weizenbaum
PROCEEDINGS

Conf 2026
06

Generative AI and Society: What is at Stake?

Weizenbaum Conference 2026 – Book of Abstracts

Weizenbaum Conference 2026

Generative AI and Society: What is at Stake?

BERLIN, 2026

DOI 10.34669/WI.CP/6 \ ISSN 2510-7666

Weizenbaum Institute
Hardenbergstraße 32 \ 10623 Berlin \ Tel.: +49 30 700141-001
info@weizenbaum-institut.de \ www.weizenbaum-institut.de/en

CONFERENCE CHAIRS & VOLUME EDITORS

Florian Butollo
Anne K. Krüger

EDITORIAL MANAGER

Moritz Buchner

WEIZENBAUM INSTITUTE

The Weizenbaum Institute conducts excellent, independent, interdisciplinary and fundamental digitalization research. We provide politics, economy and civil society with well-founded findings and value-based recommendations for action. This helps to ensure that society's digitalization is not only better understood, but can also be shaped in a sustainable, self-determined and responsible manner.

weizenbaum institut

With funding from the:
Federal Ministry of Research, Technology and Space

BERLIN
Senatsverwaltung für Wissenschaft, Gesundheit und Pflege

Contents

Generative AI and Society:
What is at Stake?

Preface

The 8th Weizenbaum Conference, entitled “Generative AI and Society,” takes place in Berlin on 10–11 June 2026. Organized by the Weizenbaum Institute, the conference brings together researchers and practitioners from a wide range of disciplines to examine how generative AI is reshaping society.

The rapid diffusion of generative AI sparks debates across academia, politics, civil society, and the economy. While new applications promise gains in productivity, creativity, and access to knowledge, they also raise pressing questions about power, inequality, dependency, and democratic governance. As generative AI becomes increasingly embedded in everyday life, its implications for work, knowledge production, research, culture, and public discourse remain the subject of intense discussion.

Against this backdrop, the Weizenbaum Conference 2026 provides a forum for critical and interdisciplinary exchange on the societal consequences of generative AI. Following the intellectual legacy of Joseph Weizenbaum, the conference explores the relationship between technological innovation and societal values, with particular attention to the ways in which AI systems may reinforce or transform existing social, economic, and political structures.

The present collection brings together 39 abstracts and offers an overview of the research and perspectives presented and discussed during the conference. Reflecting a wide range of disciplinary and methodological approaches, the contributions address the opportunities, challenges, and societal implications of generative AI across different domains of social life. We would like to thank all authors for their valuable contributions to this volume and to the conference.

Berlin, June 2026

Florian Butollo
Anne K. Krüger

Generative AI and Society:
What is at Stake?

Session 1: Shifting Epistemic Authority

Session 1

Generative AI, Epistemic Power, and the Transformation of Military Knowledge Production

Simon Egbert \\ HSU/UniBW Hamburg \\ Hamburg \\ Germany

KEYWORDS
Military knowledge production; generative AI; NATO; epistemic power; accountability

ABSTRACT
Generative Artificial Intelligence (GenAI) is transforming contemporary warfare by shifting the logic of military violence from physical destruction toward dominance over data, interpretation, and decision-making. Rather than merely optimizing existing practices, GenAI contributes to the emergence of algorithmic warfare as a socio-technical regime in which models synthesize heterogeneous data streams, generate situational images, simulate possible futures, and propose courses of action for human commanders. In this regime, violence is increasingly mediated through computational representations and probabilistic reasoning, reshaping how threats are perceived, evaluated, and acted upon.

Drawing on current military doctrines, industry narratives, and expert debates in NATO member states, this paper examines how GenAI-supported decision systems are framed as solutions to the "fog of war," promising acceleration of operational tempo, enhanced precision, and improved control over complex battlefields. At the same time, these systems reconfigure responsibility, obscure the contingencies of interpretation, and create new dependencies on proprietary infrastructures and defense-tech corporations. Empirically, the analysis is based on a qualitative review of policy documents, industry demonstrations, and strategic concept papers.

