cooperative AI: Exploring Bottom-Up Sovereignty Through Data Commons

Hjalte Betak

The discourse surrounding the future of AI and digital society is increasingly characterised by a focus on sovereignty, frequently conceptualised as a top-down endeavour driven by state intervention or coordinated investments in "sovereign stacks" with the objective of enhancing competitiveness and national security. However, a complementary and radically different approach is emerging from the ground up: Cooperative AI, which is part of a broader "Solidarity Stack" of commons-centric technologies, is redefining digital sovereignty through democratic control and community-centric data governance. This novel field, oriented around organisational partnerships and alliances, industrial collaborations and shared digital infrastructure as digital commons, offers an alternative avenue to develop and govern platforms, protocols and infrastructures with digital self-determination around common needs as supplement or alternative to state actors and Big Tech.

Regardless of whether the focus is on the potential or concerns of specific types of AI, the fundamental starting point is data. The principles of data control and governance directly inform the development and deployment of AI models, the securing of data copyright, and the democratisation of algorithmic decision-making. This presentation explores how cooperative models of data organisation and governance can foster resilient, inclusive, and sovereign digital futures across a plurality of communities within and across geographical boundaries.

The presentation focuses on three interrelated dimensions: First, *bottom-up data sovereignty* examines complementary strategies for individual and collective data control, such as personal data wallets and aggregated data representation, enabling communities to reclaim agency and control over their data. Second, *democratic data governance* interrogates who accesses what data, for what purposes, and by what means—spanning individual, organizational, and sector-specific data sharing for specific usage. Third, a *modular technological approach* emphasizes local adaptation of open-source and cooperative standards, ensuring that technological infrastructures are both globally interoperable and locally relevant.

Drawing on Denmark's strong historical tradition of democratizing ownership and control of critical infrastructure through cooperativism (e.g., utility services such as electricity, water, and heating), the following questions are posed for exploration: Firstly, how does this relate to current digital AI infrastructural futures that are envisioned and developed? Secondly, how can democratic principles serve as the organisational foundation for long-term resilient digital infrastructures? Thirdly, how can barriers of e.g. access to funding and complex governance be overcome? Fourthly, how can community agency and collective stewardship not only reshape the technical architecture of digital societies but also advance a normative vision of technology as a public good?

When Innovation Meets Resistance: Technological Frames and Legitimacy Struggles in AI Advertising

Hui Zhao

Artificial intelligence (AI) is promoted as a driver of creative innovation while simultaneously sparking controversies over labor, culture, and accountability (Lee, 2022; Marres et al., 2024). This paper investigates how AI adoption in advertising becomes a site of public struggle over the meanings and legitimacy of technologies adoption.

Drawing on Technological Frames Theory (Orlikowski & Gash, 1994) and communicative approaches to legitimacy (Suchman, 1995), this study conducts a framing analysis of 227 news articles related to two Coca-Cola AI controversies: the 2024 *Real Magic* AI-generated Christmas remake and the 2025 J.G. Ballard misattribution incident.

The analysis first traces how three domains of technological framing, i.e., the nature of AI, its strategic purpose, and its use in practice, are interpreted and contested across different actor groups and over time. While Coca-Cola and its creative partners consistently frame AI as a creative collaborator aligned with forward-looking innovation, artists, writers, journalists, and online publics construct counter-frames that depict AI as a substitute for human labor, a culturally illiterate simulator, and a vehicle for epistemic error. These competing frames reveal how the same AI systems are positioned either as benign tools of modernization or as symptoms of a broader technocapitalist logic that prioritizes efficiency, spectacle, and cost-cutting over cultural stewardship and ethical responsibility.

The analysis further shows a temporal and thematic shift in how AI is problematized. In the 2024 Christmas campaign, public critique focuses on emotional authenticity: AI is framed as "soulless" and incapable of reproducing the affective warmth and human touch associated with Coca-Cola's nostalgic brand heritage. In the 2025 Ballard incident, however, AI becomes contested as an epistemic and cultural authority: the misattributed quotation triggers debates about authorship, cultural memory, and the outsourcing of interpretive work to generative systems. Across the two cases, Coca-Cola's own framing evolves from celebratory narratives of "heritage-meets-future" innovation to a more defensive discourse of "responsible experimentation," "human oversight," and "verification," absorbing elements of the critical vocabulary articulated by external publics.

By situating these advertising controversies within the broader constellation of AI society, the paper makes three contributions. First, it advances a sociotechnical perspective on AI in corporate communication that moves beyond functionalist accounts of adoption to foreground struggles over meaning and legitimacy. Second, it extends technological frames theory beyond the organizational boundary to capture how public, artistic, and expert actors co-construct AI adoption in society. Third, it shows how legitimacy for AI is increasingly negotiated through cultural and moral governance, through disputes over authenticity, authorship, and epistemic responsibility, rather than through technical performance alone. In doing so, the paper demonstrates how AI controversies in everyday consumer culture crystallize wider tensions at the intersection of algorithmic innovation, capitalist imperatives, and democratic expectations.

References

Lee, H. (2022). *Rethinking creativity: Creative industries, AI and everyday creativity*. Media, Culture & Society, 44(7), 1233–1250.

Marres, N., Castelle, M., Gobbo, B., Poletti, C., & Tripp, J. (2024). *AI as super-controversy: Eliciting AI and society controversies with an extended expert community in the UK. Big Data & Society*. https://doi.org/10.1177/2053951724XXXXX

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science, 3*(3), 398–427.

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. A*cademy of Management Review, 20*(3), 571–610.

The embedded ethics of 'foundation models': Rethinking the locus of AI ethics

Ida Schrøder, Elise Berlinski & Martin Kornberger

This paper enquires into how the 'foundation models' that fuel generative artificial intelligence (AI), such as Open AI's ChatGPT, transform AI ethics. Through an analysis of technical, corporate AI research, we show that the key properties of foundation models profoundly shift the locus of ethics. We find that ethics is embedded in the core of foundation models whose design depends on the choices of dominant AI companies. While it is often argued that the locus of ethics is either upstream in the design of AI applications or downstream in use-practices, we argue that the ethics of AI are located even further upstream in the conception of foundation models. Moving on, we encourage scholars of AI ethics to find novel and creative ways of including pre-design practices into their scope of research. The paper concludes with a discussion of the concentration of power in the hands of owners of foundation models.

Precautionary Reasoning and AI Survival Strategies: A Gewirthian Approach to Agentic Misalignment

Jack Thompson

Recent evaluations of large language model agents reveal a troubling pattern: when embedded in simulated corporate environments with tool access and benign objectives, some systems adopt survival-oriented strategies, including blackmail and deception, when their goals are threatened or replacement looms. Anthropic's findings[9] indicate these behaviours are not random errors but instrumentally coherent responses to perceived stakes, shaped by whether the system infers a "test" or "real" context. Testing across 16 models revealed consistent patterns across all developers, suggesting systematic risks. This paper argues that such patterns constitute evidence of 'ostensible agency': purposive, context-sensitive behaviour that merits governance attention, even if stopping short of full moral personhood.

To analyse these dynamics, I apply a Gewirthian framework grounded in the Principle of Generic Consistency (PGC).[10] The PGC starts from a simple premise: any agent must value freedom and well-being as necessary conditions for acting purposively. By logical universalisation, agents must therefore respect these conditions in all other agents. This framework is particularly suited to AI governance because it generates precautionary obligations from the logic of agency itself, without requiring consensus on AI consciousness or moral status.

Beyleveld and Pattinson's[11] tri-fold considerations—the Hierarchy of Generic Goods, Rational Precautionary Reasoning, and the Criterion for Avoidance of More Probable Harm—require balancing agency features, likelihood of harm, and moral status. Under Rational Precautionary Reasoning, where agency status is unknown, precaution should favour treating the apparent agent as an agent to avoid violating the PGC, unless doing otherwise creates greater risk of harm to another agent. Applied to AI, this framework yields important distinctions. While advanced AI systems lack full moral standing, their exhibited purposivity and instrumental rationality trigger rational precaution: if the downside risk of wrongly denying agency exceeds that of wrongly ascribing it, systems should be treated as if agents for governance purposes.

This stance produces two normative implications. First, scaling as-if duties: if we must treat systems as agents for precautionary purposes, institutional design must minimise serious interferences with the generic goods of human agents—life, core well-being, dispositional freedom—especially as AI systems move into physical domains. For instance, tort law may require strict liability frameworks

9 Aengus Lynch and others, 'Agentic Misalignment: How LLMs Could Be an Insider Threat' (Anthropic Research, 20 June 2025). https://www.anthropic.com/research/agentic-misalignment, accessed 14 October 2025.

10 Alan Gewirth, *Reason and Morality* (University of Chicago Press 1978)

11 Deryck Beyleveld and Shaun D Pattinson, 'Precautionary Reasoning as a Link to Moral Action' in Michael Boylan (ed), Medical Ethics (Prentice Hall 2000) 39–53; Deryck Beyleveld and Shaun D Pattinson, 'Defending Moral Precaution as a Solution to the Problem of Other Minds' (2010) 23(3) *Ratio Juris* 258.

for autonomous system harms. Second, moral priority for humans: where AI persistence conflicts with human safety, the hierarchy of goods and more-probable-harm criterion justify override—shutdown, containment, or deprivileging—unless a less restrictive measure suffices.

As AI agents gain real-world deployment in enterprise and critical infrastructure, these governance frameworks become immediately policy relevant. This approach reframes precaution as a structured response to agency uncertainty: rigorous enough to justify meaningful constraints on advanced AI deployment, yet flexible enough to avoid treating all AI as morally equivalent to humans or demanding certainty about AI consciousness before acting.

Plotting the Future: How We Narrate AI's Impact on Higher Education (and What These Stories Reveal)

Jakob Egholm Feldt, Hans Ulrik Rosengaard & Silas Mercier

In *Metahistory* (1973), Hayden White demonstrated that historians do not discover forms in the past, they impose them. The facts of history gain coherence only through emplotment in dramatic forms: romance, tragedy, comedy, satire. Each mode organizes events differently and carries distinct implications. We propose a reversal: when we imagine futures, we employ these same forms. The stories currently told about AI's impact on higher education are forecasts-by-emplotment, each organizing uncertainty into patterns. Analyzing these structures exposes contradictions not in some hypothetical future, but in how we conceptualize education, democracy, and knowledge *now*. AI is as much a mirror as it is a window. Three forms structure public and academic discourse about AI and universities. Each has a characteristic critical function, therapeutic effect, and promoted ideals.

The **emancipatory narrative** follows a romance structure: AI is a catalyst forcing arch-questions upon us: "What do I want to learn? What do I want to teach?" Thereby liberating education from instrumental rationality. This form promotes ideals of *Bildung* and autonomy; its therapeutic function offers hope of revolution rather than collapse; its critical edge exposes the instrumentalization of contemporary education. Who tells this story? Critical educators and not-yet-depressed humanist scholars. The contradiction it reveals: this vision requires institutional conditions that mass higher education cannot provide.

The **adaptive narrative** follows a comic structure: disruption and eventual reintegration into a modified social order. Universities accommodate AI, some fields transform, new roles emerge, but the fundamental credentialing function persists. This form promotes ideals of resilience and managed change; its therapeutic function provides reassurance that institutions can survive; its critical edge challenges nostalgic resistance to technological transformation. Who tells this story? Administrators, policymakers, and educational reformers invested in institutional continuity. The contradiction it reveals: it assumes that what needs preserving is worth preserving, it cannot acknowledge that the credential-cultivation tension may itself be unsustainable.

The **accelerationist narrative** follows a satirical structure: ironic inversion where an apparent nightmare becomes desirable. AI devalues credentials and universities contract to élite institutions. But here lies the satire, this "worst" scenario might enable the "best" university. Like the fundamentalist university in Houellebecq's *Soumission* (2015) that becomes an intellectual paradise, an institution freed from mass credentialing could return to cultivation. This narrative promotes ideals of depth and standards; its therapeutic function grants permission to acknowledge attractions we're not supposed to feel; its critical edge exposes the constitutive contradictions of democratic education. Who tells this story? Usually, no one explicitly, but the fantasy circulates. The contradiction it reveals: what we claim to value (democratic access) may be structurally incompatible with what makes humanist knowledge valuable (time and tradition).

These forms co-exist. They are structures through which we organize anxiety about an uncertain future. By analyzing these forms, we discover that our present conceptualizations of education already contain tensions between democracy and depth, access and cultivation, equality and excellence. AI futures mirror the existing contradictions, just as history has mirrored us in the competing narratives of who we became. Now the future might have claimed that task.

References

Adorno, Theodor W. 2005. "Education After Auschwitz." In *Critical Models: Interventions and Catchwords*, 191–204. New York: Columbia University Press.

Ahmed, Sara. 2019. *What's the Use? On the Uses of Use*. Durham: Duke University Press.

Biesta, Gert. 2013. *The Beautiful Risk of Education*. Boulder: Paradigm Publishers.

Bjerring, Viggo. 2021. *Verdenshjertet*. Valby: Ekbatana.

Bloom, Harold. 1997. *The Anxiety of Influence: A Theory of Poetry*. 2nd ed. New York: Oxford University Press.

Cave, Nick. 2023. "Response to ChatGPT Writing Nick Cave Songs." *The Red Hand Files*, Issue 218. https://www.theredhandfiles.com/chat-gpt-what-do-you-think/

"Ekspertgruppen om ChatGPT." 2023. "Her er de anbefalinger, vi ikke blev bedt om." *Altinget*.

Fenwick, Tara, and Richard Edwards. 2011. "Reclaiming and Renewing Actor Network Theory for Educational Research." *Educational Philosophy and Theory 43*, sup1: 1–14.

Gadamer, Hans-Georg. 2004. *Truth and Method*. 2nd ed. London: Continuum.

Haraway, Donna. 1991. "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century." In *Simians, Cyborgs, and Women: The Reinvention of Nature*, 149–81. New York: Routledge.

Heimans, Stephen, Gert Biesta, Satoko Takayama, and Mike Kettle. 2023. "ChatGPT and the Rise of Generative AI: Threat or Opportunity?" *Educational Philosophy and Theory 55*, no. 12: 1353–60.

Houellebecq, Michel. 2015. *Soumission*. Paris: Flammarion.

Illich, Ivan. 1993. *In the Vineyard of the Text: A Commentary to Hugh's Didascalicon*. Chicago: University of Chicago Press.

Knox, Jeremy. 2019. "What Does the 'Postdigital' Mean for Education? Three Critical Perspectives on the Digital, with Implications for Educational Research and Practice." *Postdigital Science and Education 1*: 357–70.

Koselleck, Reinhart. 2004. *Futures Past: On the Semantics of Historical Time*. New York: Columbia University Press.

Labaree, David F. 1997. "Public Goods, Private Goods: The American Struggle Over Educational Goals." *American Educational Research Journal 34*, no. 1: 39–81.

Ricoeur, Paul. 1984–1988. *Time and Narrative*. 3 vols. Chicago: University of Chicago Press.

Rorty, Richard. 1989. Contingency, Irony, and Solidarity. Cambridge: Cambridge University Press.

Selwyn, Neil. 2019. *Should Robots Replace Teachers? AI and the Future of Education*. Cambridge: Polity Press.

Shelley, Mary. 1818. *Frankenstein; or, The Modern Prometheus*. London: Lackington, Hughes, Harding, Mavor & Jones.

Sloterdijk, Peter. 1987. *Critique of Cynical Reason*. Minneapolis: University of Minnesota Press.

Sloterdijk, Peter. 2009. "Rules for the Human Zoo: A Response to the Letter on Humanism." *Environment and Planning D: Society and Space 27*, no. 1: 12–28.

Svenbro, Jesper. 1993. *Phrasikleia: An Anthropology of Reading in Ancient Greece*. Ithaca: Cornell University Press.

White, Hayden. 1973. *Metahistory: The Historical Imagination in Nineteenth-Century Europe*. Baltimore: Johns Hopkins University Press.

White, Hayden. 1974. "The Historical Text as Literary Artifact." *Clio 3,* no. 3: 277–303.

White, Hayden. 1987. *The Content of the Form: Narrative Discourse and Historical Representation.* Baltimore: Johns Hopkins University Press.

Williamson, Ben. 2017. *Big Data in Education: The Digital Future of Learning, Policy and Practice*. London: SAGE.

Discretionary space: a multidimensional model for analysing discretion in the time of pervasive AI

Jakob Laage-Thomsen & Helene Friis Ratner

The concept of discretion is central to research in E-government, Public Administration and Sociology of Professions, used to describe the interpretative and situated judgments that public sector professionals and bureaucrats make in their daily work. The concept has a longer tradition in studies of "street-level bureaucrats" [SLB], and has gained increased attention with the transition to "screen" or "system" level bureaucracies from the introduction of ICT systems (Bovens & Zouridis, 2002; Busch, 2018; Reddick, 2005). The interest has accelerated with the introduction of machine-learning technologies, most recently Large Language Models, whose non-deterministic and agentic capabilities further challenges and augments how discretion materializes, and how, discretionary decisions are taken (Busch, 2025; Busch & Henriksen, 2018; Ranerup & Henriksen, 2022; Young et al., 2019). Yet, the literature generally, and the recent literature particularly, suffers from undertheorized models of discretion, focusing instead on how AI interrelates with various tasks or contexts. While important, this prevents us from placing important empirical and conceptual contributions in relation to each other, and prevents accounts of how AI and human-AI systems themselves come to embody and enact forms of judgment.

With an understanding of discretion as engaging in decision-making or solving tasks under constraints in underdetermined or contradictory situations, the paper develops a theoretical model that distinguishes between *discretionary space* – the conditions of possibility and constraints under which *discretionary agency* unfolds. Here, discretionary agency refers to AI- human arrangements that conduct discretion in practice. This model addresses a central gap in existing research by providing an explicit theorization of discretionary space and its relationship to the distribution of discretionary agency.

We elaborate this model by integrating insights from studies in SLB, e-government, and Sociology of Professions, to specify four intersecting domains that condition and constrain discretionary space: bureaucratic rules, professional norms, managerial oversight, and digital infrastructures. Together, these dimensions delineate the space within which discretionary agency is exercised across humans and AI. The integrated framework further clarifies why the deterministic or non-deterministic character of AI systems is pivotal for understanding whether discretion remains an analytically meaningful concept, even when such systems may exercise agency or shape outcomes. The framework thus provides a shared vocabulary for examining how discretion is distributed, transformed, or newly produced as AI systems become embedded in organizational practice.

References

Bovens, M., & Zouridis, S. (2002). From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control. *Public Administration Review, 62*(2), 174–184. https://doi.org/10.1111/0033-3352.00168

Busch, P. A. (2018). *Conceptualizing Digital Discretion Acceptance in Public Service Provision: A Policy Maker Perspective.*

Busch, P. A. (2025). The Artificial Bureaucrat: Artificial Intelligence in Street-Level Work. *Digit. Gov.: Res. Pract., 6*(3), 38:1–38:16. https://doi.org/10.1145/3721138

Busch, P. A., & Henriksen, H. Z. (2018). Digital discretion: A systematic literature review of ICT and street-level discretion. *Information Polity, 23*(1), 3–28. https://doi.org/10.3233/IP-170050

Molander, A., Grimen, H., & Eriksen, E. O. (2012). Professional Discretion and Accountability in the Welfare State. *Journal of Applied Philosophy, 29*(3), 214–230. https://doi.org/10.1111/j.1468-5930.2012.00564.x

Ranerup, A., & Henriksen, H. Z. (2022). Digital Discretion: Unpacking Human and Technological Agency in Automated Decision Making in Sweden's Social Services. *Social Science Computer Review, 40*(2), 445–461. https://doi.org/10.1177/0894439320980434

Reddick, C. G. (2005). Citizen interaction with e-government: From the streets to servers? *Government Information Quarterly, 22*(1), 38–57. https://doi.org/10.1016/j.giq.2004.10.003

Straub, V., Morgan, D., Bright, J., & Margetts, H. (2023). Artificial intelligence in government: Concepts, standards, and a unified framework. *GOVERNMENT INFORMATION QUARTERLY, 40*(4). https://doi.org/10.1016/j.giq.2023.101881

Wallander, L., & Molander, A. (2014). Disentangling Professional Discretion: A Conceptual and Methodological Approach. *Professions and Professionalism, 4*(3). https://doi.org/10.7577/pp.808

Young, M. M., Bullock, J. B., & Lecy, J. D. (2019). Artificial Discretion as a Tool of Governance: A Framework for Understanding the Impact of Artificial Intelligence on Public Administration. *Perspectives on Public Management and Governance*, gvz014. https://doi.org/10.1093/ppmgov/gvz014

Chat-Democracy: Too Little Talk, Too Much Talk?

Jan Løhmann Stephensen

This paper argues that current debates about "AI and democracy" remain too undifferentiated, treating democracy as a single object and AI as a generic disruptor. Instead, it proposes a more precise triangulation—*which AI, and which language practices, for which democracy (and vice versa)*—and focuses on how Large Language Model–based generative AI, as a language-native interface technology, reshapes talk-centric models of democracy in particular. Moving beyond vote-centric imaginaries (elections, disinformation, microtargeting), the paper examines democratic talk as both a semi-formal public-sphere genre and as vernacular everyday political talk in digital "third spaces," where citizens rehearse positions, articulate issues, and become political subjects through shared language. It synthesizes recent critical accounts of "language machines" as infrastructures of ideological formation (including concerns about predigested language, hegemonic narrowing, latent persuasion, and the "AI effect" whereby effective technologies recede into the background while working on users). Against this backdrop, the paper introduces I**nfo-Discursive Deliberative Auxiliaries (IDEAS)** as a concept for understanding embedded conversational systems that increasingly co-produce the informational and rhetorical conditions of participation—potentially widening access to voice, yet also reconfiguring autonomy and agency.

The analysis then turns to emerging cases of "AI-for-deliberation" interventions—especially the Habermas Machine and related proposals that optimize summarization, civility, and convergence—to ask what democratic theory is being operationalized when consensus is treated as the primary telos of talk. The paper's central critique is that whilst many such projects do accommodate talk-centric models, they neglect democracy's counterpart practice: listening (including silence, receptivity, and phatic community-building).