The paper foregrounds three interrelated dimensions of epistemic power in GenAI-enabled warfare. First, GenAI systems model reality by generating integrated situational pictures, predictive risk assessments, and simulated scenarios that pre-structure what can be known and anticipated. Second, they frame permissible violence by translating political and legal constraints into algorithmically suggested escalation paths, thresholds, and metrics such as "efficiency," "risk reduction," or "acceptable collateral damage." Third, GenAI systems legitimate decisions by invoking claims of algorithmic objectivity, consistency, and foresight, thereby stabilizing particular interpretations while marginalizing uncertainty, dissent, and alternative judgments.

Situated within the conference stream GenAI in Knowledge Production: Epistemic Power and the Transformation of Expertise, the paper argues that military uses of GenAI produce new forms of expertise and authority. Defense technology firms and AI providers increasingly position themselves as epistemic gatekeepers of future warfare, shifting interpretive and decision-making power away from democratically accountable institutions toward

platformized infrastructures. Human–AI co-production thus transforms military professionalism, blurring the boundary between decision support and semi-autonomous co-decision, and redefining what counts as responsible judgment under conditions of algorithmic mediation.

Theoretically, the paper connects debates on epistemic power, critical studies of automated decision-support systems, and research on military AI to show how generativity promises "total overview" and anticipatory control, while simultaneously deepening asymmetries of power and knowledge. Normatively, it asks what forms of transparency, auditability, and liability are required to democratically govern algorithmic warfare and to safeguard spaces for professional judgment, contestation, and human rights standards in an era of generative war algorithms.

Session 1

Communicative Competence in the Age of ComAI: The Case of Bereavement Support

Andrea Heisse \\ University of Vienna \\ Austria
Marvin Waibel \\ University of Vienna \\ Austria

KEYWORDS
ComAI; sociology of knowledge; ethnography

ABSTRACT
The paper explores the intersection of interpersonal and AI-based bereavement support. Drawing on empirical data from discourse analysis and ethnographic fieldwork, the paper analyses the relationship between conventional bereavement support and tech solutions, focusing on how AI developers envision good bereavement support, build on or neglect established communicative practices, and how conventional services adapt to transforming grieving practices. Developers of AI grief applications promise users a 24/7 companion who is always there to support them throughout their grieving journey. While bereavement support has often been considered a practice immune to automation, artificial intelligence is increasingly discussed in the professional field and is imagined transforming working practices, communication, and companion relations. Hence, it is crucial to understand how this technological shift both draws from and influences the profession, and how it potentially changes the companionship of bereaved individuals. The field of death, dying, and bereavement has become increasingly professionalized over the past decades. Following Pfadenhauer's (2003) notion of professionalism as performative, we will first outline how the profession of bereavement support can be characterized by its communicative patterns. Drawing on discourse analysis (cf. Keller 2013), we trace how professional competence in bereavement support first and foremost entails knowing how to communicate with the bereaved. As

generative artificial intelligence is primarily so successful because it can produce artificial communication (cf. Esposito 2017), we argue that it is precisely the capability to mimic professional communicative patterns, that enables AI developers to appropriate professional fields through communicative AI (Hepp et al. 2023). Drawing on ethnographic fieldwork, we discuss how professions distinguish their practices through redefined conceptualization of communicative competence. As artificial intelligence already capacitates various bereavement support skills, the profession engages in boundary-making processes that are collective negotiations reacting to narratives and experiences, fears and hopes of the replacement, modification or enhancement of a professional practice. Therefore, we aim to outline the interrelated processes of demarcation and appropriation in bereavement support. By analyzing the design and turn takes of AI grief companions, we reconstruct the inscribed notions of good and appropriate communication and companionship. Through a sociology of knowledge perspective, we analyze how communicative AI draws from and reshapes communicative competence in the field of bereavement support and refigures the relationships between bereaved and companion.

REFERENCES

Esposito, E. (2017). Artificial Communication? The Production of Contingency by Algorithms. Zeitschrift Für Soziologie, 46(4), 249–265. https://doi.org/10.1515/zfsoz-2017-1014

Hepp, A., Loosen, W., Dreyer, S., Jarke, J., Kannengießer, S., Katzenbach, C., Malaka, R., Pfadenhauer, M., Puschmann, C., & Schulz, W. (2023). ChatGPT, LAMDA, and the hype around Communicative AI: The Automation of Communication as a field of research in media and communication studies. Human-Machine Communication, 6, 41–63. https://doi.org/10.30658/hmc.6.4

Keller, R. (2013). Zur Praxis der Wissenssoziologischen Diskursanalyse. In R. Keller & I. Truschkat (Eds.), Methodologie und Praxis der Wissenssoziologischen Diskursanalyse (pp. 27–68). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-93340-5_2

Pfadenhauer, M. (2003). Professionalität. Eine wissenssoziologische Rekonstruktion institutionalisierter Kompetenzdarstellungskompetenz. (1st ed.). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-663-11163-4

Session 1

From Measurement to Evaluation: GenAI, Synthetic Imaging Data and Shifting Expertise in Emerging Medical Research

Paula Muhr \\ Brand University of Applied Sciences \\ Hamburg \\ Germany

KEYWORDS
synthetic medical data; epistemic expertise; evaluation practices; similarity metrics; data validity

ABSTRACT
Recent years have seen a rapid uptake of generative AI (GenAI) systems in medical research for the creation of synthetic data: artificial datasets generated to resemble real-world data, aiming to replace or reduce the need for patient-derived data. While scholars have addressed ethical, legal, and epistemic risks—especially regarding privacy, bias, and data governance—less attention has been paid to how these practices reconfigure expertise in the production of image-based medical data, such as retinal fundus images, x-ray, or MRI. Although such practices remain far removed from clinical application, I argue that the use of GenAI for synthetic data generation marks a transformation of epistemic practices within emerging medical research, shifting the locus of expertise from data acquisition to data assessment. Traditionally, expertise in medical data production centered on measurement protocols, sampling strategies, and experimental conditions through which phenomena were translated into imaging data. Questions of validity were oriented toward the relationship between empirical phenomena, instruments, and measurement conditions. When GenAI is used to synthesise artificial data, this relationship is displaced. Synthetic images are not anchored in direct empirical referents but in training datasets and opaque GenAI modelling processes. Hence, validity can no longer be established by reference to measurement but must be negotiated through comparative evaluation practices that position synthetic data relative to the real-world data from which they were derived. I argue that these evaluation practices are organised around a threefold tension between similarity, dissimilarity, and utility. First, synthetic images must be sufficiently similar to real images to preserve clinically and statistically relevant patterns, operationalised through error metrics, distributional comparisons, and feature-level similarity measures. Second, synthetic images must also be sufficiently different, avoiding replication or memorisation of training data while generating novel yet plausible instances. Third, the artificial data's epistemic utility is assessed, for example by comparing the predictive performance of machine learning models trained on real versus synthetic data. Quantitative metrics are increasingly combined with human expert assessments, such as visual inspection and plausibility checks. Together, they form a hybrid validation regime in which fidelity is no longer grounded in representational accuracy but in the ability of synthetic imaging data to perform adequately within algorithmic pipelines and to appear

visually plausible to trained experts. In the context of GenAI-based synthetic data generation, researchers' expertise increasingly shifts from producing data through measurement to evaluating synthetic outputs: defining similarity metrics, interpreting error distributions, and deciding on acceptable deviations between real and synthetic datasets. This relocates epistemic authority from clinical and experimental settings to algorithmic evaluative frameworks. Instead of treating synthetic data merely as a technical solution to data scarcity, this paper frames it as a site where the boundaries between representation and simulation are actively renegotiated, reconfiguring the requirements of human expertise. What is at stake in these early-stage experimental practices is not just the credibility of synthetic image data and their adequacy as sources for generating new medical insights but also the expertise necessary to create, evaluate and interpret such data in the emerging landscape of GenAI.