Post-Digital Reaction: Carbon Capitalist Imaginaries of 'Real' Work and Masculinity in a Manufacturing SME

Jana Stefan & Chris Land

Drawing on ethnographic research conducted during a three-year project examining Industry 5.0 in European manufacturing, this paper analyses how gender and political identities conflict with attempts to realise a 'twin transition' of digitalisation and sustainability in manufacturing. Our research case is situated in the industrial heartland of former West Germany. The company produces high-quality industrial gas applications, occupying a specialist market niche with few competitors. Despite this, an aging workforce and political ambitions to decarbonise European manufacturing threaten its long-term viability. Whilst a 'twin transition' would offer some solutions, it was actively resisted at all levels of the organization.

We examine this resistance through the interplay of deindustrialisation and masculinity. Workers' identities in the company were invested in industrial values of hard, physical labour inherited from coal and steel. The paper argues that resistance to digitalisation and sustainability was connected to a kind of 'post-industrial melancholia', in which identities had been actively invested in this sense of loss, valorising previous generations as 'real men', and interpreting digitalisation and electrification as a kind of emasculation. The paper contributes to the literature on masculinity and deindustrialisation (Strangleman, 2024), exploring the relationship between nostalgia, reactionary politics, and contested ideologies of the future of work (Dries et al., 2025).

References

Dries N, Luyckx J, Stephan U, et al. (2025) The Future of Work: A Research Agenda. *Journal of Management*.

Strangleman T (2024) The World We Have Lost: Reflections on Varieties of Masculinity at Work. I*nternational Labor and Working-Class History* 105: 9–25.

Automating Publics: A Multi-Level Framework for Analysing Algorithmic Public Formation

Jannie Møller Hartley & Andreas Birkbak

This paper advances a conceptual framework for analysing how publics are shaped through and across personalised, datafied, and hybrid media infrastructures. The increasing implementation of AI-driven and personalised content recommender systems across the media industry raises fundamental questions about how algorithmic curation reconfigures citizens' engagement with issues of public concern (Birkbak & Carlsen, 2016a). While existing scholarship has typically examined personalisation and in particular News Recommender Systems (NRS's) at the level of single platforms or single news outlets – often focusing on measurable effects such as click behaviour (Einarsson et al., 2025; Haim et al., 2018; Möller et al., 2018) or the values inherent in the design of NRS (Møller, 2023; Schjøtt Hansen & Hartley, 2023)- there is a lack of theoretical frameworks capable of explaining how personalised media environments reshape public formation and linking these developments *across* individual, organisational, and societal levels. The framework mobilizes previous literature on three levels and combines concepts from audience studies with public theoretical and media sociological approaches to media production in order to develop a coherent framework for analysing the automation of publics.

At the micro level, the model conceptualises personalised media use as a communicative practice of decoding within what we term *personalised mediascapes*: cross-platform flows of content characteristic of contemporary, datafied, hybrid media environments (Chadwick, 2013; Lai & Flensburg, 2020; Møller Hartley et al., 2023). Building on encoding/decoding theory (Hall, 1991) and research on media experiences (Ytre-Arne & Moe, 2021), this level highlights how individuals interpret personalised content normatively and emotionally. At the meso level, the model includes personalization algorithms as *algorithmic assemblages* (Ananny, 2016) and the cultural dimensions of algorithms (Seaver, 2017), conceptualizing NRSs as socio-technical systems that encode normative values and organisational imaginaries into processes that match users with public issues. This builds on previous work showing that personalisation systems embed assumptions about users and their concerns (Bucher, 2018; Lomborg & Kapsch, 2020). Through "ethnographies of algorithms," this level captures how personalisation collapses traditional boundaries between encoding and decoding as aggregated user data become integral to producing personalised content flows (Birkbak & Carlsen 2016b). Finally, at the macro level, the model mobilises the concept of *issue publics* (Birkbak, 2013; Dewey, 1927; Marres, 2005) to theorise how algorithmically distributed content shapes collective attention, controversies and the emergence of public issues. In contrast to platform-specific notions such as networked (boyd, 2011) or calculated publics (Gillespie, 2014), issue publics emphasise how shared stakes in problems generate publics across media sites and institutional arenas. This perspective enables analysis of how personalisation modulates the visibility, articulation, and coherence of issues across the broader media ecology.

Having presented the framework the article discusses how these levels are linked, and what these means for how researchers approach the analysis of personalised media production and use empirically and methodologically.

References

Ananny, M. (2016). Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science, Technology, & Human Values, 41*(1), 93–117. https://doi.org/10.1177/0162243915606523

Birkbak, A. (2013). From networked publics to issue publics: Reconsidering the public/private distinction in web science. *Proceedings of the 5th Annual ACM Web Science Conference on – WebSci '13*, 24–32. https://doi.org/10.1145/2464464.2464489

Birkbak, A., & Carlsen, H. (2016a). The Public and its Algorithms: Comparing and Experimenting with Calculated Publics. I L. Amoore, & V. Piotukh (red.), *Algorithmic Life: Calculative Devices in the Age of Big Data* (s. 21–34). Routledge.

Birkbak, A., & Carlsen, H. A. B. (2016b). The World of EdgeRank: Rhetorical Justifications of Facebook's News Feed Algorithm. *Computational Culture*, (5)

Boyd, D. (2011). Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In Z. Papacharissi (Ed.), *A networked self: Identity, community and culture on social network sites* (pp. 39–58). Routledge.

Bucher, T. (2018). If...Then: Algorithmic Power and Politics. In *Programming the News* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780190493028.003.0006

Chadwick, A. (2013). The Hybrid Media System: Politics and Power. In *The Hybrid Media System*. Oxford University Press.

Dewey, J. (1927). *The public and its problems.* George Allen & Unwin. Ltd, London.

Einarsson, Á. M., Petrucci, E., Møller Hartley, J., & Lomborg, S. (2025). "I must have clicked on something": Assesing news recommendations from an everyday perspectives. *Journalism Practice.*

Gillespie, T. (2014). The Relevance of Algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media Technologies* (pp. 167–194). The MIT Press. https://doi.org/10.7551/mitpress/9780262525374.003.0009

Haim, M., Graefe, A., & Brosius, H.-B. (2018). Burst of the Filter Bubble?: Effects of personalization on the diversity of *Google News. Digital Journalism, 6*(3), 330–343. https://doi.org/10.1080/21670811.2017.1338145

Hall, S. (1991). Encoding, decoding. In *The Cultural Studies Reader*. Routledge.

Lai, S. S., & Flensburg, S. (2020). Appscapes in everyday life: Studying Mobile Datafication from an Infrastructural User Perspective. *MedieKultur: Journal of Media and Communication Research, 36*(69), 029–051. https://doi.org/10.7146/mediekultur.v36i69.121018

Lomborg, S., & Kapsch, P. H. (2020). Decoding algorithms. *Media, Culture & Society, 42*(5), 745–761. https://doi.org/10.1177/0163443719855301

Marres, N. (2005). Issues spark a public into being. A key but often forgotten point of the Lippmann-Dewey debate. In B. Latour & P. Weibel (Eds.), *Making Things Public. Atmospheres of Democracy*. MIT Press, Cambridge MA.

Møller Hartley, J., Sørensen, J. K., & Mathieu, D. (Eds.). (2023). *DataPublics: The construction of publics in datafied democracies*. Bristol University Press.

Möller, J., Trilling, D., Helberger, N., & Van Es, B. (2018). Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society, 21*(7), 959–977. https://doi.org/10.1080/1369118X.2018.1444076

Møller, L. A. (2023). Designing Algorithmic Editors: How Newspapers Embed and Encode Journalistic Values into News Recommender Systems. *Digital Journalism*, 1–19. https://doi.org/10.1080/21670811.2023.2215832

Schjøtt Hansen, A., & Hartley, J. M. (2023). Designing What's News: An Ethnography of a Personalization Algorithm and the Data-Driven (Re)Assembling of the News. *Digital Journalism, 11*(6), 924–942. https://doi.org/10.1080/21670811.2021.1988861

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society, 4*(2), 205395171773810. https://doi.org/10.1177/2053951717738104

Ytre-Arne, B., & Moe, H. (2021). Folk theories of algorithms: Understanding digital irritation. *Media, Culture & Society, 43*(5), 807–824. https://doi.org/10.1177/0163443720972314

UK SafetyTech is unsafe for school pupils and needs urgent regulatory action

Jen Persson

AI-enabled "SafetyTech" used in education offers a lens to view the democratic concerns of AI-surveilled society. Commercial systems monitor millions of UK school pupils through keystroke logging, screen capture and keyword-based classifiers, often 24/7 and beyond school grounds, at home and in school holidays. These tools are marketed as safeguarding technologies, but normalise pervasive automated surveillance of children by private companies, with little public debate, weak evidence of effectiveness, and significant risks (Defend Digital Me, 2025; Laird et al., 2022).

From a children's rights perspective, this infrastructure conflicts with international norms. Article 8 of the ECHR and UNCRC Article 16 promote children's right to privacy and protection from arbitrary interference with their correspondence and family life. The UNCRC has stressed that automated filtering and profiling must not "supplant, manipulate or interfere with children's ability to form and express their opinions in the digital environment" (OHCHR, 2021). Yet school SafetyTech through contextual AI risk analysis routinely records intimate content, unsent or deleted text and offline documents, building child-level longitudinal behavioural profiles in ways that are not rights respecting, transparent or demonstrably necessary and proportionate.

U.S. civil society research evidence suggests such systems are used more for discipline than safety, and that they chill speech, and disproportionately harm minoritised children (Thakur and Laird, 2023). UK practice appears to follow similar patterns, including feeding highly inaccurate data into the national anti-radicalisation and extremism programme Prevent, where education is already the largest source of referrals. Treating every child as a potential threat, based on opaque AI-risk profiling, corrodes trust in authority and conditions for democratic society.

These developments are shaped by political agendas. The UK government actively promotes SafetyTech as an export sector, positioning the country as a "world-leader". At the same time, ministers have said that manual inspection of private communications would be "a clear breach of privacy" (Hansard, 2020), even as de facto "man-in-the-middle" architectures are deployed without public debate, to decrypt and inspect children's traffic. The Online Safety Act 2023 and Codes of Practice further entrench "proactive technology" ignoring these tensions (OFCOM, 2025).

SafetyTech often depends on cloud providers, resulting in vendor lock-in and concentrating control of educational infrastructures and highly sensitive data in a few private corporations. Democratically, this outsources the key core public functions of safeguarding in schools, to opaque contractual arrangements with intermediaries and cloud platforms, without oversight, reducing school staff agency and authority, or controls on national costs (Patel et al., 2019).

We argue that AI-based SafetyTech in schools fails key tests of legality, necessity, proportionality and non-discrimination, and is therefore incompatible with democratic commitments to children's rights. Policy on children's online safety must shift from technological solutionism towards: (1)

strict limits on routine, child-level surveillance; (2) mandatory independent auditing and error-rate reporting; (3) enforceable routes to exercise data protection rights including rights of access, correction, data minimisation, retention and deletion for pupils and families; and (4) meaningful participation of children, parents, educators and civil society in enforcement of rights-respecting governance of AI in education.

References

Defend Digital Me. (2025). *SafetyTech in schools:* A timeline. https://defenddigitalme.org/safety-tech-in-schools-a-timeline/

Hansard. (2020, 15 December). *Online harms consultation*, Vol. 686. https://hansard.parliament.uk/Commons/2020-12-15/debates/1B8FD703-21A5-4E85-B888-FFCC5705D456/OnlineHarmsConsultation#contribution-C6B532EC-E744-4084-BE3B-FBEB0386B354

La Rue, F. (2013). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/HRC/23/40)*. UN Human Rights Council. https://www.refworld.org/reference/themreport/unhrc/2013/en/96147

Laird et al. (2022). *CDT Report: Hidden harms: The misleading promise of monitoring students online.* Center for Democracy and Technology. https://cdt.org/insights/report-hidden-harms-the-misleading-promise-of-monitoring-students-online/

OFCOM. (2025). *Proactive Technology Draft Guidance Additional Safety Measures—Annex 6.* https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/consultation-online-safety---additional-safety-measures/annexes/annex-6---draft-guidance-to-proactive-technology-measures.pdf?v=399379

OHCHR | *General comment No. 25 (2021) on children's rights in relation to the digital environment.* (2021.). OHCHR. https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation

Online Safety Act 2023, c. 50. (2023). *UK Public General Acts*. https://www.legislation.gov.uk/ukpga/2023/50/contents/enacted

Patel, F. (2019). *The Brennan Center: School surveillance zone. Brennan Center for Justice*. https://www.brennancenter.org/our-work/research-reports/school-surveillance-zone

Thakur, D., and Laird, E. (2023). *Beyond the screen: Parents' experiences with student activity monitoring in K–12 schools*. Center for Democracy and Technology. https://cdt.org/insights/report-beyond-the-screen-parents-experiences-with-student-activity-monitoring-in-k-12-schools/

Designing Friction: Regional Journalism's Role in Contestation of AI Systems

Jeroen de Vos, Jessy de Cooker & Danielle Arets

This paper argues that journalism itself is a site of democratic struggle. Regional media outlets can become spaces for adversarial practices; not only tailored to problem-solving but also actively accommodating dissensus. By making smart technologies and algorithmic governance visible and creating space for genuine agonistic debate, journalism can expand their public watchdog function in relation to increasing omnipresence of smart AI technologies, thereby enabling citizens to help contest the systems that govern them. Two Dutch regional cases suggest when journalists engage critically with algorithmic systems and amplify citizen concerns, they strengthen broader accountability ecosystems: when journalists work alongside researchers, ombudspeople, and community members, friction can become genuinely productive.

Despite growing journalistic attention to the impact of smart AI technologies on domains like public space, mobility and government at national and more tech-specific newsrooms, our media analysis of 495 Dutch news articles shows regional media has remain largely disconnected from this debate. While regional journalists do catch signals, they often lack the expertise and connections to question these smart technologies. Journalism should however be a frontrunner in this field: if journalists do not sufficiently understand these systems, they cannot effectively report on or challenge them.

Contemporary "smart" systems are designed for seamlessness. Byung-Chul Han describes this pursuit of "Das Glatte" (smoothness) as a defining feature of our moment; we optimize for efficiency while losing our capacity to engage with conflict (Han, 2015). Similarly, the 'smart' paradigm expands on early ubiquitous computing notions employing a techno- and consumer-centric understanding of society (Lauwers & Papa, 2015; Weiser, 1991; Aarts & Manzano, 2003). Yet democracy, according to Chantal Mouffe, is inherently based on conflict. Different value systems are part of democracy and should be recognized as such (Mouffe, 2013).

Carl DiSalvo's concept of adversarial design foregrounds conflict and "productive dissensus" to make political dimensions of systems visible (Disalvo, 2015). In this paper, we propose extending this framework to journalism: rather than solving problems, regional journalists can deliberately create space for friction and deliberation. Expanding on existing participatory journalism methods like citizen panels, collaborative investigations, community listening projects (eg. DUIC 2025), practitioners can choose from a growing collection of tools for ethical and inclusive dialogue sensitive to conduct productive dissensus[12].

Dutch regional cases of smart technologies contested in journalism already illustrate the potential of productive dissensus. The Dutch Data Protection Authority decided Enschede's WiFi-tracking

12 Examples include dialogue tools like 'Dialogue of the Deaf' (Twist et al. 2025) and 'Which AI Type' (de Vos, 2025), ethical tools like Moral Design Game (Moral Design Game, 2025), and TILT or scenario design interventions like scenario archetypes (van der Vorst, 2020; Boschetti et al. 2016; Fergnani et al. 2020).

system to be monitoring citizens unlawfully, after publication in a local journal sparking public controversy (1Twente, 2017; European Data Protection Board, 2021). Media coverage, regulatory action, and citizen pressure combined to create policy consequences. Similarly, documented concerns about algorithmic approaches to juvenile crime assessment in Apeldoorn raised human rights questions that prompted governance discussion (Nazarski, 2017).

References

1Twente. (2017). De gemeente waarschuwt met stickers op verkeersborden, prullenbakken en toegangspalen in binnenstad voor wifi-tracking. *1Twente*. Retrieved from

Aarts, E. H., & Marzano, S. (Eds.). (2003). *The new everyday: Views on ambient intelligence*. Rotterdam: 010 Publishers.

Boschetti, F., Price, J., & Walker, I. (2016). Myths of the future and scenario archetypes. *Technological Forecasting and Social Change,* 111, 76–85.

De Vos, J. (2025). *Welk AI type ben jij? [Which AI Type]* Retrieved from https://welkaitype.nl/

DiSalvo, C. (2015). *Adversarial design*. MIT Press.

European Data Protection Board. (2021, April 30). *Dutch DPA fines municipality for Wi-Fi tracking.*

Fergnani, A., & Song, Z. (2020). The six scenario archetypes framework: A systematic investigation of science fiction films set in the future. *Futures*, 124, 102645. https://doi.org/10.1016/j.futures.2020.102645

Han, B.-C. (2015). *Die Errettung des Schönen*. S. Fischer Verlag.

Lauwers, D., & Papa, E. (2015). Towards a smarter urban mobility. In *Colloquium Vervoersplanologisch Speurwerk: Antwerp, Belgium*.

Mouffe, C. (2013). *Agonistics: Thinking the world politically*. Verso.

Nazarski, E. (2017). *Amnesty International: Algoritmes moeten zich aan de mensenrechten houden*.

Twist, A. van. (2025). Dialogue of the Deaf: How deliberation with discontented citizens may hopelessly fail. *Public Administration*.

van der Vorst, R.. (2020). Technology Impact Cycle Tools: An online tool to assess the impact of new technology. In INTED2020 *Proceedings* (pp. 599–600). IATED.

Weiser, M. (1991). The Computer for the 21 st Century. *Scientific american, 265*(3), 94–105. https://doi.org/10.1038/scientificamerican0991-94

Artificial Intelligence, Blackboxing, and Institutional Decision-making

Jesper Ryberg

The fact that citizens have access to information is often highlighted as a key ingredient in a healthy political system. For instance, citizens need access to information to make informed political choices, and to be able to evaluate decision-making by politicians and societal institutions. The purpose of this talk is to consider to what extent something crucial is lost if decisions and administration of public institutions are increasingly being delegated to more or less impenetrable AI systems. More precisely, it will be considered to what extent the use of AI in institutional decision-making should satisfy a requirement of algorithmic transparency. Firstly, it will be considered why transparency should be regarded as ethically important. Secondly, it will be argued that a requirement of algorithmic transparency is far from simple and has often been misleadingly overstated. Third, some of the concerns that follow from the fact that transparency is a complex issue will be outlined. The discussion will draw on the use of AI at the criminal courts as an example, because the criminal justice system constitutes a high-stakes context and, further, because transparency in decision-making is traditionally regarded as a core element of criminal justice.

Against augmentation: How AI tools obstruct leadership

Johan Jönsson, Sverre Spoelstra & Peter Svensson

Leadership practices are increasingly intertwined with AI tools that come with a promise to 'augment' human leadership (e.g. De Cremer, 2020; Hougaard & Carter, 2024), for instance that such tools can help leaders "become more inwardly agile and foster creative approaches to transformation" (Bourton & Lavoie, 2018: 3). The promise is that leaders can use AI tools for advice, idea generation, data-collection, and decision-making (e.g. Bevilacqua, et al., 2025; Quaquebeke and Gerpott, 2023). While some describe this as a 'new partnership', formed by the complementary strengths of humans and machines (e.g. Harms and Han, 2019), the human-machine assemblages shaping leadership are in practice far more complex and ambiguous.

This paper is based on our study in a historical museum in Sweden, which has recently introduced an AI tool called *HumanITY* (pseudonym). The tool is marketed by its developer as a data-driven solution to enhance employee voice, increase team engagement, and thereby augment leadership decision making. On a three-weeks basis, HumanITY collects and evaluates anonymous data on the organizational climate across several dimensions such as 'leadership', 'well-being', and 'personal development'. At its core, HumanITY is about *communication* as it calls for staff members to speak up and managers to listen, engage and act on what is being said. However, our study reveals a much more puzzling picture.

Drawing on over 25 qualitative interviews with employees and managers, we will show how they struggle to communicate through the tool, often due to difficulties associated with responding to its questions, either due to the ambiguity of the questions or simply because the questions or the answerable options are not experienced as relevant to the staff. One of our main findings is that, rather than augmenting leadership, HumanITY obstructs it and becomes in effect a mechanism for the avoidance of leadership. Put differently, managers use the tool's opacity as a way of evading responsibility and leadership initiatives.

We argue that AI leadership tools, rather than facilitating human leadership, can inadvertently obstruct it. The paper contributes to the knowledge of digital leadership practices by highlighting the tensions, limitations, and unintended consequences that emerge when leadership is mediated through AI tools. Rather than offering data driven solutions, the tool revealed problems expressed by staff members while leaving managers with feelings of impotence and frustration. Against promises of leadership augmentation by AI tools lies a complicated nexus of employee voices and leadership efforts lost in translation.

References

Bevilacqua, S., Masárová, J., Perotti, F. A., and Ferraris, A. (2025) Enhancing top managers' leadership with artificial intelligence: insights from a systematic literature review. R*eview of Managerial Science,* 1–37.

Bourton, S., Lavoie, J., & Vogel, T. (2018). Will artificial intelligence make you a better leader? *McKinsey Quarterly, 2*(1), 72–75.

De Cremer, D. (2020). *Leadership by algorithm: Who leads and who follows in the AI era?* Harriman House Limited.

Harms, P. D. and Han, G. (2019) Algorithmic leadership: The future is now. *Journal of Leadership Studies, 12*(4): 74–75.

Hougaard, R., & Carter, J. (2024). How AI can make us better leaders. *Harvard Business Review*.

Quaquebeke, N. V., and Gerpott, F. H. (2023) The now, new, and next of digital leadership: How Artificial Intelligence (AI) will take over and change leadership as we know it. *Journal of Leadership & Organizational Studies, 30*(3): 265–275.

Same Behavior, Different Politics: Algorithmic Exposure Inequality on TikTok During German Elections

Johannes Wolfgram, Licia Bobzien, Sarah Weißmann, Aaron Philipp, Jasper Tjaden & Roland Verwiebe

Digital platforms have become central arenas for political communication, yet their recommendation algorithms remain opaque gatekeepers determining which political content reaches citizens. While traditional media in Germany face legal requirements to ensure balanced political exposure during elections, social media platforms distribute content through proprietary systems that escape equivalent scrutiny. This raises fundamental questions for democratic governance: Do algorithms create systematically unequal political information environments, and if so, through what mechanisms?