Session 1

From Language Models to Reasoning Behaviours

Konstantin Mitrokhov \\ Leuphana University Lüneburg \\ Germany
Alexander Campolo \\ Durham University \\ United Kindgom

KEYWORDS
LLM; reasoning; technical genealogy

ABSTRACT
One of the distinctive features of generative AI is the way that technical operations and idiosyncratic conceptualizations from the research community interact to transform longstanding categories related to knowledge and expertise. In this paper, we develop a genealogy to ground one of the most significant of these concepts: reason. By developing a technical genealogy of "reasoning models" we provide a critical account of this emerging form of epistemic power. This account pays close attention to the science of machine learning, particularly reinforcement learning and inference paradigms, with a critical conceptualization of "reason" informed by the history of philosophy and cognitive science. Rather than evaluating these models by some inherited standard of reason, we instead track the contingent development of novel reasoning paradigms, which are already reshaping knowledge and expertise.

This very recent genealogy begins around 2020, when large language models (LLMs) based on transformer architectures had unlocked a wealth of new capabilities through the power of scaling, where performance increased as a function of the size of pre-training datasets, compute, and number of parameters (Kaplan et al. 2020). Although these models could adapt to an astonishing range of linguistic tasks (Brown et al. 2020), researchers, initially working at the margins, began to identify surprising limitations. Many of these involved problems of mathematical reasoning itself, an irony given the centrality of this domain in the history of computing writ large.

The drive to improve the performance of LLMs led to the emergence of models optimised for such "reasoning" tasks. Despite the wide-ranging epistemic critique (Huang et al. 2025; Hong et al. 2025; Kosmyna et al. 2025; Shojaee et al. 2025), such LLMs are being explored for inclusion or informally integrated into existing workflows in scientific, cultural, and political domains (Holtzman & Tan 2025; Naddaf 2025; Moseley 2025; Westwood 2025; Becker et al. 2025; Lin et al. 2025). "Reasoning" as instantiated by these technical systems thus moves across a wide range of scientific specializations. This paper constructs a recent technical genealogy that moves from the automation of language in pre-trained LLMs to so-called reasoning models that have emerged in their wake. It critically characterizes this emerging form of machine learning reason by "reading" a set of post-training techniques for inducing or eliciting reasoning behaviours (Amoore et al. 2023), from chain-of-thought-prompting (Wei et al. 2023) to group-relative policy optimization (Deepseek-AI 2025). The stakes of our genealogy concern a specific shift away from a computational sense of reason that emerged during the twentieth century alongside digital computing (Erickson et al. 2013). In the specific sense we analyze, reasoning emerges less as an explicit aim of rule-based algorithmic formalisation than as a problem that was initially an incidental limitation of LLMs—their inability to handle certain forms of mathematical reasoning. More broadly, accounting for the emergence of this form of "reason" in GenAI provides a theoretical grounding for studies of emerging constellations of epistemic power and expertise as they engage in more complex, multistep tasks.

REFERENCES

Amoore, L., Campolo, A., Jacobsen, B., & Rella, L. 2023. 'Machine learning, meaning making: On reading computer science texts. Big Data & Society 10(1). https://doi.org/10.1177/20539517231166887

Becker, Joel, Nate Rush, Elizabeth Barnes, and David Rein. 2025. 'Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity'. arXiv:2507.09089. Preprint, arXiv, July 25. https://doi.org/10.48550/arXiv.2507.09089

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). 'Language Models are Few-Shot Learners'. arXiv:2005.14165 [cs.CL]. http://arxiv.org/abs/2005.14165

DeepSeek-AI, Guo, D., Yang, D., et al. 2025. 'DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning'. arXiv:2501.12948 [cs.CL]. https://doi.org/10.48550/arXiv.2501.12948

Erickson, P., Klein, J.L., Daston, L., Lemov, R., Sturm, T., & Gordin, M.D. (2013). How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality. The University of Chicago Press.