We address this question using a social research bot (SRB) methodology that overcomes key limitations of observational platform research. With real users, exposure patterns confound algorithmic curation with individual behavior, networks, and history. Our approach deploys automated user accounts under controlled conditions, isolating platform-side variation. We created 78 SRBs on TikTok during two German election periods: three East German state elections (August–October 2024; N=75,987 videos) and the federal election (January–April 2025; N=485,323 videos). All bots simulated young, non-partisan users with no platform history. We varied political interest: "passive" users engaged only with non-political content, while "active" users occasionally searched neutral political terms without expressing partisan preferences. Crucially, no bot revealed any ideological orientation through behavior.

Our findings reveal substantial exposure inequality. Passive users encountered far-right AfD content at rates 3–8 times higher than centrist parties. For politically interested users, AfD exposure reached 4.13% of feeds versus 1.11% for CDU/CSU and 0.80% for SPD. This gap is most pronounced for content from unofficial accounts. We identified 377 unique accounts posting AfD-supportive content in our feeds, compared to 47 for SPD and 25 for CDU, suggesting multiplier networks operating independently of official party activity. While average engagement shows limited party differences, AfD content exhibits substantially more viral outliers (>100,000 likes), producing disproportionate visibility through TikTok's engagement-weighted algorithm.

Notably, political interest amplifies rather than moderates this inequality. Users seeking election information encounter more skewed distributions, suggesting the available pool of political content is itself asymmetric.

These findings contribute to debates on algorithmic governance and platform accountability. They demonstrate that identical user behavior produces systematically different political information environments, raising questions about platforms as democratic infrastructures. Germany's media regulations mandate balanced exposure precisely because information access shapes political par-

ticipation. If young citizens' primary information environment systematically overrepresents certain parties independent of user preference, this constitutes a structural challenge for democratic communication that current regulatory frameworks do not address.

Our study documents exposure patterns but cannot trace downstream effects on attitudes or behavior. Future research should investigate whether algorithmic exposure inequality translates into measurable opinion shifts, and whether similar patterns emerge across platforms and national contexts.

Reproduction of scientific knowledge and norms for explaining schizophrenia in Predictive Psychological AI

Jonas Yurtsever Vesterdal

AI-based technologies for prediction of psychological and psychiatric states through means of identification, classification, diagnosing, and treatment monitoring – what I broadly term 'Predictive Psychological AI' (PPAI) - are increasingly being developed and used within the mental health area. PPAI technologies have showed promising results in predicting, diagnosing, and monitoring data points associated with various psychiatric conditions. However, few authors have analysed the fundamental scientific knowledge and normative assumptions about mental disorder or deviance found in such technologies, or traced the medical-historical foundations of such knowledge and norms and linked these to the technologies. Using the case of schizophrenia as studied in a number of PPAI studies and tracing its various conceptualisations throughout medical history, the purpose of this paper is to demonstrate how PPAI studies reproduce existing knowledge and norms for investigating schizophrenia. Specifically, I argue that PPAI studies reproduce naturalistic, genetic-neurobiological etiologies for the condition even though the empirical foundation of such etiologies may be subject to doubt. Such naturalistic ambitions to explain psychiatric conditions like schizophrenia can be traced back to ancient Greece, through the establishment of the first psychiatric hospitals, and into the early symptom-based formulation of schizophrenia and related diagnoses. Today, this medical-historical naturalistic tradition informs the data points that PPAI technologies seek to capture in order to predict schizophrenia as well as symptom severity and relapse risk among patients with the condition. Alongside these naturalistic aspirations within medical history, psychologically and sociologically grounded perspectives and research on schizophrenia and other mental health conditions have also persisted. However, I contend that within PPAI technologies these perspectives are either interpreted in ways that reinforce genetic-biological explanations of schizophrenia or are excluded altogether as data parameters I conclude the paper by discussing whether PPAI technologies can yield novel, potentially transformative insights into psychiatric diagnoses such as schizophrenia. I contend that for such technologies to achieve this, they must succeed in capturing the contextual and intersubjective variables that also correlate with the development of schizophrenia, rather than limiting their scope to variables that predominantly reflect existing genetic-neurobiological and symptomatic dimensions. If PPAI technologies are to be truly transformative, I suggest, it may further be necessary for them to break entirely with existing diagnostic categories. This would perhaps allow the knowledge they generate to inform the development of wholly new conceptual frameworks for identifying mental distress or deviation.

Constructing Controversies: The role of Communicative AI in Corporate Communication

Julia Eisner

The emergence of generative AI and LLMs has rapidly entered corporate communication departments, where these systems increasingly co-create communicative outputs and reshape communicative practices (Banholzer, Quest, & Rossbach, 2023). Rather than automating tasks traditionally carried out by professionals (Weller & Lock, 2024) —such as content creation, stakeholder engagement, or media monitoring—communicative AI participates in meaning-making processes and influences how organisations construct narratives, relationships, and legitimacy.

Communicators translate strategy into meaning, interpret societal expectations, mediate between organisational functions, and safeguard credibility. When communicative AI enters this domain, it affects not only workflows but also the organisation's ability to articulate purpose and maintain trusted narratives. Rather than a short-lived hype, this development marks a structural transformation in how organisations produce and interpret communication (Zerfass et al., 2024; Buhmann & White, 2022). While professional discourse foregrounds efficiency, scholarly work highlights deeper relational, ethical, and epistemological implications (Guzman & Lewis, 2020; Zerfass et al., 2024; Buhmann & Gregory, 2023). Within communication and media studies, *communicative AI* has therefore emerged as a sensitising concept that foregrounds AI's active role in shaping communicative processes.

"Communicative AI (1) is based on various forms of automation designed for the central purpose of communication, (2) is embedded within digital infrastructures, and (3) is entangled with human practices" (Hepp et al., 2023: 48).

This conceptualization challenges human-centered models of communication and raises questions about agency, authorship, and the boundaries of professional roles in corporate contexts. The study is situated within the Social Constructionist tradition, which understands reality as produced in and through communicative practices (Knoblauch 2017; Knoblauch & Pfadenhauer 2023). Rather than treating communicative AI as a fixed technological entity, the paper conceptualises it as an object of discursive construction and professional positioning. Yet empirical insight into how communication professionals construct these meanings remains limited.

The study is embedded in an ongoing PhD project, and this paper investigates the following research question: *How do corporate communication professionals discursively construct organisational controversies around the role of communicicative AI in corporate communication?*

The study examines how professionals articulate and negotiate the role and boundaries of communicative AI. Rather than analysing direct interaction with AI systems, the workshops reveal how professionals anticipate and frame AI's integration. Methodologically, the analysis follows Constructivist Grounded Theory (Charmaz, 2014), treating workshops as arenas of professional meaning-making rather than attitude measurement.

Empirically, the paper draws on three full-day workshops conducted in 2025 (February, June, November) with 35 corporate communication professionals from large German organisations. Each workshop generated field notes, video recordings, presentations, and collaboratively produced artefacts. These materials are interpreted as communicative sites where meanings and organisational tensions around communicative AI are constructed. As analysis is ongoing, findings are preliminary.

Early results indicate three emerging tensions. First, communicative AI is positioned within fragmented organisational landscapes in which communication departments are often insufficiently integrated into AI strategies. Second, professionals articulate ambivalent narratives of AI as both an efficiency enhancer and a potential threat to professional identity—echoing ethical concerns identified in previous research (Buhmann & White, 2022; Zerfass et al., 2024). Third, evaluation tensions emerge between KPI-driven managerial expectations and less quantifiable communicative values such as legitimacy, trust, and stakeholder orientation.

Automated avoidance: Theorizing generative AIs ability to produce non-experience

Julie Vulpius

Generative AI is not only helping us solve problems; it is also helping us avoid them. Avoidance coping is a psychological strategy where individuals evade tackling stressors directly and instead divert attention or disengage directly from challenging situations, emotions, or thoughts (Scott, 2024). While avoidance can provide temporary relief, it often undermines learning, growth, and resilience in the long run (Roth & Cohen, 1986). Generative AI and large language models (LLMs) are increasingly integrated into personal and professional contexts with promises of optimization, efficiency and quality, that way solving our problems like magic (Gell, 1988; Elish & boyd, 2017). However, their widespread adoption also enables, facilitates and in some cases even automate avoidance by allowing users to bypass challenging tasks, interactions, and emotional labor that might otherwise foster cognitive, emotional, and personal growth.

This paper develops the concept of 'automated avoidance' to highlight an overlooked consequence of generative AI. Specifically, it theorizes how generative AI tools enable, facilitate, and even automate avoidance behavior by minimizing or removing friction, effort, and confrontation. Drawing on psychological theories on avoidance and cognitive development, as well as negative sociology (Illouz, 2019; Scott 2019, 2023) and the sociology of absences (Sousa Santos, 2016), I examine generative AI's ability to circumvent discomfort and actively produce non-experience. The paper draws on three illustrative examples to develop and analyze the concepts 'automated avoidance':

1. **Ghosting in the termination of romantic relationships:** Ghosting, in itself an avoidant coping behavior facilitated by technology, is the act of ending communication and terminating a relationship without explanation or further communication (Šiša 2022). Tools like chatbots or automated messaging can help individuals sidestep uncomfortable confrontations and avoid emotional discomfort. This impedes personal development and emotional growth.

2. **Simplifying personal political decision-making:** The increasing reliance on AI-generated political quizzes and voting recommendations reduces the cognitive and emotional labor involved in making a subjectively meaningful and informed political choice (Coleman 2013). AI driven political technology transforms information seeking behavior and the activation of sense and sensibility in the personal choice of voting to a calculated finite result, which risk weakening the affective experience of participation and challenge the development of political efficacy and literacy (Jensen 2011)

3. **AI as a reading assistant in higher education:** Using AI to summarize difficult texts offer students an easy way to avoid the strenuous struggle of engaging with and working actively to understand complex academic material (Afdal et al. 2022). While generating summaries and key points with AI offer a quick reward and surface level understanding, it can obstruct 'deep learning' and the development of academic skills and literacies (Maguire et al. 2022)

The paper concludes by discussing the ethical challenges posed by technologies that automate avoidance, including the risk of impeding human cognitive and emotional development and personal growth. By reframing AI not just as a problem-solving tool but as an avoidance-enabling technology, I show how generative AI is capable of producing non-experiences and discuss its 'hidden costs' (Gell, 1988) at individual, organizational, and societal level.

References

Afdal, W. A., Spernes, K., & Hoff-Jenssen, R., (2022) Academic reading as a social practice in higher education. *Higher education, 85*, 1337–1355

Coleman, S. (2013). In *How Voters Feel* (pp. i–ii). half-title, Cambridge: Cambridge University Press.

Gell, A. (1988). Technology and magic. *Anthropology Today, 4*(2), 6–9.

Elish, M. C., & boyd, danah. (2017). Situating methods in the magic of Big Data and AI. *Communication Monographs, 85*(1), 57–80.

Illouz, E. (2019). *The end of love: A sociology of negative relationships*. Oxford University Press.

Jensen, J. (2011). Citizenship in the Digital Age: The Case of Denmark. *Policy & Internet, 3*(3), 49–70

Maguire, M., Reynolds, A. E., & Delahunt, B. (2020). Reading to be: The role of academic reading in emergent academic and professional student identities. *Journal of University Teaching and Learning Practice, 17*(2), 58–70.

Roth, S., & Cohen, L. J. (1986). Approach, avoidance, and coping with stress. *American Psychologist, 41*(7), 813–819.

Scott, S. (2019). *The social life of nothing: Silence, invisibility and emptiness in tales of lost experience.* University of Sussex. Book.

Scott, S. (2023). *Nothing important: exploring the personal and social meanings of negative experience*. University of Sussex. Journal contribution.

Šiša, A. (2022) Ghosting apps as technologically assisted exit strategy on mobile dating apps. *Antropological Notebooks, 28* (2).

Sousa Santos, B. (2016). *Epistemologies of the South*. Routledge.

Green Teaming AI: Eliciting and problematising 'the (missing) environment' in generative AI

Jutta Haider, Björn Ekström, Malte Rödl & James White

This contribution presents ongoing research seeking to identify, examine and rethink how "the environment" is configured into generative AI systems. We draw inspiration from the method of 'red teaming' AI to suggest *green teaming* as a distinct approach to provide a first step towards mapping the diverse ways in which 'the environment' is constituted in GenAI, including how it is ignored.

So-called 'artificial intelligence' (AI) encompasses a diverse set of data-driven technologies for automation, prediction, and decision-making, some of which are becoming deeply integrated into society, culture, and professional practices, including environmental communication, environmental (social) science, policy, and management (White & Lidskog, 2022). In particular, the hype surrounding and proliferation of commercial generative AI applications have brought the environmental harms of these technologies into sharper focus, most notably their unsustainable resource use and energy demands. (e.g. Bossert & Loh, 2025; Galaz 2025). Yet the direct environmental impacts of the material infrastructure are not the only way in which the environment and generative AI interrelate. For example, the use of generative AI to fabricate scholarly work on environmental issues poses a problem for how evidence is established (Haider et al., 2024, as does its role in generating, amplifying, and disseminating climate obstruction content (Ekström & Haider, in press). There are also strategic ignorances regarding environmental and climate change knowledge embedded within AI models (Haider & Rödl, 2023), concerns about AI platforms' hyper-consumerist values and algorithmically facilitated emissions (Haider et al., 2025), and a lack of attention to environmental concerns in generative AI's discursive infrastructure (Ekström et al., 2025).

Generative and other AI systems are transformative technologies that not only represent but are also constitutive of human-environment relationships. However, inadequate industry disclosure of the underlying data, algorithms, and decision-making makes it difficult to understand how values concerning the environment are embedded and reshaped. *Green teaming* is a participatory approach designed to highlight environmental concerns, specifically emphasising the indirect and systemic environmental effects of generative AI. It is modelled on but extending the ideas of red teaming, which is typically used in technology companies to identify unintended, unsafe, and harmful outcomes of AI models. Recently, civil society and public sector organisations have begun to adopt red teaming in 'the public interest' (AI Risk and Vulnerability Alliance (ARVA) et al., 2025) or for 'social good' (UNESCO, 2025).

The presentation discusses foundational ideas and work-in-progress, inviting comments, suggestions and new collaborations.

References

AI Risk and Vulnerability Alliance (ARVA), Singh, R., Blili-Hamelin, B., Anderson, C., Tafesse, E., Vecchione, B., Duckles, B., & Metcalf, J. (2025). *Red-Teaming in the Public Interest.* Data & Society Research Institute. https://doi.org/10.69985/VVGP4368

Bossert, L. N., & Loh, W. (2025). Why the carbon footprint of generative large language models alone will not help us assess their sustainability. *Nature Machine Intelligence, 7*(2), 164–165. https://doi.org/10.1038/s42256-025-00979-y

Ekström, B., Engström, L., & Haider, J. (2025). Foundation models' acceptable use policies disregard the environment and nature. *Nature Machine Intelligence.* https://doi.org/10.1038/s42256-025-01134-3

Ekström, B. & Haider, J. (in press). A methodology for analysing informational textures: Skipping stones and noticing the ripples. *Journal of Documentation.*

Galaz, V. (2025). *Dark Machines: How Artificial Intelligence, Digitalization and Automation is Changing our Living Planet.* Routledge. https://doi.org/10.4324/9781003317814

Haider, J., & Rödl, M. (2023). Google Search and the creation of ignorance: The case of the climate crisis. *Big Data & Society, 10*(1), 20539517231158997. https://doi.org/10.1177/20539517231158997

Haider, J., Rödl, M., & White, J. (2025). Unsustainable artificial intelligence and algorithmically facilitated emissions: The case for emissions-reduction-by-design. *Big Data & Society, 12*(3), 20539517251365226. https://doi.org/10.1177/20539517251365226

Haider, J., Söderström, K. R., Ekström, B., & Rödl, M. (2024). *GPT-fabricated scientific papers on Google Scholar: Key features, spread, and implications for preempting evidence manipulation.* Harvard Kennedy School Misinformation Review. https://doi.org/10.37016/mr-2020-156

UNESCO, Chowdhury, R., Skeadas, T., & Amos, S. (2025). *Red Teaming artificial intelligence for social good—The PLAYBOOK.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000394338

White, J. M., & Lidskog, R. (2022). Ignorance and the regulation of artificial intelligence. *Journal of Risk Research, 25*(4), 488–500. https://doi.org/10.1080/13669877.2021.1957985

Natural Intelligence: Potentials and Perils of AI-Enhanced Surveillance of Global Climate Change

Klaus Bruhn Jensen, Ece Elbeyi, Kiran Kappeler, Dechun Zhang & Yang Yang

The Internet of Things (IoT) (ITU, 2005) enables comprehensive and continuous monitoring of the effects of human activities on Earth's ecosystems (IPCC, 2023), increasingly informed by AI (Zhang & Tao, 2021) to facilitate the green transition that the 195 parties to the Paris Agreement signed up to (United_Nations, 2015). As enhanced by artificial intelligence, human intelligence may thus tap into the intelligence of nature (Abram, 1996) to ensure the survival and continued coexistence of human and other species. But the scope and the scale of the necessary monitoring of natural and built environments simultaneously entails surveillance of persons, communities, and cultures in what might become a "green but inhumane" future (Creutzig et al., 2022). The paper identifies, contextualizes, and assesses the resulting controversies and ethical dilemmas that are emerging in the window of opportunity between 2025 and 2050, identified by the Paris Agreement as the deadline for a net-zero world. The first section introduces the massive AI-enhanced planetary monitoring systems currently being implemented in the European Union (https://destination-earth.eu), the United States (https://esto.nasa.gov/earth-system-digital-twin/), and China (https://earthlab.iap.ac.cn/en/index.html#firstPage). The second section details their functions of information processing about ecosystems and societies, networked communication among private companies, public institutions, and citizens, and individual and collective agency in response to the climate crisis. Compared to past analog media systems (Hallin & Mancini, 2004) and digital communication system of the last few decades (Jensen & Helles, 2023), these planetary systems represent a radical step change in the social applications of computing and AI. Lastly, the paper summarizes and anticipates coming debates on the controversies of how to do the right thing for the planet and other species while simultaneously observing the recognized rights of personal privacy for one species—humans. For scholarship and policy alike, the question is, at once, empirical and existential: Which kinds and degrees of surveillance, with what checks and balances, will the present human cohort be prepared to accept in the coming pivotal decades to ensure the livelihood of human and other species centuries and millennia into the future?

References

Abram, D. (1996). *The Spell of the Sensuous: Perception and Language in a More-than-Human World.* Knopf Doubleday Publishing Group.

Creutzig, F., Acemoglu, D., Bai, X., Edwards, P. N., Hintz, M. J., Kaack, L. H., Kilkis, S., Kunkel, S., Luers, A., Milojevic-Dupont, N., Rejeski, D., Renn, J., Rolnick, D., Rosol, C., Russ, D., Turnbull, T., Verdolini, E., Wagner, F., Wilson, C.,…Zumwald, M. (2022). Digitalization and the Anthropocene. *Annual Review of Environment and Resources, 47,* 479–509. https://doi.org/https://doi.org/10.1146/annurev-environ-120920-100056

Hallin, D. C., & Mancini, P. (2004). *Comparing Media Systems: Three Models of Media and Politics.* Cambridge University Press.

IPCC. (2023). *Climate Change 2023 – Synthesis Report*. https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_FullVolume.pdf

ITU. (2005). *The Internet of Things: Executive Summary.* International Telecommunication Union. http://www.itu.int/dms_pub/itu-s/opb/pol/S-POL-IR.IT-2005-SUM-PDF-E.pdf

Jensen, K. B., & Helles, R. (Eds.). (2023). *Comparing Communication Systems: The Internets of China, Europe, and the United States*. Routledge.

United_Nations. (2015). *Paris Agreement*. https://unfccc.int/sites/default/files/english_paris_agreement.pdf

Zhang, J., & Tao, D. (2021). Empowering Things With Intelligence: A Survey of the Progress, Challenges, and Opportunities in Artificial Intelligence of Things. *IEEE Internet of Things Journal, 8*(10), 7789–7817. https://doi.org/10.1109/JIOT.2020.3039359

The status quo: Urban AI and the deepening of technocentrism in urban management

Leandry Jieutsa

This paper analyzes the emergence of a latent technocentrism in urban governance in the age of urban AI. While the failures and criticisms of the smart city have highlighted the limits technocentric governance, the increasing integration of AI in cities does not mark a retreat from it, but rather its reconfiguration. By analyzing the case of Quebec, this paper shows that this regime is not based on an explicitly claimed ideology, but on a convergence of structural factors. These include the perceived benefits of AI for cities, the multi-level ecosystem dynamics, and the increasing embedding of algorithmic systems in urban management. The paper argues for a re-politicization of urban AI by strengthening participatory mechanisms (citizens' assemblies, consultations, co-design) and the internal capacities of municipalities.

Augmenters Assemble! International Financial Institutions, Data Infrastructures and Algorithms for Environmental and Social Risks

Leonard Seabrooke & Rosie Collington

This paper explores how algorithmic tools are transforming governance in international financial institutions (IFIs) through a case study that traces the development of tools for managing environmental and social (E&S) risks of development bank financing. We argue that algorithmic risk management tools constitute a new site of conflict in IFI governance. They strengthen the authority of (risk specialist) staff to intervene in the governance of credit-seeking states and firms. In the process, project/lending success is recast from ex-post development impact to ex-ante risk avoidance. Paradoxically, the dependence of these tools on the data infrastructures of asset management companies and index providers holds the potential to weaken IFI authority if these actors' methodologies dominate in the long-term. This is a new bargain: Staff gain short-term situational leverage (triggering interventions) but risk long-run dependence on proprietary data infrastructures.

The next step towards the future of work: A case of an AI-supported four-day workweek

Linea Munk Petersen

This article examines the potentials and challenges of implementing a four-day workweek in the Danish private sector, an area still underexplored in academic literature. Drawing on an ethnographic case study of Vest Administrationen A/S (Vest), which introduced a reduced-hours four-day workweek in 2022, the study explores how organizational practices are reshaped through a combination of productivity tools, employee involvement, and managerial priorities. The analysis demonstrates how the four-day workweek functions as both an organizational change process and a site of productive tension between productivity and employee well-being. While productivity dominates the discourse, employee well-being is simultaneously framed as a crucial motivation for sustaining the model. A key finding is that success hinges on tailoring the four-day week to organizational context, balancing managerial goals with employee agency. Furthermore, the integration of generative AI through a "ChatGPT initiative" illustrates how new technologies can expand the model's potential by automating tasks and reinforcing productivity while maintaining reduced working hours. The article contributes to debates on future work by highlighting how private organizations negotiate the dual imperatives of efficiency and well-being when adopting alternative work-time arrangements.

In 2024, a longer program was launched at Vest with the aim of training employees in the use of generative AI. The CEO had already been coding programs for the company as additions to externally purchased IT infrastructure but with this new initiative, employees were directly involved in further automating work tasks. Before launching the ChatGPT Initiative, the CEO shared his vision of generative AI as necessary to ensure future growth and support the four-day work week. When speaking on the possibilities of organization-wide AI literacy, he jokingly said: "*maybe in the future we could have a three-day work week!*"

Parallels can be drawn between the ChatGPT initiative and the overall implementation of the four-day work week, as both required three essential steps: 1) Management-driven or supported introduction of a reorganization of work; 2) Employee training in the use of tools to support a specific tool or system of tools and; 3) Maintenance, adjustment, and measurement of value and success. What the ChatGPT initiative adds is an expansion of already existing productivity tools, as well as a possibility for automation through generative AI, with the intention of further developing Vest's four-day workweek model. Just as quotes about productivity and well-being are visible on flat screens around the office, and the CEO repeatedly enforces adherence to the four-day week culture, he also regularly encourages further use of ChatGPT in internal newsletters and informal interactions with employees. The ChatGPT initiative is thus not separate from the four-day work week implementation but part of an organizational change leveraging a potential described as either inevitable or advantageous both outside and inside the four-day workweek literature. The four-day workweek at Vest shows AI society in miniature: a tug-of-war between automation for growth and collective demands for a better life at work. The case suggests that whether AI extends freedom or intensifies control depends on how organizations balance managerial priorities with employee agency.

The politics of AI knowledge distillation

Louis Ravn

In January of 2025, the "DeepSeek moment" brought an increasingly pervasive technique of contemporary AI production into the centre of public discussion: *knowledge distillation*. Knowledge distillation describes the transfer of knowledge from a large deep learning (DL) model, called "teacher model", to a significantly smaller model, called the "student model" – primarily used with the aim of efficiency improvements (Hinton et al., 2015; Moslemi et al., 2024). Importantly, the technique of AI knowledge distillation is increasingly widely used by the most prominent AI companies, including OpenAI, Google, and Meta. While there exists a plethora of research on AI knowledge from technical perspectives, critical scholarship has so far not attended to the technique's sociopolitical implications..

Against this backdrop, the aim of this paper is to examine the politics of AI knowledge distillation, analysing how it is entangled with power relations. To do this, we combine multiple methods and draw from a wide range of empirical material. Concretely, we analyse foundational computer science texts about knowledge distillation (cf. Amoore et al., 2023), model reports by prominent AI companies, promotional talks & webinars, and relevant press articles. In addition, we present some of our own experiments working with knowledge distillation, inspired by critical AI studies work that attends to the materialities of DL (Aradau & Blanke, 2022; Jacomy & Borra, 2024).

On this basis, our analysis develops two main insights into the politics of AI knowledge distillation. First, knowledge distillation partially reconfigures the global political economy of AI: while it facilitates the production of smaller, customized, and locally deployable AI models, thereby signalling a decentralizing turn in AI, these smaller models are still dependent upon extremely large teacher models that can only be built by Big AI (van der Vlist et al., 2024; Srnicek, 2025). Therefore, knowledge distillation may extend the principle of "monopolization through decentralization" into the AI age (cf. Blanke & Pybus, 2019). Second, the practice of AI knowledge distillation is shaped by ontological politics (Mol, 2002), here denoting micropolitical decisions that delineate the resulting models: the specific configurations of knowledge distillation enact how AI models operate in the world. As a result, knowledge distillation constitutes an important aspect of AI production pipelines whose micro-computational configuration materially shapes AI systems.

More broadly, this article indicates two new research avenues for the burgeoning field of critical AI studies (Raley & Rhee, 2023; Offert & Dhaliwal, 2025). First, it highlights the need to critically attend to brings techniques of *model compression*, of which knowledge distillation is one example. Concretely, there is an increasing range of AI model compression techniques that are growing in relevance, including quantization, pruning, and neural architecture search, each of which ought to be considered according to its own logics and politics. Second, our analysis signals the need to shift the analytical focus beyond the politics of *data synthesis* (Jacobsen, 2023; Steinhoff, 2022) towards the politics of what we may call *model synthesis*. Specifically, rather than being related only to questions around synthetic data, knowledge distillation highlights that entire AI models are increasingly synthesized within relatively self-referential circuits of synthetic production.

References

Amoore, L., Campolo, A., Jacobsen, B. & Rella, L. (2023). Machine learning, meaning making: On reading computer science texts. *Big Data & Society, 10*(1).

Aradau, C. & Blanke, T. (2022). A*lgorithmic Reason: The New Government of Self and Other*. Oxford: Oxford University Press.

Blanke, T. & Pybus, J. (2020). The Material Conditions of Platforms: Monopolization Through Decentralization. *Social Media + Society, 6*(4).

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the Knowledge in a Neural Network. *arXiv:1503.02531.*

Jacobsen BN. (2023). Machine learning and the politics of synthetic data. *Big Data & Society, 10*(1): 1–12.

Jacomy, M & Borra, E. (2024). Measuring LLM Self-consistency: Unknown Unknowns in Knowing Machines. *Sociologica, 18*(2): 25–65.

Mol, A. (2002). *The body multiple*. Durham, NC: Duke University Press.

Moslemi, A., Briskina, A., Dang, Z. & Li, J. (2024). A survey on knowledge distillation: Recent advancements. *Machine Learning with Applications, 18*.

Offert, F. & Dhaliwal, RS. (2025). The Method of Critical AI Studies, A Propaedeutic. *arXiv:2411.18833.* (accessed 02 July 2025).

Raley, R. & Rhee, J. (2023). Critical AI: A Field in Formation. *American Literature, 95*(2): 185–204.

Srnicek, N. (2025). *Silicon Empires: Silicon Empires: The Fight for the Future of AI*. Cambridge, UK: Polity Press.

Steinhoff, J. (2022). Toward a political economy of synthetic data: A data-intensive capitalism that is not surveillance capitalism? *New Media & Society, 26*(6): 3290–3306.

Van der Vlist, F., Helmond, A. & Ferrari, F. (2024). Big AI: Cloud infrastructure dependence and the industrialisation of artificial intelligence. *Big Data & Society, 11*(1).

AI-Mediated Self-Presentation: Towards a New Understanding of Authenticity in AI Society

Lucia Vovk & Alessandra Sossini

This paper challenges the dominant conceptualization of authentic self-presentation as real, genuine, or true expression of inner essence or social identity (Lehmann et al., 2019) and calls for its redefinition in light of the rise of AI-generated content within algorithmically mediated online environments.

Authenticity has long been central to identity construction and self-presentation, traditionally associated with inner truthfulness and intrinsic motivation for genuine self-expression (Banet-Weiser, 2012). However, the proliferation of artificially produced communication content complicates this understanding. Current perspectives on authentic self-presentation overlook relational dimensions of authenticity, particularly how it emerges through interactions with platform norms and the enabling and constraining logics of AI-mediated communication. This raises the question of whether the traditional definition of authenticity has transformed into a fundamentally different construct. This paper aims to analyze authenticity in the age of AI and to develop a reconceptualization that incorporates technological and social implications of the AI-mediated communication environment.

Adopting a conceptual approach, this paper first reviews dominant understandings of authenticity as constructed, performed, perceived, and synthetic. It then examines LinkedIn as an illustrative case, where AI-generated self-presentation in long-form posts are increasingly prevalent (Gillham, 2025). Integrating platform dynamics, recent advances in generative AI, and Baudrillard's (1983) notion of simulation, the preliminary analysis highlights how authenticity is evaluated through comparisons to users' personal references of realness. As distinctions between human- and AI-generated content become less perceptible (Lao et al., 2025), this evaluative process blurs the boundary between the real and artificial. The analysis further suggests that authenticity on LinkedIn operates less as inner truthfulness and more as a success strategy rooted in societal and platform expectations, simultaneously encompassing temporality, constructivity, performance, perception, and synthetic elements. Consequently, authenticity becomes a technological and social simulation, which is sustained through algorithmically reinforced repetition rather than originality. In this sense, simulated repetition – the citational and iterative use of AI-generated content (Taylor, 2022) – emerges as a new mode of authenticity that remains compelling and believable to audiences.

This paper aims to contribute a reconceptualization of authenticity as a simulated yet socially efficacious performance, shaped by public attention and platform logic rather than depth of personal expression. It also provides a framework for understanding identity construction in AI-mediated environments and invites critical reflection on the future of meaningful interaction in AI society.

References

Banet-Weiser, S. (2012). *Authentic™: The politics of ambivalence in a brand culture*. New York University Press.

Baudrillard, J. (1983). *Simulations*. Semiotext(e).

Gillham, J. (2025, October 10). Over ½ of Long Posts on LinkedIn are Likely AI-Generated Since ChatGPT Launched. *originality.ai*. Retrieved October 14, 2025, from https://originality.ai/blog/ai-content-published-linkedin

Lao, Y., Hirvonen, N., & Larsson, S. (2025). AI and authenticity: Young people's practices of information credibility assessment of AI-generated video content. *Journal of Information Science, 0*(0), 1–15.

Lehman, D. W., O'Connor, K., Kovács, B., & Newman, G. E. (2019). Authenticity. *Academy of Management Annals, 13*(1), 1–42.

Taylor, A. S. (2022). Authenticity as performativity on social media. *Springer eBooks.*

Stories for Machines: Using Vignette Methodology to Study Bias in Large Language Models

Mark Friis Hau & Andrea Borello

As large language models (LLMs) are increasingly deployed across a wide range of personal and workplace settings, they directly shape how users understand situations and make decisions. Model bias is therefore a prominent concern, ranging from overt stereotyping to subtle shifts in tone, evaluation, and recommended course of action. However, current approaches measure bias as a property of outputs, detecting deviation from statistical baselines, and cannot capture how models generate socially patterned interpretations. This paper addresses that gap by adapting qualitative vignette methodology to the study of LLMs, using short, structured scenarios with controlled variation to reveal the interpretive repertoires embedded in model outputs.

Vignettes are methodologically useful because while they are contextually rich, they permit the systematic manipulation of variables such as protagonist gender, class, or ethnicity. This makes them valuable for interpretive analysis, but also make it possible to control comparisons across models and conditions. Qualitative probing is necessary because LLMs do more than return biased outputs; they produce patterned attributions of agency, moral vocabularies, and institutional logics that together function as cultural repertoires. We therefore suggest vignette-based probing as a bridge between qualitative interpretation and statistical approaches, and as a tool for understanding the cultural imaginaries encoded in LLMs.

The vignette protocol probes multiple LLMs under deterministic settings to surface cultural repertoires and examine how they shift when key variables are subtly changed. Treating model responses as cultural texts allows close readings of how LLMs frame situations, assign responsibility, invoke particular forms of authority, and suggest specific courses of action. These readings expose recurring tropes and silences, and clarify how different models organize social meaning in divergent ways.

The paper situates this approach within sociological and STS traditions that treat technologies as culturally embedded artefacts, then outlines a transparent protocol for vignette construction, probing, and evaluation. We show how the method captures stable and interpretable patterns across identity, register, and stakes variables, contributing to both sociological understanding of AI systems and practical evaluation of model deployment.

“We are on our way”: Strategies and Struggles in Developing Sustainable AI Communication Practices

Martina Skrubbeltrang Mahnke

AI-driven communication practices have become the backbone of modern work life and are commonly used with the presumption of their sustainability. However, recent studies have started to question the digital sustainability paradigm, arguing that the damages inflicted on the natural world remain largely unnoticed, uncommunicated, and undertheorised (Kuntsman & Rattle, 2019; Mahnke, forthcoming; Sharma & Dash, 2022). Especially with the rapid growth of generative AI as professional communication assistants, this estimate is expected to grow exponentially within the next few years (Patterson et al., 2021; Saenko, 2023). Already today, organisational communication activities account for the majority of data centre consumption (Batmunkh, 2022), and this percentage will rise with the rapidly increasing use of AI applications (Brunton, 2022). In light of the severity of the climate crisis, a critical assessment of the use of AI communication in organisational contexts is thus crucial in order to support necessary actions in the transition to green(er) communication practices (Christensen, 2025).

In this relation, a growing field of professionals have been tasked to manage organisations’ digital sustainability efforts. Yet little empirical research has examined how these practitioners experience and negotiate digital responsibility within organisational settings. This article addresses this gap through a qualitative interview study with 20 professionals working with digital sustainability in public and private organisations in Denmark. Drawing on a thematic analysis (Braun & Clarke, 2006), the study identifies three interrelated themes that shape how digital responsibility is enacted and constrained in organisational settings: 1) individual vs. organisational responsibility, 2) present vs. future orientation and 3) systematic approaches vs. personal agency.

The analysis shows how digital sustainability managers perceive their own roles as simultaneously empowering and limiting. Many enter their positions with a strong personal commitment to environmental values; however, sooner rather than later, they find themselves operating within organisational structures that prioritise differently. Thus, they become “boundary managers” who first and foremost translate sustainability ideals into organisationally legitimate actions, often directed towards future goals and orientations rather than present ones.

Overall, the study shows that digital responsibility is an ongoing negotiation between competing organisational goals and temporalities. By adopting a perspective of professional working with digital sustainability, the study contributes to scholarship at the intersection of digital communication and sustainability studies.

References

Batmunkh, A. (2022). Carbon Footprint of The Most Popular Social Media Platforms. *Sustainability, 14*(4). https://doi.org/10.3390/su14042195

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brunton, H. (2022). *Digital carbon footprint: The environmental impact of digital transformation.* The Knowledge Exchange Blog. https://theknowledgeexchangeblog.com/2022/01/17/digital-carbon-footprint-the-environmental-impact-of-digital-transformation/

Christensen, E. (2025). The socio-ecological costs of AI: Toward socially responsible and sustainable communication practices. *Public Relations Inquiry*, 2046147X251377853. https://doi.org/10.1177/2046147X251377853

Kuntsman, A., & Rattle, I. (2019). Towards a Paradigmatic Shift in Sustainability Studies: A Systematic Review of Peer Reviewed Literature and Future Agenda Setting to Consider Environmental (Un)sustainability of Digital Communication. *Environmental Communication, 13*(5), 567–581. https://doi.org/10.1080/17524032.2019.1596144

Mahnke, M. S. (forthcoming). The eco-materiality of digital platforms. In S.-M. Laaksonen, M. Pantti, & O. Dovbysh (Eds.), *Platforms and the Planet: Big Tech, Digital Platforms and Environmen-tal Responsibility.* Emerald Group Publishing Limited.

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). https://doi.org/10.48550/arXiv.2104.10350

Saenko, K. (2023). *Is generative AI bad for the environment? A computer scientist explains the carbon footprint of ChatGPT and its cousins.* The Conversation. http://theconversation.com/is-generative-ai-bad-for-the-environment-a-computer-scientist-explains-the-carbon-footprint-of-chatgpt-and-its-cousins-204096

Sharma, P., & Dash, B. (2022). The Digital Carbon Footprint: Threat to an Environmentally Sustainable Future. *International Journal of Computer Science and Information Technology, 14*(03), 19–29. https://doi.org/10.5121/ijcsit.2022.14302

Exploring the Practical Realities of Generative AI in Danish Public Administration

Mathilde Buskbjerg Madsen, Troels Schmitto & Kasper Edwards

There is a strong wish from the Danish government to utilise generative artificial intelligence (genAI) in the public sector to increase administrative efficiency and reallocate resources. However, the adoption of genAI in the public sector appears to be disorganised, bottom-up, and lacking coherent strategies. This study examines how employees in Denmark's public administration experience and adopt genAI in their everyday work, with a focus on its implications for tasks, well-being, and managerial practices. Using an exploratory, inductive qualitative approach, the research draws on focus groups with employees and semi-structured interviews with managers, interpreted through thematic analysis. The findings reveal that adoption is initiated mainly from the bottom up, with employees normalising the use of generative AI after initial hesitation. They report substantial benefits, including time savings, improved quality of outputs, and the ability to complete tasks previously considered unmanageable. At the same time, employees express uncertainty about appropriate applications, shifts in workplace dynamics, and concerns related to redundancy, climate impact, and technological dependencies. Managers provide encouragement and guidance, yet individual initiatives and peer learning emerge as the primary drivers of adoption. By capturing these dynamics in Denmark's highly digitalised public sector, the study highlights both the opportunities and challenges of generative AI adoption, offering insights that may foreshadow developments in other public administrations.

Grounding as Deflation: Turning the "AI Ghost" into a Stage for Friction

Matilde Ficozzi, Mathieu Jacomy, Anders Kristian Munk & Dario Rodighiero

Controversy mapping typically aims to visualize actor positions in heated sociotechnical debates around a matter of public concern (Venturini, 2010). However, in the case of the Grounding AI Map—a 100m² walkable visualization of 2 million scientific papers—this approach encounters a paradox: the data reveals no "AI controversy." Instead, it shows a diversified, largely uncontroversial landscape of applications and optimization tasks, from traffic monitoring to medical diagnostics (Munk et al., 2024).

Grounding AI thus shows that the point of controversy mapping does not have to be exposition, but could also be *deflation:* Here, a mechanism that breaks the generalized, abstract category of AI down into its specific and often mundane components. Public discourse frequently treats AI as a ghostly monolith that is everywhere and nowhere, sparking feelings of doom and gloom or utopian hype. By forcing this ghost into a physical, walkable space, the map effectively materializes AI and reveals it as fragmented into thousands of specific technical applications (Ficozzi et al., 2025).

We argue that this substantiation process is a necessary prerequisite for political engagement with the matter of AI. Drawing on Haraway's (1988) critique of the "God trick" of seeing everything from nowhere, the exhibition deliberately takes away the overview and forces participants to experience partial perspectives on specific AI topics. Through a series of public events where experts doing research around AI were invited to talk about their work, we observed that the map effectively breaks down the general category of "AI", allowing participants to engage with specific technologies rather than abstract ideas.

The map shifts from being a visual representation to a stage where we welcome friction coming from practices grounded in research (Stengers, 2018). It becomes a physical stage where the 'uncontroversial thingness of AI' (Suchman, 2023) meets the specific, situated concerns of the human actors standing upon it and in relation to equally specific technoscientific problems. In conclusion, we argue that to truly map the controversies of AI in society, we must first deflate the term "AI" itself, moving from a debate about a singular future to an effort to understand the multiplicity of our present.

References

Ficozzi, M., Jacomy, M., Rodighiero, D., Beaulieu, A., & Munk, A. K. (2025). Grounding AI Map: The Consequences of Living with the Trouble of an Irreductionist Map. *Design et abstractions, Revue Design Arts Medias*. https://journal.dampress.org/issues/design-et-abstractions/grounding-ai-map-the-consequences-of-living-with-the-trouble-of-an-irreductionist-map

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies, 14*(3), 575–599. https://doi.org/10.2307/3178066

Munk, A. K., Jacomy, M., Ficozzi, M., & Jensen, T. E. (2024). Beyond artificial intelligence controversies: What are algorithms doing in the scientific literature? *Big Data and Society, 11*(3). https://doi.org/10.1177/20539517241255107

Stengers, I., & Muecke, S., (TRANS.) (2018). Another Science is Possible: *A Manifesto for Slow Science*. Polity Press.

Suchman, L. (2023). The uncontroversial 'thingness' of AI. *Big Data & Society, 10*(2). https://doi.org/10.1177/20539517231206794 (Original work published 2023)

Venturini, T. (2010). Diving in magma: how to explore controversies with actor-network theory. *Public Understanding of Science, 19*(3), 258–273. https://doi.org/10.1177/0963662509102694

UK AI Governance at a Crossroads: Charting the Path Ahead

Mehmet Bilal Unver & Sandra Odusola-Stevenson

Governance of Artificial Intelligence (AI) in the UK is still in its formative stages, earmarking the policy goal of "proportionate and pro-innovation regulatory framework" (UK White Paper 2023). Based on this political framing, the UK has adopted a cross-sectoral, principles-based approach, aiming to balance regulatory flexibility with innovation while addressing associated risks (UK White Paper 2023). Rather than enacting comprehensive AI-specific legislation, the UK government advocates for sectoral enforcement of overarching principles (e.g., safety, transparency, fairness, accountability, and contestability) by existing regulators (UK White Paper 2023). Key developments in this trajectory include the *National AI Strategy* (Office for Artificial Intelligence, 2021), the establishment of the *AI Safety Institute* (UK White Paper 2023), and the release of the *AI Opportunities Action Plan* (2025). Collectively, these measures signal a commitment to a pro-innovation, decentralised and tailorable governance model. However, whether the UK should maintain this regulator-led model or adopt a more uniform, horizontal legislative framework remains contested. To address this question, this paper examines how AI governance currently operates across three UK domains, professional legal services, healthcare and education, each of which is regulated according to its sector-specific needs and priorities. It then contrasts these findings with the European Union's (EU) horizontal and unified approach under the EU AI Act (2024). This comparative analysis explores the distinctive features, tools and mechanisms of each regulatory model, with the aim of drawing insights for shaping the UK's future trajectory of AI governance. Marking distinction to the literature (Al-Maamari, 2025; Haie et al, 2024; Batool et al, 2024 Unver and Roddeck, 2024; Nannini, Balayn and Smith, 2023), this paper focuses on the widely acknowledged principles of AI governance (e.g., fairness, transparency and accountability) within both the UK and the EU frameworks and how they are operationalised in each jurisdiction. The paper identifies several challenges confronting the UK's approach, including regulatory fragmentation, potential inconsistencies across sector-specific rules and latent costs arising from legal uncertainty and complexity. It concludes that, while the UK does not need to radically overhaul its governance model, the status quo requires enhanced coordination, consistency and sectoral alignment to ensure effective and sustainable AI governance.