Holtzman, Ari, and Chenhao Tan. 2025. 'Prompting as Scientific Inquiry'. arXiv:2507.00163. Preprint, arXiv, July 8. https://doi.org/10.48550/arXiv.2507.00163

Hong, Kelly, Anton Troynikov, and Jeff Huber. 2025. 'Context Rot: How Increasing Input Tokens Impacts LLM Performance'. July 14. https://research.trychroma.com/context-rot

Huang, Lei, Weijiang Yu, Weitao Ma, et al. 2025. 'A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions'. ACM Transactions on Information Systems 43 (2): 1–55. https://doi.org/10.1145/3703155

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. 2020. 'Scaling Laws for Neural Language Models'. arXiv:2001.08361 [cs.LG]. https://doi.org/10.48550/arXiv.2001.08361

Kosmyna, Nataliya, Eugene Hauptmann, Ye Tong Yuan, et al. 2025. 'Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task'. Version 1. Preprint, arXiv. https://doi.org/10.48550/ARXIV.2506.08872

Lin, Hause, Gabriela Czarnek, Benjamin Lewis, et al. 2025. 'Persuading Voters Using Human–Artificial Intelligence Dialogues'. Nature, December 4, 1–8. https://doi.org/10.1038/s41586-025-09771-9

Moseley, Simon. 2025. Automating Deception: AI's Evolving Role in Romance Fraud. Alan Turing Institute. https://cetas.turing.ac.uk/publications/automating-deception-ais-evolving-role-romance-fraud

Naddaf, Miryam. 2025. 'Major AI Conference Flooded with Peer Reviews Written Fully by AI'.

Nature, ahead of print, November 27. https://doi.org/10.1038/d41586-025-03506-6

Shojaee, Parshin, Iman Mirzadeh, Keivan Alizadeh, Maxwell Horton, Samy Bengio, and Mehrdad Farajtabar. 2025. The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity. June. https://machinelearning.apple.com/research/illusion-of-thinking

Wei, J., Wang, X., Schuurmans, D. et al. 2023. 'Chain-of-Thought Prompting Elicits Reasoning in

Large Language Models'. arXiv:2201.11903 [cs.CL]. https://doi.org/10.48550/arXiv.2201.11903

Westwood, Sean J. 2025. 'The Potential Existential Threat of Large Language Models to Online Survey Research'. Proceedings of the National Academy of Sciences 122 (47): e2518075122. https://doi.org/10.1073/pnas.2518075122

Generative AI and Society:
What is at Stake?

Session 2: Data Work, Inequality, Reproduction

Session 2

Prompted by Power: Generative AI, Skill Formation, and Subjectivities in Kenya's AI Work Infrastructures

Philip Mong'Are Achoki \\ University of Essex \\ United Kingdom

KEYWORDS
generative AI; AI labour and work infrastructures; skill formation and training regimes; epistemic power and expertise; human–machine interaction; Big Tech dependency; Global South AI work

ABSTRACT
Generative AI (GenAI) is frequently framed as a productivity-enhancing technology that augments knowledge work and democratizes access to expertise. Yet far less attention has been paid to the labour infrastructures and skill regimes that make GenAI systems operational, particularly in the Global South. This paper examines how skills, subjectivities, and work quality are reshaped through everyday encounters with GenAI within Kenya's AI labour ecosystem, a key site in global AI value chains.