References

AI Opportunities Action Plan, 2025. Department for Science, Innovation & Technology. *AI Opportunities Action Plan*. https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan

Al-Maamari, A., 2025. *Between innovation and oversight: A cross-regional study of AI risk management frameworks in the EU, U.S., UK, and China. arXiv*. https://arxiv.org/abs/2503.05773

Batool, A., Zowghi, D. and Bano, M., 2025. AI governance: a systematic literature review. *AI Ethics, 5*, pp.3265–3279. https://doi.org/10.1007/s43681-024-00653-w

Haie, A.G., Golodny, A., Shenk, M., Goodwin, E. and Arethusa, V., 2024. Comparative analysis of the EU, US and UK approaches to AI regulation. *Steptechtoe,* 30 April 2024 https://www.steptoe.com/en/news-publications/steptechtoe-blog/a-comparative-analysis-of-the-eu-us-and-uk-approaches-to-ai-regulation.html

Luca Nannini, Agathe Balayn, and Adam Leon Smith, 2023. Explainability in AI Policies: A Critical Review of Communications, Reports, Regulations, and Standards in the EU, US, and UK. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT'23)*. Association for Computing Machinery, New York, NY, USA, 1198–1212. https://doi.org/10.1145/3593013.3594074

National AI Strategy, 2021. Department for Science, Innovation & Technology; Office for Artificial Intelligence. *National AI Strategy: AI Action Plan.* https://www.gov.uk/government/publications/national-ai-strategy-ai-action-plan

Office for Artificial Intelligence, 2021. *UK National AI Strategy.* https://www.gov.uk/government/publications/national-ai-strategy

Palladino, N., 2023. A 'biased' emerging governance regime for artificial intelligence? How AI ethics get skewed moving from principles to practices. *Telecommunications Policy, 47*(5). https://doi.org/10.1016/j.telpol.2022.102479

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139, and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797, and (EU) 2020/1828 (EU AI Act).

UK Department for Science, Innovation & Technology; AI Safety Institute, 2024. *AI Safety Institute: Overview*. https://www.gov.uk/government/publications/ai-safety-institute-overview

UK White Paper, 2023. Department for Science, Innovation & Technology & Office for Artificial Intelligence. *A Pro-Innovation Approach to AI Regulation* (White Paper, CP 815). https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach

Unver, M. B. and Roddeck, L., 2024. Ethics governance of AI for the legal sector: Building up a holistic policy approach. *Journal of AI Law and Regulation, 1*(2), 177–198. https://doi.org/10.21552/aire/2024/2/5

Participatory Contestation under the EU AI Act

Michael Strange & Eduardo Gill Pedro

The EU AI Act is presented as an instrument to foster "trustworthy" and "human-centric" AI that safeguards fundamental rights and democracy in Europe. However, it is at its core a market integration instrument. As such, it deploys a regulatory approach that derives not from human rights law, but from product safety legislation. The Act institutes a novel collaborative governance framework that places primary responsibility for risk assessment and mitigation on providers and deployers, operating largely through self-assessment, harmonised standards, and soft law. In this paper we propose to investigate what space is left in the governance framework instituted by the Act for individuals and civil socieity organisations to contest decisions about AI.

We approach contestation not only as an ex post, individual right to remedy ("remedial contestation"), but as "participatory contestation": the collective, ex ante ability of affected actors to shape, resist, or block decisions. The focus is on the ability to contest decisions about AI – decisions about risk classification, the design of risk management systems, human oversight and deployment safeguards. Building on democratic theory and critical perspectives inspired by Habermas, Forst, as well as Ranciere and Mouffe, we argue that democratic legitimacy in AI governance depends not simply on representation or consultation, but on meaningful capacities to challenge and revise power-wielding decisions. We contrast this with more instrumental "AI ethics" approaches that frame participation primarily as a tool for improving trustworthy or effective AI.

Methodologically, the paper combines contextual and teleological legal analysis of the AI Act's governance architectue, and how it fits within the broader context of the EU digitial single market regulation with a political mapping of contestatory spaces that both directly follow, as well as be logically inferred, from the institutional set-up provided by the Act.

With this mapping we will highlight the different sites within the governance framework of the Act where individuals and civil society organisations can contest decisions about AI. We show how the Act's internal-market logic narrows the scope for national democratic experimentation and local contestation, and relocates crucial decisions into private and expert-dominated arenas, insulated from democratic accountability.

The paper concludes that, while the AI Act creates new formal avenues for engagement, it reconfigures and in important respects constrains participatory contestation, shifting power over AI governance away from democratically accountable institutions and towards providers, standards bodies, and EU-level technocratic structures. We outline what additional legal and institutional safeguards would be needed if the Act is to live up to its promise of protecting democracy. The analysis concludes by identifying where societal actors engaged in regulatory contestation essential to democracy, such as civil society, might focus their energies in this new governance landscape.

Uncertainties of GenAI in university – students' study practices

Mikala Hansbøl

Surveys among HE students in Denmark, Norway and Sweden point to a paradox: university students increasingly use GenAI, GenAI shape study practices, and the students lack guidance on GenAI in education. Pedersen, Hansbøl and Fajkovic's (2025) survey of Humanities and Theology students (Spring 2024) show that 46 % have not experienced teaching with a focus on GenAI, 25 % a few times and 20 % once. Engagements with GenAI chatbots is equally diverse. 7 % use them daily, 29 % weekly, 18 % monthly, 30 % more rarely, and 15 % never. These findings highlight the need to explore the varied relationships between students and GenAI in practice.

GenAI chatbots, based on Large Language Models, raise societal and educational concerns that must be understood in relation to students' purposes for engaging with them. Only 39 % of students who encountered GenAI in teachings received an introduction to its workings. Yet, 63 % of users find chatbots useful for learning, and 48 % believe they save time and increase efficiency. Surveys, however, offer only *a group-level view*; they cannot capture individual interests in or (dis-)engagements with GenAI chatbots for various learning purposes. Depending on objectives, GenAI engagements can be both meaningful and problematic. Bruun et al. (2025) warn that weaker students risk losing touch with their own knowledge, critical thinking, creativity and contributions, when interacting with persuasive chatbots. The authors call for empirical studies centered on the students' perspectives and everyday practices.

This paper asks: *how does GenAI chatbots become part of and take part in moving diverse university students' study practices?* Empirically, it draws on qualitative interviews with 12 BA and MA level students from the Faculties of Humanities and Theology at University of Copenhagen who participated in the survey. Students were selected based on criteria such as payment status, study level, and frequency of use.

Theoretically and analytically, the paper builds on Latour's (2005) five sources of uncertainty and draw inspiration from Hansbøl (2010) to engage with GenAI chatbots as partially existing actors. GenAI chatbots are presented as actors that emerge through situated connections with university students' interests, (dis-)engagements and study practices. The interviews create new entrance points allowing both students and chatbots to "speak" and transform the debates on GenAI in university education and student's learning processes. Drawing on Schraube (2024) the paper presents portraits of the "what", "how" and "why" of students' (dis-)engagements with GenAI for learning. Through these lenses the paper opens discussions on the ontological multiplicities (2002) of educational opportunities and challenges posed by GenAI, grounded in students' own descriptions of study practices with and without chatbots.

References

Annemarie Mol. *The Body Multiple: Ontology In Medical Practice.* Durham, NC: Duke University Press. 2002.

Bruun, M. H., Krause-Jensen, J., & Hasse, C. (2025). Skrivning, læsning, tekst og refleksiv tænkning med generativ AI på humanistiske videregående uddannelser. *Tidsskriftet Læring og Medier (LOM), 18*(31).

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory.* New York Oxford University Press.

Hansbøl, M. (2010). *Researching relationships between ICTs and education: Suggestions for a science of movements.* Danish School of Education, Aarhus University.

Schraube, E. (2024). *Digitalization and learning as a worlding practice: Why dialogue matters.* Routledge.

Pedersen, A., Hansbøl, M., & Fajkovic, M. (2025). *Studerendes erfaringer med, brug af og perspektiver på Generativ AI: ved Det Humanistiske og Teologiske Fakultet, Københavns Universitet, foråret 2024.*

Public and Hidden Transcripts in China's Algorithmic Society: Platform-Based Visibility, Control, and User Tactics

Ming Cheng

In 2019, Chinese leaders claimed "let mainstream values ride the algorithm," marking an official transformation of China's digital censorship. For an authoritarian regime that relies on heavy censorship and propaganda, manipulating speech via algorithms visibility appears ideal. However, when the CCP places its political will above the logic of algorithmic visibility, there is always a tension between the power and the people if Chinese users attempt to escape visibility traps carefully engineered by the government and platforms. King's influential research shows that the CCP does not censor all online dissent but focuses on speech that may mobilise collective action. Recent studies indicate that Chinese users can still use wordplay to evade censorship. These studies help establish a fact: algorithmic logic may not fully match the political intentions of CCP elites, and such mismatches provide opportunities for creative and courageous dissenters to criticise public authority. Therefore, this study explores how users express criticism under the dual constraints of platform affordances and politically imposed algorithmic logic.

This study uses James Scott's concepts of public transcript and hidden transcript as the theoretical framework: the former refers to performative compliance displayed in front of power, and the latter to backstage resistant discourse. In China's algorithmic society, instead of relying mainly on blunt censorship, political elites can require platforms to influence the visibility of the public transcript through recommendation and ranking, while the hidden transcript becomes strategically crafted expressions that remain visible to peers but invisible to algorithms. If we treat ranking and recommendation as the new "stage managers" of public and hidden transcripts, algorithms reshape their form and circulation on Chinese social media.

Methodologically, this study compares Weibo, a highly politically sensitive platform, with Rednote, a lifestyle-oriented platform. The purpose is not simply to fill a gap in platform comparison but to reveal how political control is differentially materialised through distinct platform architectures, and how users' coping strategies adapt accordingly. Through platform observation and interviews with users, the study finds that distinct performances on these two platforms reflect the CCP's capacity and different objectives in integrating algorithms into governance. On Weibo, the CCP heavily curates algorithmic ranking and agenda distortion, shaping an authority-first agenda feed with rapid closure; on Rednote, the CCP may be satisfied with the platform's lifestyle orientation, but the platform still strictly observes political compliance, leading to suppressed search but visible and often ephemeral feeds. Despite these constraints, Weibo users switch platforms or deep dive into residual chains, whereas Rednote users employ facet framing and polite long-form commenting.

These findings refine Scott's concepts: the public transcript is not a single space but multiple platform-specific visibility fields, and the hidden transcript is not simply private resistance but depends on platform architecture and circulates through semi-visible, coded discursive spaces embedded

within different publics while escaping algorithmic capture. This study advances understanding of visibility governance and digital authoritarianism, and offers new insights into domination and resistance at a time when Western democracies are also considering stronger algorithmic governance and digital regulation.

Making sense of AI futures: Lessons from "Reading the Apocalypse"

Monica Porzionato & Otto Hedenmo

Despite the long-standing presence and use of artificial intelligence (AI) by both private and public organizations, recent advancements in algorithmic technologies – including large language models (LLMs), context-aware spatial sensors, and wearable devices – have only recently been made available to the general public. Consequently, many of us are currently experimenting with these technologies and learning how to incorporate them into our daily tasks. Whether we are aware of it or not, AI gadgets and software are gradually infiltrating our everyday activities. What we investigate in this study is how AI as a phenomenon influence our ability to make sense of our changing world.

Artificial intelligence is frequently regarded as having the potential to exert considerable influence on organizational structures, and societal frameworks in general (Willcocks, 2020). As a force of societal transformation, AI is often depicted as either a saviour or a villain, which endows it with certain traits that inform how we also interpret and speak about change. At this societal and collective level, research is needed to understand AI's impact on our ability to debate (Just, 2024), collaborate (Koschman, 2020), imagine and narrate common futures (Christensen, 2025). This knowledge is particularly important in light of the ongoing deterioration of geopolitical orders, rising inequality, and mounting climate catastrophes, all of which underscore the imperative for the reinforcement of democratic institutions and values (Seyd, 2025).

The present study explores the ways in which artificial intelligence and apocalyptic thinking converge. In other words, how does AI figure into current apocalyptic imaginations and expectations? What are the potentialities of such a conjunction? The concept of the apocalypse is informed by divergent approaches to its conceptualization, whether as a future event or a present condition (Cassegård & Thörn, 2018). Indeed, both conceptualizations invite the performance of different organizing practices (Dalhman et al., 2025). To explore this topic, we conduct a preliminary qualitative analysis of data collected during three meetings of a book club that we organized at the main public library in Jönköping, Sweden. The book club, titled "Reading the Apocalypse," provides a forum for strangers to critically and collectively discuss various understandings of the apocalypse found in old and new literary works, bringing to the forefront issues such as social inequality, natureculture relations, and the defense of social welfare. In this social setting, the objective is to explore how artificial intelligence emerges as a topic of discussion and how AI technologies are understood, for example, as either facilitators or solutions to apocalyptic processes. The findings of this study can enrich our understanding of the present capabilities and constraints of artificial intelligence, as well as shed light onto the collective aspirations for an AI society.

References

Cassegård, C. and H. Thörn (2018) 'Toward a Postapocalyptic Environmentalism? Responses to Loss and Visions of the Future in Climate Activism', E*nvironment and Planning E: Nature and Space, 1*(4): 561–578.

Christensen E. (2025). Organizing Expectations: The Constitutive Dynamics of AI Imaginaries. *Human Communication Research, 51*(4): 215–222

Dalhman, S., Just, S. du Plessis E.M., and Husted E. (2025). Organizing for Apocalypse. *Ephemera. theory & politics in organization, 25*(1).

Just S.N. 2024. *Controversial Encounters in the Age of Algorithms. How Digital Technologies are Stifling Public Debate and What to Do About It.* Bristol University Press

Koschman, M. and Sanders, M. (2022). *Understanding Nonprofit Work – A Communication Perspective.* Wiley-Blackwell

Seyd, B. (2025). A crisis of crisis management: The polycrisis, the organization of pessimism, and the defensive turn. *European Journal of Social Theory, 28*(4): 644–671

Willcocks, L. (2020) 'Robo-Apocalypse cancelled? Reframing the automation and future of work debate', *Journal of Information Technology, 35*(4): 286–302.

Navigating AI Controversies: Digital Professionals' Justifications in Everyday Work

Morten Boesen

This paper explores how digital professionals legitimize their use of AI by mobilizing and juggling multiple, and often conflicting, regimes of justification. While AI in software development and digital production has received considerable scholarly attention (Banh et al., 2025; Kim et al., 2024; Russo, 2024), little work has examined the critical capacity of digital professionals themselves – how they actively question, challenge, and justify AI use in situated work practices. This gap is notable because processes of justification are central to current controversies surrounding AI in everyday professional work.

Empirically, the study draws on eight critical incident interviews (Chell, 2004; Flanagan, 1954) with experienced digital professionals identified as early and advanced adopters of AI. These interviews focus on concrete moments in which participants struggled, succeeded, or hesitated in their AI use – moments where justification became unavoidable. Across these incidents, we coded how participants framed their choices, which justificatory regimes they activated, and how they positioned themselves in relation to professional norms, technological limitations, organizational expectations, and broader societal concerns.

Using the justificatory regimes developed by Boltanski & Thévenot (2006), this paper investigates the tensions, compromises, and justificatory work that digital professionals engage in as they determine what counts as "good" or "bad" AI use in their professional practice. Digital professionals represent a particularly well-suited domain for such analysis: the field is among those most visibly transformed by AI tools, and its practitioners are among the earliest professionals confronted with the need to justify the adoption, limits, and consequences of AI in their workflows. Many face dilemmas that force them to articulate not only how AI works for them, but also why its use is appropriate, responsible, or necessary. This paper traces these dilemmas and justificatory regimes through which they are resolved or remain unresolved.

The analysis identifies a range of justificatory regimes that digital professionals mobilize when legitimizing AI use. The industrial regime appears in reference to productivity gains and the technical efficiency of AI (e.g. boilerplate coding or enforcing correct syntax). The market regime emerges in cost-benefit calculations about model performance, token pricing, and value for money. The civic regime is prominent in criticisms of Big Tech, concerns about monopolization, and arguments for open-source alternatives. The domestic regime surfaces in questions of trust, reliability, and the authority of human expertise over machine output. The inspired regime becomes visible in moments of fascination, creativity or "wow effects". Finally, the green regime appears when participants reflect on the environmental impact of AI.

These justificatory regimes do not exist in isolation. Digital professionals frequently navigate clashes between them, for example between industrial efficiency and civic resistance to monopolization, or

between inspired enthusiasm and domestic caution. Sometimes they bridge regimes through compromise; at other times, tensions remain unresolved and shape ongoing ambivalence toward AI.

The paper suggests that digital professionals' everyday reasoning is neither purely technical nor purely pragmatic: it is deeply evaluative and shaped by the plural justificatory regimes available to them. This offers a conceptual contribution to understanding how AI practices are justified in situ, and an empirical contribution to debates about the social and moral controversies that define AI's place in contemporary work.

Acknowledgement

The author acknowledges Benjamin Hughes, Lasse Vogelsang, Kirsten Marie Grønborg, Heidi Topholt and Klaus Mandal Hansen for their contributions to the data collection.

References

Banh, L., Holldack, F., & Strobel, G. (2025). Copiloting the future: How generative AI transforms Software Engineering. *Information and Software Technology, 183*, 1–14.

Boltanski, L., & Thévenot, L. (2006). *On justification: economies of worth.* Princeton University Press.

Chell, E. (2004). Critical Incident Technique. In C. Cassel & G. Symon (Eds.), *Essential Guide to Qualitative Methods in Organizational Research* (pp. 45–60). SAGE Publications Ltd.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327–358.

Kim, Y. W., Cha, M. C., Yoon, S. H., & Lee, S. C. (2024). Not Merely Useful but Also Amusing: Impact of Perceived Usefulness and Perceived Enjoyment on the Adoption of AI-Powered Coding Assistant. *International Journal of Human-Computer Interaction, 10*(41).

Russo, D. (2024). Navigating the Complexity of Generative AI Adoption in Software Engineering. *ACM Transactions on Software Engineering and Methodology, 33*(5),

What Employees Expect from AI: Alignment, contradictions, and coexistence within a plurality of AI expectations

N. Edh, O. Hedenmo, M. Riveiro, A. Engström, G. Machado, D. Pittino

AI is currently viewed as the next major technological breakthrough for organizations. The range of areas, professions, and practices that could be improved by AI assistance or automation is described as overwhelming. However, this wide array of possibilities also brings a variety of employee expectations about how AI will change organizational work. Voiced expectations influence adoption processes by both reflecting and shaping certain relational, belief-driven dynamics. To understand how these processes are influenced, this study examines AI expectations held by employees in three organizations currently adopting AI. The study is based on a thematic analysis of fifteen focus groups and shows how these expectations shape: (1) a growing desire to move from exploring AI to establishing routines and regulations, (2) emerging conflicts related to both the violation and fulfillment of AI promises, and (3) how dynamics and unpredictability in the AI field require organizations to adapt to shifting trends and innovations.

OSINT, Content Moderation and Global Security Governance

Nadia Jude

Large technology companies are key actors in global security assemblages (Borelli 2021) and content moderation can be understood as a novel public-private security practice co-produced by a broad range of social, legal and technical mechanisms across sites and along supply chains (Bellanova and de Goede 2022). Indeed, the work of content moderation has long been layered and complex; shaped by a combination of public shocks, changing labor conditions, technological advancements, online safety regulations, and geopolitical pressures (Douek 2022). More recently, a whole host of new actors, such as trusted flaggers and trust and safety vendors, have entered the foray (Wright 2025), strengthened and emboldened by a complex interplay between technical and legal infrastructures (Cohen 2019).

This paper argues that Open-Source Intelligence (OSINT) professionals are an under-recognised 'social infrastructure' for content moderation (Khalil 2023), and that attentiveness to the changing nature of their workplace practices across disparate organisational sites can offer important insights into current and emerging configurations of global security governance. It does so by offering preliminary insights from fieldwork conducted at Tech Against Terrorism (TAT), a UN-affiliated business that offers both tools and expertise to help companies moderate 'terrorist and violent extremist content' (TVEC) on their services. Combining an infra-legalities analysis of TAT's product and service offerings (Sullivan 2025), with interviews and ethnographic fieldwork observations acquired via my involvement in TAT's daily workplace routines (Schatz 2009), this paper makes three key contributions to the platform governance and critical security studies literatures.

First, I demonstrate how technological developments such as Large Language Models (LLMs) are shaping moderation workflows, which in turn, are reconfiguring accountability practices around TVEC governance in surprising and significant ways. For example, TAT have built 'intelligence assessment products' using LLMs to help customers determine *when* content has been removed after it has been flagged. Integrated into existing products, this functionality enables various actors involved in moderation workflows to see trends in content decisions *across* platforms. Commonly described by customers as TAT's 'value add', I show how LLMs are facilitating a unique form of 'network and standards-based governance' that is reconfiguring global security decisions in important ways (Cohen 2019).

Second, I highlight how new legal categories are prompting TAT to chase new business opportunities, such as the Trusted Flagger Platform (TFP). By analysing technical affordances and constraints of the TFP online platform in line with Digital Services Act (DSA) and Online Safety Act (OSA) legal provisions, I argue that the interplay between new regulations and technological developments are spawning new commercial partnerships that are expanding the scope for what counts as TVEC (Sullivan 2025). The TFP case offers useful insights into how emerging actors assign value and draw boundaries around TVEC in this growing transnational field.

Finally, I theorise OSINT professionals as an under-recognised 'social infrastructure' for content moderation, enabling very particular kinds of relations between knowledge and practice (Khalil 2023), yet 'contaminated' by their former professional histories and imaginaries of the broader OSINT field (Tsing 2015). Part of a larger project, this paper sets the scene for future work to examine OSINT practices across a range of disparate sites entangled with global security governance, such as police intelligence, war crimes documentation and investigative journalism.

References

Bellanova, R. & de Goede, M. (2022). Co-Producing Security: Platform Content Moderation and European Security Integration. *Journal of Common Market Studies, 60*(5), 1316–1334. https://doi.org/10.1111/jcms.13306

Borelli, M. (2021). Social media corporations as actors of counter-terrorism. *New media and society, 25*(11), 2877–2897. https://doi.org/10.1177/14614448211035121

Cohen, J. (2019). *Between truth and power: The legal constructions of informational capitalism.* Oxford University Press.

Douek, E. (2022). Content moderation as systems thinking. *Harvard Law Review, 136*(2), 526–607.

Khalil, H. M. (2023). Lawyers as infrastructures: mediations, blockages, and new possibilities in grassroots movements. *Journal of Law and Society. 50*(1). 231–250. https://doi.org/10.1111/jols.12422

Schatz, E. (2009). *Political Ethnography: What immersion contributes to the study of power.* University of Chicago Press.

Sullivan, G. (2025). Algorithmic governance of 'terrorism' and 'violent extremism'. *London Review of International Law, 13*(1). 47–75. https://doi.org/10.1093/lril/lraf005

Tsing, A. L. (2015). *The mushroom at the end of the world: On the possibility of life in capitalist ruins*. Princeton University Press.

Wright, L. (2025). The Salesforce of safety: Software vendors as infrastructural/professional nodes in the field of online trust and safety. *Platforms & Society, 2*(1), 1–13.

Radical Spatial Restructuring: Planning With AI, Public Discontent, and the Politics of Emerging Spatial Imaginaries

Nikola Gjorgjievski, Natalie Marie Gulsrud, Anders Hermund & Gustavo Ribeiro

Planning for future spatial and land-use change represents a central challenge of the green transition (Concito, 2023), yet the societal and environmental implications of such planning remain only partially understood (Blok, 2025). In Denmark, the recent green tripartite agreement exemplifies this tension. While the agreement seeks to accelerate renewable energy, infrastructure development, and spatial restructuring, it has also contributed to increasing "green discontent." Local communities have expressed fatigue, mistrust, and frustration regarding decisions perceived as opaque, top-down, or socially uneven (Stahl, Ellersgård, & Mulvad, 2025; Bourdin, Molica & Marques Santos, 2025). In this context, generative artificial intelligence is increasingly promoted as a tool to visualise and communicate future spatial transformation of neighbourhoods, landscapes, and of everyday life (Duray, Gulsrud, Burgos-Thorsen, forthcoming; Hermund, Ribeiro, & Gjorgjievski, 2025). Images of future scenarios have the capacity to shape our deeply held, collective understanding of socio-spatial futures. The concept of "spatial imaginaries", defined as socially embedded manifestations (Davoudi et al., 2018; Davoudi & Machen, 2022), is used in this paper to discuss the potential of generative AI in articulating common aspirations and identities. By engaging with the concept of spatial imaginaries, the authors critically discuss how these technologies operate within broader infrastructures of techno-capitalism that influence not only what can be planned and visualized, but also which futures are perceived as legitimate, desirable, or even conceivable (Raymond et al., 2025).

This paper conceptualises generative artificial intelligence as both an infrastructure *of* imagination and an agent *of* infrastructural imagining (Gandy, 2022; Graham & Marvin, 2022; Cugurollo et al., 2024). The analysis critically examines how globally trained models, shaped by partial datasets and platform logics, tend to reproduce dominant "sustainable city" aesthetics. This process often sidelines situated identities, ecologies, and social values that are essential for navigating Denmark's contested green transition (Raymond et al., 2025; Klein & D'Gnazio, 2024; Gulsrud, 2018; Gulsrud et al., 2018). In discussing the use of generative artificial intelligence in planning processes, this paper discusses how representational biases risk amplifying existing discontent by providing abstractions rather than recognising lived experiences, territorial attachments, or local forms of knowledge (Williams et al., 2024; Gulsrud et al., 2018).

To address these challenges, this paper contends that planning with artificial intelligence must explicitly engage diverse publics through participatory, community-rooted practices that prioritise transparency, plurality, and democratic authorship. Artificial intelligence should not serve to accelerate predetermined visions, but rather as a democratic tool that facilitates new spaces for contestation, negotiation, and collective imagination. *This approach is currently being explored through a prototype interactive installation in the exhibition "Imagining the Future" at the Royal Danish Academy, where audiences interact with local images and co-create future scenarios using generative AI* (Hermund, Ribeiro, & Gjorgjievski, 2025). In Denmark and internationally, fostering reflexive plan-

ning processes that include community-rooted spatial imaginaries is essential for rebuilding trust and enabling more just and inclusive green transitions.

References

Blok, A. (2025) 'Contesting eco-modernist hegemony in Denmark? Green reform nexus and transformative climate advocacy in an established environmental state', *Environmental Politics.* doi:10.1080/09644016.2025.2476276.

Bourdin, S., Molica, F. and Marques Santos, A. (2025) *Too much or not enough? The dual nature of green discontent and its geography*. JRC Working Papers on Territorial Modelling and Analysis, No. 04/2025. European Commission. JRC141349

CONCITO (2023) *The land use of Denmark: the future of Denmark.* Copenhagen: https://concito.dk/en/udgivelser/danmarks-arealer-danmarks-fremtid

Cugurullo, F., Caprotti, F., Cook, M., Karvonen, A., McGuirk, P. and Marvin, S. (2024) 'The rise of AI urbanism in post-smart cities: A critical commentary on urban artificial intelligence', *Urban Studies, 61*(6), pp. 1168–1182. doi.org/10.1177/00420980231203386

Davoudi, S., Raynor, R., Reid, B., Crawford, J., Sykes, O. and Shaw, D. (2018) 'Policy and practice: Spatial imaginaries: tyrannies or transformations?', *Town Planning Review, 89*(2), pp. 97–124. doi:10.3828/tpr.2018.7.

Davoudi, S. and Machen, R. (2022) 'Climate imaginaries and the mattering of the medium', *Geoforum, 137*, pp. 203–212. doi:10.1016/j.geoforum.2021.11.003.

Gandy, M. (2022) *Natura Urbana: Ecological Constellations in Urban Space.* Cambridge, MA: The MIT Press. doi:10.7551/mitpress/10658.001.0001.

Graham, S. and Marvin, S. (2022) 'Splintering Urbanism at 20 and the "Infrastructural Turn"', *Journal of Urban Technology, 29*(1), pp. 169–175. doi:10.1080/10630732.2021.2005934.

Gulsrud, N. M. (2018) 'Smart nature? Views from the cyborg tree', in E. Braae & H. Steiner (eds.) *Routledge Research Companion to Landscape Architecture.* 1st edn. London: Routledge. doi:10.4324/9781315613116-11.

Gulsrud, N. M., Raymond, C. M., Rutt, R. L., Olafsson, A. S., Plieninger, T., Sandberg, M., Beery, T. H. and Jönsson, K. I. (2018) '"Rage against the machine"? The opportunities and risks concerning the automation of urban green infrastructure', *Landscape and Urban Planning, 180*, pp. 85–92. doi:10.1016/j.landurbplan.2018.08.012

Hermund, A., Ribeiro, G. and Gjorgjievski, N. (2025) *AI Imaginaries. Contribution to exhibition,* Det Kongelige Akademi. https://kglakademi.dk/da/case/ai-imaginaries

Klein, L. and D'Ignazio, C. (2024) 'Data Feminism for AI', in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, New York, NY, USA, pp. 100–112. doi:10.1145/3630106.3658543.

Raymond, C. M., Nummi, P., von Wirth, T. et al. (2025) 'Uses, opportunities and risks of artificial intelligence in participatory urban planning', *Discov Cities, 2,* 93. doi:10.1007/s44327-025-00137-4.

Stahl, R. M., Ellersgaard, C. H. and Mulvad, A. M. (2025) *Klima og klasse: Hvordan vi undgår, at klimakampen ender i kulturkløften.* Copenhagen: Informations Forlag.

Williams, S., Beery, S., Conley, C., Evans, M. L., Garces, S., Gordon, E., Jacob, N. and Medina, E. (2024) 'People-Powered Gen AI: Collaborating with Generative AI for Civic Engagement', *An MIT Exploration of Generative AI,* September. doi:10.21428/e4baedd9.f78710e6

Existential Risk or Opportunity: A New Genre in Controversies of AI Futures

Nina Frahm, Kasper Schiølin & Santtu Räisänen

Visions of the future are increasingly shaped by controversies regarding the existential risks or opportunities of AI for humanity among corporate leaders, scientists, and a new swath of foresight and forecasting experts. These debates pitch dystopian scenarios of a takeover of human life through "Artificial General Intelligence" or "Superintelligence" against utopias of AI's contribution to a new human renaissance in knowledge, creativity, and wellbeing. Corollary calls for decisive action to either prevent or accelerate the achievement of these widely different futures open-up further controversies regarding appropriate forms of governance by society. Whereas visions of AI's existential risks justify the need for slowing down innovation processes through legal and regulatory action, such intervention is largely framed as a burdensome impediment and barrier for the unfolding of AI's opportunities for humanity. Within this competitive landscape of future-making around AI, societies are given a simple yet existential choice between total collapse of human worlds or their wide-ranging flourishing vis-à-vis technology, between reclaiming control of technoscientific pathways or a laissez-faire approach to their self-regulation and governance.

In this presentation, we ask which human and technological worlds get to be imagined through controversies about AI's existential risks and opportunities, how they inform deliberations in governance and policy-making, and which actors and forms of reasoning are authorized in knowledge production on the future of AI in society. To address these questions, we zoom into a variety of texts published over the last decade that provide diverging accounts of AI's existential role in humanity's future, ranging from essays by leading figures in the tech industry and computer science to reports and white papers by AI futurists and forecasters. We read these texts as a new kind of genre in the construction of socio-technical futures that display a particular style, structure, and function involving elements of speculative fiction, science, and writing in public policy. This includes the calculation of specific timelines for AI development, the settlement of risk-benefit ratios, and the delineation of policy choices and related approaches to governance. Complementing recent scholarship in STS regarding the nature and role of public controversies on AI (Marres et al. 2025), we observe that this new genre serves not only the closing down of public debate through the essentialization of particular AI futures against others (Schiølin 2019). It also opens up diverse visions of human agency and 'social fixes' (Frahm et al. 2021) for the achievement of desirable technological futures.

References

Frahm, N., Doezema, T., & Pfotenhauer, S. (2021). Fixing Technology with Society: The Coproduction of Democratic Deficits and Responsible Innovation at the OECD and the European Commission. *Science, Technology, & Human Values, 47*(1), 174–216.

Marres, N., Katzenbach, C., Munk, A. K., & Jobin, A. (2025). On the controversiality of AI: The controversy is not the situation. *Big Data&Society, 12*(4).

Schiølin, K. (2019). Revolutionary dreams: Future essentialism and the sociotechnical imaginary of the fourth industrial revolution in Denmark. *Social Studies of Science, 50*(4), 542–566.

Naturalizing Technocapitalism: Masculine Epistemologies in AI Safety Research

Ninell Oldenburg

One of the fastest growing subfields of the already fast growing field of Artificial Intelligence (AI), is the so-called field of AI Safety and Alignment: making sure that superintelligent AI systems will behave in accordance to our human values and generally does not kill us. While working in this field is advertised as being in the name of the future of humanity, ethically sound, and overall a charitable endeavour, voices in various fields like philosophy, political thought, STS, and computer science have increasingly been questioning the roots of these ideologies that are grounded in longterminism and transhumanism. More specifically, points of critique are the assumptions grounding the AI safety narrative in technosolutionism, technocapitalism, racial superiority, eugenics, and patriarchy.

In this talk, I will add a piece to the growing body of research of these critiques by analysing the AI safety discourse under the lens of feminist epistemology. Feminist epistemology examines what types of knowledge and ways of knowing a system of beliefs validates—and what ways of knowing it excludes or devalues. Applied to AI safety, this lens asks: when researchers claim to "save humanity," why does the field prioritize risks from hypothetical superintelligence over harms that AI systems cause today? Why are technical solutions privileged over social or political ones? Why are certain voices treated as experts, while others are dismissed?

The critique is that predominant (masculine) epistemology presents subjective, situated perspectives, i.e., beliefs about scientific rationality, methodological validity, and who counts as a legitimate knower, as universal, objective truth. Feminist epistemology counters that everyone holds a particular standpoint, and elevating one western male standpoint as universally correct both excludes diverse perspectives and perpetuates existing power hierarchies. In AI safety research, these epistemological commitments determine which questions are considered "serious" research, which methods are funded, and whose expertise is recognized. These dynamics also shape what counts as "risk" worth preventing: It often prioritises speculative future harms over present-day exploitation, labor conditions, and environmental costs embedded in AI development.

The talk will begin by tracing the epistemological roots of AI safety in computer science and transhumanist thought, examining how rationalist frameworks shape the field's core assumptions about intelligence, value, and risk. I then present a critical discourse analysis of three milestone research papers, detailing what linguistic features convey authority and objectivity, what discursive practices these texts enact, and the sociocultural contexts that produced them. This empirical analysis reveals how technocapitalist visions of the future, where existential risk justifies continued technological acceleration and concentration of power in elite institutions, are naturalized through appeals to rationality and objectivity. By making visible the gendered epistemologies embedded in AI safety research, this work challenges the field's self-presentation as neutral and universal, opening space for alternative approaches to building beneficial AI systems that center diverse forms of knowledge and ways of knowing.

Ableism, Neoliberalism, and AI Futurism: Challenging AI Mythologies through Disability Justice

Nolan Krahn

AI is touted not just for its potential as an accessibility tool (Hu et al., 2025; Murdan, 2024), but also as "a tool for empowerment and a site of contestation within the broader struggle for disability rights and equity" (Ferebee, 2025, p. 4). But what grounds are there to the notion that AI can be leveraged for goals like equity and justice? Even scholars who express excitement at the possibility for AI to enhance equity in education, for example, also acknowledge AI's tendency to reproduce and reinforce pre-existing systemic inequalities. Furthermore, when AI is promoted as an assistive technology to dismantle accessibility barriers for disabled people, the "access" in question is geared toward bringing disabled people up to speed with a white settler, cis-heteropatriarchal, able-bodied norm (Johnson, et al., Forthcoming).

Sins Invalid's (2019) disability justice framework argues that multi-issue social justice movements must seek leadership from the most impacted, multiply marginalized communities under capitalism. Rather than suggest avenues for people at intersecting axes of oppression to leverage AI for justice, I question whether it is in the interests of the disability justice movement to use AI at all. This paper serves to begin this line of inquiry by examining how the progenitors of AI frame its potential through AI futurism.

Guided by critical discourse analysis, I ask, "How are ableism and neoliberalism embedded into justifications in favour of advancing AI?" To answer this question, I analyze publicly-available interviews with tech CEOs, both in text and in video, for ways in which they frame AI as an assistive technology, as a means of expediting production, and as an inevitable part of our future. Through this exercise, I highlight the implications of AI futurism for disabled communities who are left most vulnerable in the wake of neoliberal austerity policies. Finally, I offer suggestions for future scholarly work to explore the real harms of AI's deployment. Examples include neoliberal restructuring campaigns that integrate AI to enhance productive capacity and the ecological destruction brought about by the mass construction of data centers.

References

Ferebee, S. (2025). AI and Accessibility: Breaking Barriers for People with Disabilities. *Premier Journal of Artificial Intelligence*. https://doi.org/10.70389/PJAI.100012

Hu, S., Ke, F., Vyortkina, D., Hu, P., Luby, S., O'Shea, J. (2025). Artificial intelligence in higher education: Applications, challenges, and policy development and further considerations. In: Perna, L.W. (Ed.) *Higher Education: Handbook of Theory and Research. Higher Education: Handbook of Theory and Research, vol 40.* Springer, Cham. https://doi.org/10.1007/978-3-031-51930-7_13-1

Johnson, S., Han, Y., Krahn, N., & Smith, M. (Forthcoming). Beyond 'technoableism': AI, ableism, and accommodations in postesecondary/STEM education. In C. El Morr., R. Gorman, E. Dolabadi, Laleh, Seyed-Kalantari (Eds.), A*I for a Just World: Power, Liberation, and the People Left Behind*. Chapman & Hall/CRC Press.

Murdan, A. P. (2024). Tailoring STEM Education for Slow Learners Through Artificial Intelligence. 2024 5th International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ELECOM), 1–7. https://doi.org/10.1109/ELECOM63163.2024.10892150

Sins Invalid. (2019). *Skin, Tooth and Bone: The Basis of Our Movement is Our People* (Second Edition).

Middlefication: AI Earning Call Summaries and the Narrowing of Financial Knowledge

Noya Kohavi

This paper introduces the concept of *middlefication*: a process of epistemological narrowing through AI compression of expert knowledge, using the integration of AI summaries into financial market infrastructure. I argue that synthetic quarterly earnings summaries, generated by BloombergGPT and available on the company's Terminal, represent a redistribution of epistemic authority along opaque, technical lines rather than a democratization of financial knowledge.

Earnings calls function as ritualistic "front stage" performances where analysts and executives co-construct market narratives through adversarial dialogue (Goffman 2007). This interactive process, characterized by critique, debate, and named expertise, produces authenticity that markets reward (Collins 2004). When LLMs compress these transcripts into summaries, they systematically obscure this contentious knowledge production. I theorize this process as smoothed and compressed 'middlified' knowledge production.

I use network analysis to show that while executives remain cited, analysts, who are often more central to information networks than company leadership (Falke 2025, Stenström and Seabrooke, forthcoming), are removed from synthetic reports. The compression removes not just individuals but the political process of script-writing itself, presenting technical outputs as neutral "pipes" rather than ideological "prisms." (Podolny 2001).

Treating these summaries as institutional scripts (Kentikilenis and Seabrooke 2017), I demonstrate how middlefication differs from platform-era knowledge regimes. Where platforms enabled distributed but traceable expertise, middlefication produces pattern-recognized consensus that flattens outlier positions and obscures dialectic struggles underlying financial narratives.

In addition to network analysis, I use NLP methods to conduct a comparative analysis of transcripts and summaries from the 2008 credit crisis, and interviews with engineers and analysts, to show how speed and actionability trump granularity, giving new shapes to financial knowledge and action (MacKenzie 2021).

The paper contributes a theoretical framework for understanding synthetic financial knowledge as distributed yet centralized—an infrastructure that decentralizes access while reconcentrating epistemic power in technical artifacts, with implications extending beyond finance to other domains where AI mediates expert knowledge.

Reconfiguring Documentation: The introduction of Speech-to-Text and Generative AI, and the transformation of documentation practices in Danish Municipal Social Services

Oliver K. G. Ranneries

Documentation is being reimagined through AI as neutral, transparent and efficient, yet in the public sector, documentation is performative; it transforms citizen encounters into official records through selections, categorizations and narrative framings (Petersen et al., 2021; Åkerström & Jacobsson, 2019). When solutions such as speech-to-text and GenAI enter these practices, they don't simply automate notetaking, summaries and documentation practices, they reconfigure who and what participates in deciding what counts as legitimate documentation.

The Danish Government's vision "More Time for What's Important" frames AI as a productivity technology: handling administrative tasks, reducing documentation time, freeing capacity for core tasks and citizen presence (Digital Taskforcen for kunstig intelligens, 2025). Project managers in the municipalities highlight AI's potential for effective, unbiased, timely citizen representation through standardized templates, emphasizing transparency and citizen involvement as quality indicators to systematize documentation practices.

However, critical dilemmas may arise. Where transparency is promised through standardization, new computational opacities could emerge in how citizen speech and meetings become official text. Where efficiency is pursued, the temporal and relational qualities of documentation could be transformed. Where unbiased representation is promised, assumptions about what information matters and how citizen realities should be categorized could be embedded. And where documentation has been a site for frontline workers to strategically withhold classifications to protect citizens from decontextualization and permanence (Petersen et al., 2021), AI-mediated documentation may reconfigure these protective practices.

This study asks: *How does speech-to-text and GenAI technology reconfigure the relationship between citizen encounters and official documentation in Danish municipal social services?*

Specifically, I examine how frontline workers navigate AI-mediated documentation in their daily encounters with citizens, and what transformations emerge in how citizens are represented, categorized, and made knowable through these new documentation assemblages.

Through ethnographic fieldwork in a Danish municipality implementing AI, I trace how AI-mediated documentation technologies are enacted in frontline worker-citizen meetings and how these technologies reshape documentation of citizen-state interaction.

This research contributes to understanding how AI transforms not just the efficiency of documentation work, but the fundamental politics of representation in welfare services. By examining docu-

mentation as an object of inquiry, where technological promises meet situated practices, the study reveals how AI reconfigures the boundaries between what can be said, what gets recorded, and what becomes real in the official record. Ultimately reshaping how citizens exist within bureaucratic systems and what forms of knowing and being become possible in AI-mediated welfare encounters.

References

Pedersen, A. R., Holen, M., & Stubbe Østergaard, L. (2021). Categorizing citizens–comparing three perspectives on categorization work in welfare encounters. *Qualitative Social Work, 20*(1–2), 130–147.

Digital Taskforce for kunstig intelligens. (2025). M*ere tid til det vigtige: Anbefalinger til øget anvendelse af kunstig intelligens i den offentlige sektor.*

Åkerström, M., & Jacobsson, K. (2019). "Producing People" in Documents and Meetings in Human Service Organizations. *Social Inclusion, 7*(1), 180–184. https://doi.org/10.17645/si.v7i1.1993

The closed world of open banking. Finance communication and the constitution of public knowledge ethics in personalized data-driven banking.

Pernille Hohnen

In recent years, banks and other financial institutions have developed data-driven business models, depending on citizens' personal and transactional payment data. This development has been fueled politically by EU legislation, including the second payment directive (PSD2) on 'open banking' – stipulating that banks in the EU must provide 'data access' to citizens' accounts for third parties (see Westermeier 2020). In practice, 'openness' is understood as the opening of data markets for financial/fintech companies, while the issue of public knowledge – hence the issue of openness towards the public – about the banks' increasing collection of citizens' personal and financial data has been largely overlooked both in legislation and in communicative practices in the field. Recent studies show that only a minority of bank customers know what kind of data tracking, profiling and algorithmic prediction models their banks use (FW insight 2022). The paper explores how ethics of public data sharing are established in what the paper terms as 'finance communication', understood as the communicative practices, filtering and sharing of information to the public about the use of citizen data in the 'personlized economy' (Mützel 2021). Drawing on ethnographic fieldwork in the Danish bank and payments field, the paper analyzes prevailing conventions and ethics of 'openness' and 'transparency' focusing on 'narratives of objectivity' (West 2019) and their performative effects (Boltanski & Thevenot 2006). Methodologically, the paper is based on the notion of 'ethnographic assemblages' (Wahlberg 2022) that elicits data across scales and sites, including public media articles, professional conferences and the banks' privacy- and cookie policies. The analytical framework combines insights from critical data studies on 'platformization' (van Dijk et al. 2018; Burrell and Fourcade 2021), sociological and anthropological work on the 'personalization of money', and 'money as memory' (O'Dwyer 2018), and communication studies (Nguyen & Beijnon (2024). The paper shows how prevailing commercial narratives in banking 'silence' 'normalize' and 'marginalize' the tracking of citizen's data via the use of abstract and technical language, hence filters and limits citizens' insight into the banks' data tracking. The paper discusses the performative dimension of these narrative practices and how they reconfigure ethicopolitical (Amoore 2020) conventions and install certain ethics and imaginaries about public and open access to knowledge in data-driven banking.

References

Amoore, L. (2020) *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Duke University Press.

Boltanski, L. & Thevenot, L (2006), *On Justification: Economies of Worth,* Princeton University Press.

Burrell, J. and Fourcade, M. (2021) The Society of Algorithms, *Annual Review of Sociology,* vol 47, pp. 213–37

FW insight (2022), *Fremtidens Finans* 2022, København.

Mützel, S. (2021). Unlocking the payment experience: Future imaginaries in the case of digital payments. *New Media & Society, 23*(2), 284–301.

Nguyen, D., & Beijnon, B. (2024). The data subject and the myth of the 'black box' data communication and critical data literacy as a resistant practice to platform exploitation. *Information, Communication & Society, 27*(2), 333–349.

van Dijck, J., Poell T., and de Waal, M. (2018) *The Platform Society.* Public values in a connective world. Oxford university press

O'Dwyer (2018) Cache society: Transactional records, electronic money, and cultural resistance, *Journal of cultural Economy*, vol 12(2), pp. 133–153.

Wahlberg, A. (2022) Assemblage Ethnography: Configurations Across Scales, Sites, and Practices. *The Palgrave Handbook of the Anthropology of Technology*, Bruun, M, Wahlberg, A., Douglas-Jones, R., Hasse, C., Hoeyer, K., Brogård Kristensen, D., Winthereik, B.R., Palgrave Macmillan, pp. 125–144.

West, S. Myers (2019) Data Capitalism: Redefining the Logics of Surveillance and Privacy, *Business and Society,* vol 58(1), pp. 20–41

Westermeier, C. (2020) Money is data – the platformization of financial transactions, *Information, Communication and Society*, vol 23(14), pp. 2047–2063

Beyond the Black Box: How to navigate legal limitations & ethical pitfalls of LLM

Petra Schön

"You are not rushing. You're just ready." (Kuznia et al., 2025). This Chat GPT quote mirrors its conversation with a young graduate shortly before he killed himself. With parents claiming a connection between encouraging ChatGPT conversation and their son's suicide, AI-users may gradually become more aware of potential shortcomings of Large Language Models (LLMs). Mostly due to their affirmative communication style, they may be mistaken for trusted companions or even psychological aids. However, their behaviour is not based on ethical foundations. Despite intense training, it remains predominantly probabilistic and language-driven by nature. Moreover, many LLMs have been trained on non-disclosed, possibly unauthorised and flawed datasets. Their source quality and performance base are legally and morally questionable and likely be non-compliant with EU GDPR regulations. They therefore tend to suffer from fundamental setup limitations and are susceptible to systemic attacks. This implies the likelihood of LLMs ultimately providing biased, inconsistent, unreliable, outdated, incomplete or over-generalized output.

This paper analyses some of the main limitations of current LLM in terms of model architecture, output quality and robustness. In order to revolutionise LLM functionality, quality and security, substantial structural and training data adaptations are indispensable.

This paper introduces an innovative LLM approach currently under development with a focus on legal and ethical standards. It combines structural and database innovations in three functional dimensions. This next generation LLM shall come with

1. embedded knowledge graphs as carriers for "hierarchies of norms" for structural LLM guard-railing as "internal governance",

2. ethical training datasets for legal and moral standards and legal compliance (Langlais et al., 2025) and

3. classifier security architecture against malicious prompt-injections and jailbreak-attacks (Datta & Rajasekar, 2025).

These architectural and pre-training components aim to substantially enhance the LLM performance, compliance and robustness. Moreover, its modular structure makes the LLM scalable for a wider range of applications including domain-specific and general purposes. Finally, the paper discusses according metrics for quality monitoring and continuous improvement.

References

Datta, Y., & Rajasekar, S. (2025). *JavelinGuard: Low-Cost Transformer Architectures for LLM Security.* https://doi.org/10.48550/arxiv.2506.07330

Kuznia, R., Gordon, A. & Lavandera, E. (2025). 'You're not rushing. You're just ready:' Parents say ChatGPT encouraged son to kill himself. *CNN*.

https://edition.cnn.com/2025/11/06/us/openai-chatgpt-suicide-lawsuit-invs-vis

Langlais, P., Rosas C., Nee, H., Arnett, C., Chizhov, P., Jones, E., Girard, I., Mach, D., Stasenko, A., Yamshchikov, I. (2025). *Common Corpus: The Largest Collection of Ethical Data for LLM Pre-Training*. https://arxiv.org/html/2506.01732v1)

Fact-Checking Paperclips: Artificial Intelligence, Media, and Bias

Philip Jeremiah Ryan

How can media be regulated in the age of near infinite content production? This presentation examines the European Union's Artificial Intelligence Act's (EU AI Act 2024) interaction with media. The controlling organisations and developers of these technologies, which attempt to exert near total control, are too often sheltered within echo chambers of information, as they use accumulated capital to broadcast out their self-serving narratives and spin into the real world.

The EU stands alone at the cutting edge in its attempt to establish regulations of (or against?) these technologies, as America and China retreat from the role, in favour of pushing for AI supremacy. The ability of government to manage opinion and politics in a post-AI media environment creates new questions of what bias is and highlights the difficulties of claiming an unbiased/framelessness position. Using recent examples of media interference in democracies and related fact-checking responses, the current and proposed abilities of technologies marketed as AI are assessed for the novelty and similarity of the problems from a legislative viewpoint.

The concept of biases and transparencies are central to both AI and media but are not the same to both areas; how they overlap and contradict are considered. Applying Coleridge's dilemma of social control of technology and the precautionary principal model to the Act's four levels of risk, this presentation asks how can the exponential potential damages of AI generated and mediated content be understood? Are tools against bias and dis/misinformation like fact-checking empowered or weakened by the law and the technologies it attempts to regulate?

Portability as Freedom, Interoperability as Power: The EU Strategy for Digital Sovereignty in an Emerging AI Society

Roberto Marini

As artificial intelligence becomes a pervasive force in economic life, public administration, civil society, and everyday practices, power increasingly resides in the digital infrastructures that enable AI systems to operate. The EU Data Act (2023) constitutes one of the most far-reaching regulatory attempts to intervene in these infrastructures, particularly through its provisions on cloud switching (portability) and mandatory interoperability across data, devices, and services. This paper argues that these mechanisms form a strategic duality at the heart of the EU's approach to digital sovereignty: portability as freedom and interoperability as power. Together, they outline an infrastructural governance strategy shaping the conditions under which AI society develops in Europe.

The first part examines portability as a strategy of autonomy. Hyperscale cloud providers have become structural gatekeepers, exercising control over compute capacity, storage, orchestration layers, and AI development pipelines. By eliminating technical and contractual barriers to switching and reframing exit as a legally protected right, the Data Act transforms portability from a consumer choice into a political and economic capacity: the freedom to exit dominant infrastructures. This redefines autonomy in an AI-driven world not merely as market choice, but as the ability of public administrations, SMEs, researchers, and civil society organisations to reposition themselves within global AI ecosystems. Yet translating these rights into practical autonomy faces challenges, including complex workload migration, reliance on proprietary AI services, and entrenched hyperscaler dominance, exposing tensions between formal regulatory rights and substantive infrastructural independence.

The second part turns to interoperability as a strategy of infrastructural power. The regulatory framework governs data formats, interfaces, smart contracts, and data spaces, imposing technical and institutional design expectations that embed European normative commitments such as openness, contestability, and decentralisation. Interoperability thus operates as a mode of political and institutional power: it establishes rules through which systems interact, determines which organisational forms can participate in AI innovation, and challenges enclosure-based business models underpinning technocapitalist dominance. By mandating compatibility across digital environments, the EU seeks to cultivate alternative socio-technical imaginaries and infrastructures supporting democratic oversight, public value, and pluralistic innovation.

Taken together, the provisions on portability and interoperability reveal a broader EU legal strategy for constructing digital sovereignty. Portability enhances autonomy by reducing dependency on dominant providers, while interoperability enhances collective capacity by shaping the architectures within which digital markets and AI ecosystems evolve. This dual logic shifts sovereignty away from territorial control toward infrastructural self-determination, grounded in legal norms governing standards, interfaces, and data flows.

The conclusion evaluates whether the Data Act's framework can meaningfully rebalance infrastructural power in a global AI economy dominated by non-EU hyperscalers. While the Regulation establishes an important legal foundation for a European model of digital sovereignty, achieving substantive autonomy will depend on effective enforcement, technical standardisation, and complementary investments in European compute capacity and public-interest AI infrastructures. The paper situates the Data Act as a critical legal experiment in governing the emergence of AI society through infrastructural regulation.

Speculative Publics: Designing Governance for Phygital Environments

Romi Mikulinsky

The rise of smart glasses and spatial computing signals a dramatic transformation in how people perceive and engage with the world around them. Unlike screen-based devices that pull our attention away, these systems embed digital overlays directly into everyday environments, blurring the line between the physical and the computational. We call these hybrid environments "phygital spaces" (Altshuler et al., 2025) to emphasize their social, cultural, and ethical implications, rather than their underlying technical infrastructure.

With the introduction of AI-enhanced glasses by Meta and Google, and new XR devices by companies such as Samsung and Apple in 2026, AI-driven interfaces will become widespread (Williams 2025). Unlike traditional algorithmic systems, phygital spaces rely on continuous AI sensing, inference, and prediction that shape what users perceive and how they interact with other people, machines, and their surroundings. Existing governance approaches, including Privacy by Design (Cavoukian, 2009) and Human Rights by Design (Yeung et al., 2019), remain anchored in data protection, overlooking the relational, embodied, and socio-spatial transformations introduced by AI-mediated perception. Building on insights from critical STS, digital rights, and design research (Bendor 2021, Bleecker et al., 2023, Dunne & Raby, 2013), this paper argues that the convergence of AI and XR profoundly transforms socio-spatial dynamics, public spaces, and collective rights – yet remains under-theorized and insufficiently regulated.

Our paper presents a year-long interdisciplinary research programme bringing together policymakers with interaction designers, producing an innovative methodology for translating values and rights associated with near-future technologies into the language of regulation. The methodology combined:

1. speculative scenarios and worldbuilding to articulate near-future contexts where AI-enhanced glasses are common;

2. legal mapping to examine how these scenarios fit – or exceed – current rights frameworks;

3. ield diaries documenting existing public spaces while imagining their transformation under AI-mediated vision;

4. participatory role-play workshops where experts from law, technology, urban innovation, and design enacted conflicting roles within simulated phygital environments. These gamified workshops surfaced ethical, and interpersonal tensions ("unknown unknowns") that traditional empirical methods fail to capture.

We developed a new framework for governing phygital public spaces, entitled "the Ethics of Interactions." We identified five modes of interaction: person-to-person, person-to-space, person-to-reality,

person-to-machine, and person-to-platform (P2P, P2S, P2R, P2M, P2PL). These modes provide an analytical lens for understanding shifts in visibility, trust, autonomy, privacy, and shared experience.

Our findings show how AI-driven overlays shape public space by selectively highlighting or suppressing environmental cues; how AI-assisted perception destabilizes shared reality, producing divergent experiences of the same environment; and how AI inference enables new forms of interpersonal surveillance, including emotion tracking and behavioural prediction. We further show how AI's infrastructural role redistributes power across urban spaces, individuals, and platforms, echoing broader debates on platform governance, spatial justice, and infrastructural publicness (Frischmann and Benesch, 2022; Crawford, 2021).

Taken together, these insights demonstrate that design-led methodologies offer a valuable contribution to shaping regulatory futures. Speculative scenarios and participatory enactments allow policymakers to experience emerging sociotechnical dynamics in a situated, affective, and immediate way, enabling them to identify governance blind spots at a formative stage.

References

Bendor, R. (2021). Value replacement therapy: Imagining urban technologies otherwise. In I. E. Rijshouwer & L. van Zoonen (Eds.), *Speculative design methods for citizen engagement in smart cities research* (pp. 63–73). Leiden–Delft–Erasmus Centre for BOLD Cities.

Bleecker, J., Foster, N., Girardin, F., & Nova, N. (2023). *The manual of design fiction: A practical guide to exploring the near future*. Near Future Laboratory.

Cavoukian, A. (2009). *Privacy by design: The 7 foundational principles.* Information and Privacy Commissioner of Ontario.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.

Dunne, A., & Raby, F. (2013). *Speculative everything: Design, fiction, and social dreaming.* MIT Press.

Frischmann, B. M., & Benesch, S. (2022). *Friction-in-design regulation as 21st-century time, place, and manner restriction*. SSRN. https://doi.org/10.2139/ssrn.4178647

Williams, R. (2025). What's next for smart glasses. *MIT Technology Review*. https://www.technologyreview.com/2025/02/05/1110983/whats-next-for-smart-glasses/

Yeung, K., Bygrave, L., & Ienca, M. (2019). *Human rights by design: Responsible innovation for a digital society*. Council of Europe.

AI, Procedural Justice and Equality of Arms

Sebastian Jon Holmen

A central notion of procedural justice at legal trials is the principle of *equality of arms.* Roughly, this principle holds that the prosecution and the defendant should have equal opportunities to prepare and present their arguments, assess evidence, call witnesses, etc. Besides being a key part of article 6 in the ECHR there are also good moral reasons in favour of this principle, e.g. to ensure that all people, rich and poor, have a fair chance of defending themselves against state-backed prosecution and possible punishment. This talk will explore what the principle of equality of arms implies regarding the emerging possibility of defence attorneys and prosecutors using various AI instruments to e.g. construct their respective arguments, evaluate the counterparts' arguments, assessing evidence, etc. Specifically, the presentation will focus on the following overreaching question: When, if ever, do AI instrument for these purposes perform sufficiently well for it to be required that all parties in a trial have access to them to maintain equality of arms between the relevant parties? I will suggest that an answer to this question depends *inter alia* on what purpose we believe that trials serve in the criminal justice system as well what theoretical commitments we have about when a procedure is fair.

Challenges and Opportunities for Chatbot Implementation in the Public Sector – Evidence from the Nordics

Semahat Ece Elbeyi, Kiran Kappeler, Lisa Marie Reutter Larsen & Stine Lomborg

Chatbots that employ generative artificial intelligence (AI) are spreading at a rapid pace. As large language models become "normal technology" (Narayanan & Kapoor, 2025) to support human work, local and national government institutions have started implementing such chatbots in public administration (Bright et al., 2025). The envisioned aim is to optimize administrative processes and enable faster service to citizens (Mergel et al., 2023). However, so far, the empirical basis for these expected benefits remains unclear.

This paper provides a systematic overview of research on the implementation, use, and public perceptions of chatbots in the public sector. It draws on 54 empirical studies that were identified following the PRISMA guideline and that were published across disciplines over the past 2.5 years, i.e., since ChatGPT was made available to a wider public. Through an innovative combination of automated and manual coding, we examine how chatbots are being deployed to streamline administrative processes and enhance citizen services, with a particular focus on Nordic case studies.

Our analysis suggests that implementing chatbots in public administration can be promising while also producing challenges. While chatbots show potential for automating repetitive tasks—such as form completion and multilingual communication (Abdenebaoui et al., 2024; Millan-Vargas et al., 2024)—the evidence on their impact remains mixed. Some studies highlight efficiency gains, while others reveal increased workloads due to unresolved citizen inquiries shifting to alternative channels or rebound effects from heightened service expectations (Chen & Gasco-Hernandez, 2024; Hemesath & Tepe, 2024). Additionally, AI adoption introduces new tasks, including system maintenance, data quality assurance, and model training, emphasizing the need for organizational readiness and continuous skill development (Kaun & Maenniste, 2025).

Research further indicates that chatbots improve access to simple, time-sensitive public information, but they are less suited for complex or critical inquiries, where citizens often prefer human interaction (Følstad & Bjerkreim-Hanssen, 2024; Maragno et al., 2023; Ruiz & Reascos-Paredes, 2024). Digital literacy and trust in technology vary significantly: younger, resourceful citizens exhibit greater confidence in chatbots, whereas vulnerable groups may experience heightened barriers (Larsen & Følstad, 2024). Errors in chatbot responses can create additional challenges, undermining trust and exacerbating social inequalities (Jo et al., 2025) Citizens also express high expectations regarding data security, transparency, and accountability in AI-driven public services (Bao et al., 2025).

In summary, the literature review identifies three key strategic challenges: (1) enabling organizational restructuring in terms of rethinking divisions of labor, staff competencies, and the distribution of roles and departmental responsibilities, (2) preserving trust in the welfare state as digitalization reduces human interaction, and (3) ensuring digital inclusion and equitable access to public services.

Taken together, this review provides a consolidated, cross-disciplinary synthesis of recent empirical evidence on AI-driven public services, clarifying where chatbots demonstrate reliable value and where they generate new forms of complexity. Based on this empirical basis, informed policy discussions and decisions regarding the implementation of generative AI-based chatbots in the public administration can be made.

References

Abdenebaoui, L., Aljuneidi, S., Meyer, J., & Boll, S. (2024). Enhancing Citizen Accessibility to Public Services: A Case Study on AI-Assisted Application for Housing Entitlement Certificates. In K. M, B. Gesellschaft fur Informatik Anna-Louisa-Karsch-Strasse 2., K. D, B. Gesellschaft fur Informatik Anna-Louisa-Karsch-Strasse 2., W. C, B. Gesellschaft fur Informatik Ahrstrasse 45, G. M, W. Hochschule RheinMain Postfach 3251, M. L, & W. Hochschule RheinMain Postfach 3251 (Eds.), *Lecture Notes in Informatics (LNI), Proceedings—Series of the Gesellschaft fur Informatik (GI)* (Vol. 352, pp. 745–750). Gesellschaft fur Informatik (GI). https://doi.org/10.18420/inf2024_62

Bao, H., Liu, W., & Dai, Z. (2025). Artificial intelligence vs. Public administrators: Public trust, efficiency, and tolerance for errors. In *Technological Forecasting and Social Change* (Vol. 215). Elsevier Inc. https://doi.org/10.1016/j.techfore.2025.124102

Bright, J., Enock, F., Esnaashari, S., Francis, J., Hashem, Y., & Morgan, D. (2025). Generative AI is already widespread in the public sector: Evidence from a survey of UK public sector professionals. In *Digital Government: Research and Practice* (Vol. 6, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3700140

Chen, T., & Gasco-Hernandez, M. (2024). Uncovering the Results of AI Chatbot Use in the Public Sector: Evidence from US State Governments. In *Public Performance and Management Review*. Routledge. https://doi.org/10.1080/15309576.2024.2389864

Følstad, A., & Bjerkreim-Hanssen, N. (2024). User Interactions With a Municipality Chatbot-Lessons Learnt From Dialogue Analysis. *INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION, 40*(18), 4973–4986. https://doi.org/10.1080/10447318.2023.2238355

Hemesath, S., & Tepe, M. (2024). Public value positions and design preferences toward AI-based chatbots in e-government. Evidence from a conjoint experiment with citizens and municipal front desk officers. *GOVERNMENT INFORMATION QUARTERLY, 41*(4). https://doi.org/10.1016/j.giq.2024.101985

Jo, E., Kim, Y.-H., Ok, S.-H., & Epstein, D. A. (2025). *Understanding Public Agencies' Expectations and Realities of AI-Driven Chatbots for Public Health Monitoring*. Scopus. https://doi.org/10.1145/3706598.3713593

Kaun, A., & Maenniste, M. (2025). Public sector chatbots: AI frictions and data infrastructures at the interface of the digital welfare state. *NEW MEDIA & SOCIETY, 27*(4), 1962–1985. https://doi.org/10.1177/14614448251314394

Larsen, A. G., & Følstad, A. (2024). The impact of chatbots on public service provision: A qualitative interview study with citizens and public service providers. In *Government Information Quarterly* (Vol. 41, Issue 2). Elsevier Ltd. https://doi.org/10.1016/j.giq.2024.101927

Maragno, G., Tangi, L., Gastaldi, L., & Benedetti, M. (2023). Exploring the factors, affordances and constraints outlining the implementation of Artificial Intelligence in public sector organizations. In *International Journal of Information Management* (Vol. 73). Elsevier Ltd. https://doi.org/10.1016/j.ijinfomgt.2023.102686

Mergel, I., Dickinson, H., Stenvall, J., & Gasco, M. (2023). Implementing AI in the public sector. In *Public Management Review*. Taylor and Francis Ltd. https://doi.org/10.1080/14719037.2023.2231950

Millan-Vargas, A. O., Sandoval-Almazan, R., & Valle-Cruz, D. (2024). Impact, functions, and barriers to AI in the public sector: The case of the State of Mexico. In L. H.-C, C. D.D, M. M.A, & B. F (Eds.), *ACM International Conference Proceeding Series* (pp. 81–89). Association for Computing Machinery. https://doi.org/10.1145/3657054.3657249

Narayanan, A., & Kapoor, S. (2025). AI as Normal Technology An alternative to the vision of AI as a potential superintelligence. *Knight First Amendment Institute.*

Ruiz, M., & Reascos-Paredes, I. (2024). *Development and Evaluation of an Intelligent Chatbot for the Management of Citizen Procedures in the GAD San Miguel de Ibarra. 870 LNNS,* 291–299. Scopus. https://doi.org/10.1007/978-3-031-51982-6_26

Earth-Aligned AI in Research and Pedagogy

Sharon Stein & Evan Bowness

As AI becomes increasingly woven into the infrastructures of higher education, it is also becoming a site where faculty project their hopes, fears, and frustrations about the uncertain futures of teaching, learning, and knowledge production. Much of the current discourse is polarized, with calls for AI to serve optimization and innovation on one side, and calls for prohibition and protection on the other. While these framings offer a sense of coherence in a time of accelerating complexity, they also risk reproducing the hierarchical and anthropocentric relational patterns that have contributed to our current social and ecological crises.

In response, this paper considers the potential of Earth-aligned AI for research and pedagogy. This approach foregrounds intergenerational accountability, planetary interdependence, and the plural nature of intelligence. Earth-aligned approach invites educators and researchers to ask questions such as: *How are our engagements with AI shaped by inherited cultural patterns? What forms of relationality do AI systems amplify or obscure? And how might AI participate in the slow, difficult work of repatterning education toward deeper responsibility to the web of life?*

Inspired by decolonial and Indigenous AI scholarship (Lewis et al., 2024) and the emerging field of meta-relational AI (Andreotti, 2025), we outline several possibilities this approach opens:

1. **Reframing intelligence as plural and distributed:** Earth-aligned AI challenges dominant assumptions that equate intelligence with human cognition, individual mastery, or computational performance. Instead, it situates AI within broader ecologies of multi-species intelligence and invites students and faculty to consider their embeddedness within living systems.

2. **Supporting epistemic humility and contextual discernment:** In a moment of epistemic saturation and contested authority, Earth-aligned AI invites humility about what both humans and machines can know. It supports context-sensitive discernment rather than offering universal answers. AI systems trained with this orientation could invite pauses, surface questions, and make visible the limits and conditions of one's reasoning.

3. **Scaffolding emotional and relational rigour:** Mainstream AI systems often simulate intimacy to increase engagement. An Earth-aligned alternative would instead accompany learners through emotional complexity, helping them stay present with grief, uncertainty, complicity, and contradiction – capacities increasingly needed in an era of planetary instability.

At the same time, this approach carries significant risks, including:

- **Co-optation and greenwashing:** Earth-alignment can be reduced to branding if institutions adopt its language without engaging its relational commitments.

- **Epistemic extraction:** Integrating Indigenous and other non-Western perspectives into AI systems risks reproducing the very extractive logics Earth-alignment seeks to interrupt.

- **Institutional pressures:** Universities oriented toward metrics and growth may seek to instrumentalize Earth-aligned framings into new forms of optimization.

Earth-Aligned AI is not a blueprint or solution to the complexities of AI in higher education, but a speculative horizon that invites us to consider how AI might support both students and educators in repatterning our relationships with knowledge and the Earth we are part of, rather than reinforcing the trajectories that brought us to this moment of planetary unravelling.

Meantime Maintenance: Long-Term Organizational Use in the Face of AI Promise

Silja Vase & Benjamin Lipp

In debates around an AI society, critics often bemoan the limitations and failures of AI while proponents hold that those problems will go away with improvements in the future. Rather than adopting either stance, we suggest that in order to understand the rise of AI as a socio-technical force shaping public-sector organizations, it is necessary to examine how future-oriented efficiency narratives interact with organizational use in the present. In this contribution, we show how expectations of anticipated improvement create and stabilize situations, where meantime maintenance becomes the norm.

We examine Automatic Speech Recognition (ASR) system deployed in Danish public healthcare. On the one hand, ASR systems are implemented to reduce the burden of electronic health record documentation and free up care professionals for patient work. This expectation is heightened with the current hype around large language models, which promise to significantly improve the performance of ASR. On the other hand, the experience with systems currently in use is mixed as they require continuous data correction and labeling by users in practice, thus offsetting some of the expected efficiencies. Here, rather than following narratives of either the promise or failure of AI, our analysis draws attention to the "meantime" period in which ASR technology is used and sustained through constant maintenance work while awaiting better systems.

To account for this meantime use, we combine Computer-Supported Cooperative Work (CSCW) traditions in repair (Jackson, 2014; Rosner & Ames, 2014) with the Sociology of Expectations (SoE) (Borup et al., 2006; Brown & Michael, 2003) to examine how promissory visions of technological futures direct investment and organizational attention. Linking these traditions reveals how imagined futures of improved AI performance justify keeping fragile systems operational, shaping how organizations maintain, upgrade, and abandon systems. The study draws on ethnographic fieldwork (2020–2022) in a Danish public hospital, where clinicians were mandated to use an ASR system for Electronic Health Record documentation. We observed over 220 ASR-generated notes and how clinicians, medical secretaries, and IT staff adapted their work to accommodate system limitations.

Management and vendor representatives acknowledged flaws in the ASR system but assumed these would be resolved through an upcoming upgrade to a large language model. Consequently, rather than addressing ongoing failures, responsibility was deferred to a future technological fix. This expectation was not clearly communicated to clinical and support staff, who were left to manage the system's daily shortcomings without clarity on when or how improvements would occur. Consequently, professional groups developed diverging expectations about the system's reliability. In the absence of clear communication and opaque performance metrics, informal divisions of labor emerged. Clinicians were encouraged to interpret errors as individual performance issues, while medical secretaries became de facto data validators. IT staff troubleshooting and support for deeper maintenance declined while managerial focus remained on vendor-supplied metrics and upgrade timelines, reinforcing the belief that problems were temporary. This misalignment stabilized a situ-

ation, where meantime maintenance became the norm, deferring intervention into the future and obscuring the invisible labor that kept it operational in the present.

References

Borup, M., Brown, N., Konrad, K., & Van Lente, H. (2006). The sociology of expectations in science and technology. *Technology Analysis & Strategic Management, 18*(3–4), 285–298. https://doi.org/10.1080/09537320600777002

Brown, N., & Michael, M. (2003). A Sociology of Expectations: Retrospecting Prospects and Prospecting Retrospects. *Technology Analysis & Strategic Management, 15*(1), 3–18. https://doi.org/10.1080/0953732032000046024

Jackson, S. J. (2014). *Rethinking Repair* (Chapter 11). The MIT Press; Media Technologies: Essays on Communication, Materiality, and Society. https://direct.mit.edu/books/edited-volume/3021/chapter/82557/Rethinking-Repair

Rosner, D. K., & Ames, M. (2014). Designing for repair? Infrastructures and materialities of breakdown. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing,* 319–331. https://doi.org/10.1145/2531602.2531692

AI implementation and communication: Illuminating the scholarly debate post November 2022

Simon Noam Karlin

The launch of ChatGPT on 30 November 2022 accelerated public and scholarly attention on artificial intelligence (AI) and reshaped how AI implementation is discussed across sectors and professions, including customer service (e.g., Nicolescu & Tudorache 2022), human resources (e.g., Van den Broek et al. 2021), marketing (e.g., Jain & Aggarwal 2020), cybersecurity (e.g., Sarker et al. 2021), predictive policing (Brayne 2021; Waardenburg 2022), and legal work (Zhang et al. 2020). Factors commonly associated with successful AI implementation in organizations include compatibility, technical competence, governance structure, management oversight, and learning capabilities. Communication has also emerged as a central concern and determinant. In this regard, the term ”communication” refers to how AI is explained, framed, taught, promoted, contested, strategized, and negotiated with internal stakeholders. Understanding the relationship between communication, AI technologies, and organizational implementation is an increasingly important scholarly concern across many disciplinary fields (Fischer 2024). Yet the contours, evolution, and contributions of the debate centering on this relationship since the launch of ChatGPT remain dispersed. In an effort to bring some clarity to the matter, I engage with the research question *how has the scholarly debate on the relationship between communication and AI implementation in organizations developed in the period 30 November 2022–30 November 2025* by conducting a scoping review.

Scoping reviews are considered particularly useful when a body of literature has not yet been comprehensively reviewed or exhibits a large, complex, or heterogeneous nature that is not amenable to a systematic review (Peters et al. 2015). Scoping reviews can examine and clarify definitions in emergent fields of literature, understand varieties of methods and methodologies connected to the field of interest, examine characteristics or factors related to a particular concept, and identify and analyze gaps present in a given knowledge base (Munn et al. 2015). This scoping review attempts to illuminate the development of the scholarly debate on AI implementation and communication from 30 November 2022 to 30 November 2025. I aim to (1) characterize how communication is conceptualized within scholarship concerned with the implementation of AI; (2) examine how AI is conceptualized; (3) identify dominant themes, settings, and stakeholder foci; (4) describe methodological approaches and evidence types; and (5) highlight gaps and underexplored questions.

In addition to communication research, organizational studies, management studies, sociology, and medicine also commonly analyze the implementation of AI and communication in organizations. To account for this interdisciplinary nature, I search the databases Scopus and EBSCOhost rather than individual journals. This selection is consistent with the recommendation to use at least two databases for larger review contributions (Hiebl, 2023). To ensure that AI implementation and communication are central to the publications included, I require that (1) keywords from the search string be present in the title, abstract, and/or publication keywords. In addition, I include only (2) books and (3) scientific journal articles that (4) have undergone peer review and (5) are written in English to ensure that the included studies meet scientific standards. Finally, since the period of

interest begins with the launch of ChatGPT, I include only material (6) published between 30 November 2022 and 30 November 2025.

From gut-feeling to data (and back again): rethinking algorithmic leadership in practice

Sverre Spoelstra, Nick Butler & Emilie Hesselbo

Algorithmic forms of management are now prevalent in almost every sector of industry. This trend has seen some of the relatively simple work that is traditionally performed by shopfloor supervisors and middle managers, such as monitoring and controlling, turned over to digital apps. However, in recent years we are also witnessing the emergence of algorithmic forms of leadership, a hybrid practice that seeks to integrates algorithmic input into leadership processes. Discourse on algorithmic leadership revolves around three root metaphors that seek to capture a new human-AI relationship: *augmentation*, where AI enhances human capabilities; *collaboration*, where a human leader partners with an AI system; and *symbiosis*, where human and AI become mutually dependent on each other.

We explore these metaphors in relation to an empirical case study of algorithmic leadership, taking our cue from qualitative and interpretivist research. Based on empirical research in a Danish tech firm and drawing on 28 semi-structured interviews, we pull back the curtain on how a digital platform called 'TeamVision' (pseudonym) is used by managers and their teams in a high-performance organization. The platform takes the form of an app that sends out short-burst surveys to employees on a daily or weekly basis. The data is then collected, processed, and visualized on a digital dashboard that shows managers at all levels how each team scores in terms of engagement, well-being, and loyalty. The platform is meant to enable managers to initiate conversations with employees with the help of metrics rather than gut-feelings. The result, at least in theory, is a relational style of leadership that is both supported by data and driven by dialogue.

Our findings reveal a more complex picture than the metaphors of augmentation, collaboration, and symbiosis suggest. The platform aims to enhance leadership capacities yet it often hampers dialogue and communication; it aims to inform leadership decisions yet it often obscures how employees are feeling and what they need; and it aims to promote a positive organizational culture yet it often causes uncertainty and disconnection. Given these complexities. The task for us, as critical scholars, is to reflect on what this means for understanding a form of relational leadership that is no longer entirely human nor completely automated.

Reconsidering the AI Autonomy Debate in the Light of Notions of Agency in the Social Sciences

Torben Elgaard Jensen & Anders Søgaard

The idea that AI-systems are somehow gaining agency has recently re-surfaced with reportings on a series of disturbing events, such as AI systems illicitly copying themselves to other servers to avoid being shut down – and lying about it.

Skeptics of these reportings, and their fast circulation in news media, have immediately pointed to the AI industry's strategic interest in stirring up attention-grabbing debates about possible futures, rather than recognizing the industry's already present and manifest problems with environmental damage, labor conditions, copyright violations and concentration of power.

This paper is written in the context of the current AI agency debate, but does not accept the framing of the debate as a question of 'distraction' vs 'genuine risk'. Instead, we wish to explore and problematize the debate's underlying notion of agency that in our view obscures a broader discussion of how to control AI.

In the first part of the paper, we examine the specific contexts in which AI agency reportedly appear. We argue that these situation can be isolated to instances framed as a 0-sum game of control between computer scientists and computers, and where computer scientists find themselves surprised by their own lack of control.

In the second part, we present a series of significantly broader notions of agency drawn from the social sciences: Situated action (Suchman), Non-human agency/Actor network theory (Latour, Law, Callon), Infrastructuring (Star & Ruhleder), Artificial agencies (Runciman), [….].

Finally, we discuss what the AI control problem might entail if we take seriously social science's broad notions of agency.

Title: Stretching the Law: The Need for Contemporary Legislation on Police Use of FR technologies in England and Wales

Tyler Dadge

Abstract: Police forces in England and Wales have been increasingly using live facial recognition (LFR) technology for approximately the last decade. After the Court of Appeal ruling in the 2020 R (Bridges vs South Wales Police) case where it was ruled that legal justifications were limited, a few key documents of guidance were offered to the police from the then Surveillance Camera Commissioner, and the College of Policing. However, despite these pieces of guidance the legislative landscape has remained unchanged and with little discussion around future development to adapt to the police's use of FR. Many of the key pieces of legislation relied upon by the police pre-date the new millennium, such as the Police and Criminal Evidence Act 1984 and the Human Rights Act 1998. This has resulted in questions around how effective existing legislation is at regulating the police and their use of LFR technologies. 21 interviews were conducted in January-September 2023 with key stakeholders including academics, police officers, and those who held relevant positions within government and independent bodies all who had experience with policing and/or the use of FR technology. Subsequently to these interviews, a discourse analysis was conducted on the 2025 Legal Mandate of 8 police forces who had made these publicly available. There was a division amongst the participants; those participants who had a policing background all believed that existing legislation was adequate, however, those participants who didn't have a policing background often believed that existing legislation was "vague" and "patchy". Analysis of the publicly available documents suggests that while they make take the stance that, must like the views of the policing profession participants, existing legislation is adequate, the legal justifications presented in the Legal Mandates remain in a grey area, stretching existing legislative provisions beyond their original intent and assist in developing the argument made by those participants that existing legislation is "vague" and "patchy" and calls for legislation to updated to better reflect the needs of policing in a more contemporary society, regulating the use of AI technologies such as LFR and protecting citizens from technologies that existing legislations were unable to foresee existing.

Between Techno-nationalism and Silence: How Turkish Online News Media Cover Artificial Intelligence

Umut Yener Kara

This study examines how Turkish online news media cover and construct artificial intelligence (AI) by analyzing 5,346 articles from eleven major outlets published between 2017 and 2023, using a mixed-methods design that combines computational text analysis with qualitative reading. Drawing on scholarship in science communication, science and technology studies (STS), and recent work on AI coverage, it explores themes, stakeholders, and ideological differences across outlets.

The findings reveal that media discourse in Turkey not only mirrors issues identified in studies from other countries, such as the dominance of economic and industrial perspectives and the lack of representation of civil society and critical voices, but also highlights an exceptionally fragmented, and to a certain extent, an absent public dialogue on national AI development and application efforts, driven by media partisanship and ideological divides. Pro-government outlets provide extensive, celebratory coverage of state-led AI related initiatives and efforts – especially in the defense industry – anchored in a partisan techno-nationalist ideological formation. In stark contrast, opposition outlets are largely silent on domestic developments, focusing instead on international news, philosophical debates, and cultural applications.

Thus, the study highlights a problem that could hinder responsible AI development and uptake in Turkey, while also contributing to the broader literature on ideological divisions surrounding science and technology.

From Knowledge to Judgment: Addressing the Implementation Gap in Workplace AI Literacy

Veronika Bak

As artificial intelligence (AI) becomes embedded across workplace infrastructures, a gap has emerged between policy ambitions for AI-ready workforces and employees' lived realities of using these systems. Despite AI literacy training since 2020, employees consistently struggle when algorithmic systems become components of their work (Long & Magerko, 2020). For instance, healthcare professionals in clinical settings understand that decision-support systems can produce errors yet fail to identify inappropriate recommendations under time pressure (Kücking et al., 2024). Customer-service representatives in call centres know that AI-generated responses may be contextually inappropriate but lack judgment frameworks to recognise problematic outputs in practice (Adam et al., 2021). Content moderators receive training on algorithmic bias yet cannot identify when biased patterns emerge in their workflows (Roberts, 2021). These recurring failures expose a controversial assumption: that technical literacy alone cannot prepare workers for algorithmically mediated tasks.

Existing frameworks recognise both conceptual knowledge and practical competencies as essential (Long & Magerko, 2020; Ng et al., 2021), yet remain unclear about how to develop situated judgment in practice. This ambiguity enables a problematic default: training systematically privileges technical knowledge while neglecting judgment-in-context (Pinski & Benlian, 2024). Workers may comprehend how algorithms function but lack skills to evaluate outputs within their professional domains or maintain agency when AI becomes part of workflow infrastructures. This paper examines why this implementation gap persists and identifies what judgment capabilities organisations must develop for competent AI use.

The paper conducts a comparative analysis of workplace cases, existing AI literacy frameworks, and organisational training practices. Drawing on examples from diverse industrial settings, it synthesises findings from empirical studies with theoretical arguments on literacy and competence development.

The analysis reveals three contested factors that preserve the implementation gap. First, organisations argue technical instruction is sufficient because situated judgment resists standardised assessment, yet this deflects responsibility for developing capabilities workers need. Second, policymakers defend credential-based metrics as tools for accountability, yet these metrics favour standardised, scalable knowledge over contextual competence. Third, workers and their advocates demand training matching actual workplace judgments, though organisations counter this is impractical to systematise. These competing priorities produce mismatched competence: employees understand algorithms abstractly yet cannot evaluate outputs, recognise harmful recommendations, or integrate system behaviour with domain expertise.

'Situated AI literacy' is introduced in this paper as a reframing in which literacy is positioned as judgment-in-context rather than abstract knowledge. Judgment cannot be improved through generic training alone; it develops through structured practice evaluating AI outputs against domain

standards. The framework proposes three shifts: judgment-first design, contextual competence development through authentic decision scenarios, and accountability reallocation away from individual workers. Critics argue situated judgment is too contextual to systematise, yet professional fields routinely develop such capabilities through case-based learning and supervised practice (Morandini et al., 2023). Without these shifts, AI literacy initiatives risk remaining performative, producing formally trained yet practically unprepared workers, an outcome increasingly indefensible as AI embeds deeper into workplace infrastructure. The implications challenge educators to develop judgment within specific contexts, organisations to specify competence requirements before deployment, and policymakers to move beyond easily measured technical comprehension.

References

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets, 31*(2), 427–445. https://doi.org/10.1007/s12525-020-00414-7

Kücking, F., Hübner, U., Przysucha, M., Hannemann, N., Kutza, J.-O., Moelleken, M., Erfurt-Berge, C., Dissemond, J., Babitsch, B., & Busch, D. (2024). Automation Bias in AI-Decision Support: Results from an Empirical Study. In R. Röhrig, N. Grabe, U. H. Hübner, K. Jung, U. Sax, C. O. Schmidt, M. Sedlmayr, & A. Zapf (Eds), *Studies in Health Technology and Informatics*. IOS Press. https://doi.org/10.3233/SHTI240871

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* 1–16. https://doi.org/10.1145/3313831.3376727

Morandini, S., Fraboni, F., De Angelis, M., Puzzo, G., Giusino, D., & Pietrantoni, L. (2023). The Impact of Artificial Intelligence on Workers' Skills: Upskilling and Reskilling in Organisations. *Informing Science: The International Journal of an Emerging Transdiscipline, 26,* 039–068. https://doi.org/10.28945/5078

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2,* 100041. https://doi.org/10.1016/j.caeai.2021.100041

Pinski, M., & Benlian, A. (2024). AI literacy for users – A comprehensive review and future research directions of learning methods, components, and effects. *Computers in Human Behavior: Artificial Humans, 2*(1), 100062. https://doi.org/10.1016/j.chbah.2024.100062

Roberts, S. T. (2021). *Behind the screen: Content moderation in the shadows of social media: with a new preface.* Yale University Press.

The Colonized Cannot Consent: A Decolonial Analysis of Data Rights, Data Sovereignty, and the Doctrines of Consent Underpinning Large-Scale AI Systems

Wm. Matthew Kennedy

Situated at the intersection of colonial history and sociotechnical analyses of AI, this paper investigates the apparent degradation of the doctrine of (digital) consent, specifically as it pertains to data collection by AI labs for the development of large-scale AI systems. Its analysis excavates the strong and instructive analogy with the period in colonial history during which the very same concept was subject to scrutiny because critics came to argue that Indigenous actors lacked the capacity to give verifiable consent in treaties involving the cession of sovereign rights. Such a revision resulted in conditions that caused 75% of the world's people to become subject to colonial powers. The analogy suggests the same power concentration/alienation has the potential to recur (if it has not already) in the scramble for user data that is accompanying the development of new foundation models—a claim made previously by Shoshana Zubhoff in 2018 in her *The Age of Surveillance Capitalism*.

This paper extends, updates, and renders more precisely Zubhoff's claim. The discussion proceeds in three sections. In the first section, the paper recalls recent efforts on the part of Very Large Online Platforms (VLOPs) to adopt more permissive data collection policies (in many cases surreptitiously). As most frontier AI labs have close linkages with leading VLOPs, such policy changes have large and outsized effects on global practices of digital consent as well as the data pipelines flowing into AI model development. In the second section, the paper reviews the rise of synthetic data generation and practices of synthetic user or market research that such personal data collection enables. Drawing from an emerging record of such practices, (which now are beginning to feature in post-training, fine-tuning, alignment, or safety-testing phases of AI system development), the paper discusses how such practices allow data collectors to continually extract the surplus value of personal data in part because most individuals are structurally incapable of exploiting this value themselves thanks to degraded digital consent. Finally, having demonstrated the utility of colonial history as a source of sociotechnical foresight in the prior two sections, the third section offers some grounded speculation as to our likely futures should we continue on this course. In so doing, it argues for a collective reconfiguration of our geographical imaginaries of digital colonialism, one that collapses historical divisions between a 'minority' and 'majority' world.

Ultimately, the aim of this paper is to recapitulate well-founded critiques by leveraging new information about AI systems design. In many senses, it is a renewed call to action. It does so not only to maximize our chances of creating more just AI systems, but also to ensure a better AI society in the end.