Drawing on qualitative interviews with AI data workers, digital-skills trainers, and platform intermediaries, complemented by digital ethnography and critical discourse analysis of corporate training materials, this study investigates how workers are trained to interact with, evaluate, and "prompt" GenAI systems. In the spirit of Joseph Weizenbaum's critique, the paper treats Generative AI not as an autonomous intelligence, but as a socio-technical system whose moral and epistemic consequences are inseparable from the labour relations that sustain it. It shows that GenAI does not simply augment human labour but reorganises it through a corporatised skill formation regime, in which platform-specific certifications, proprietary tools, and opaque performance metrics define what counts as legitimate expertise.

The findings reveal four interrelated dynamics. First, GenAI skill formation is increasingly standardised around Big Tech infrastructures, privileging compliance with platform logics over transferable or critical forms of knowledge. Second, workers experience a paradox of empowerment and dispossession: while acquiring new technical competencies, they lose control over task boundaries, evaluation criteria, and the downstream uses of their labour. Third, human–machine interaction is deeply asymmetrical; workers are required to adapt their cognition, language, and judgment to GenAI systems whose inner workings remain inaccessible, reinforcing epistemic dependency. Finally, GenAI reshapes worker subjectivities, producing forms of self-disciplining, aspirational compliance, and affective attachment to technological futures that remain structurally out of reach.

Conceptually, the paper advances the notion of a Power–Skills Regime to capture how GenAI mediates accumulation, productivity, and labour control through the governance of skills and expertise. By foregrounding the experiences of AI workers in Kenya, the paper contributes a Global South perspective to debates on GenAI and society, challenging dominant narratives that treat GenAI as a neutral tool or purely cognitive innovation.

The paper speaks directly to current concerns about new dependencies on Big Tech, the transformation of knowledge work, and the uneven distribution of benefits in GenAI-driven economies. It argues that understanding what is at stake in GenAI requires close attention not only to models and outputs, but also to the human labour, skills, and subjectivities that sustain them.

Session 2

Beyond Uniform Adoption: Gender, Inequality, and Task-Specific AI Use in Journalism

Sina Thäsler-Kordonouri \\ Ludwig-Maximilians-Universität München \\ Germany
Andreas Riedl \\ University of Augsburg \\ Germany
Tobias Rohrbach \\ Université de Fribourg \\ Switzerland

KEYWORDS
AI; journalism; AI journalism; gender; inequality

ABSTRACT
The increasing use of AI (artificial intelligence) in journalism is often driven by expectations of more efficient workflows for journalists, who are assumed to gain greater professional autonomy through the automation of low-level tasks (e.g., Borchardt et al., 2024). This expectation has given rise to organisational strategies across various media contexts that promote the integration of AI into editorial processes and reward the acquisition of AI skills (Schaetz & Schjøtt, 2025). This strategic logic assumes equal access to AI for all newsroom staff, facilitated by organisational structures. However, research on digital innovation adoption in journalism questions this assumption, demonstrating gendered inequalities in access to opportunities and resources (De Vuyst & Raeymaeckers, 2019) and in the mobilisation of related cultural capital (Holman & Perreault, 2023). These gendered differences are strongly shaped by organisational factors that not only embed inequality within their structures but actively reproduce it (Acker, 1990).

Research on AI use in journalism has so far paid limited attention to inequality-related factors. First empirical findings are contradictory, indicating both unequal (Thurman et al., 2025) and equal levels of AI use among women and men journalists (Fürst et al., 2025). However, these gender comparisons do not differentiate between the type of editorial tasks

supported by AI, for example whether men and women use AI differently for research or fact-checking. As these studies also show that generally journalists' AI use varies substantially by the task type, the existing picture and gender-based comparisons remain insufficiently differentiated.

The present study aims to help fill this research gap. Based on a representative online survey of professional journalists in Germany (n = 1,116), we examined (1) whether—and if so, how—gender-specific differences in AI use exist, and (2) what role perceptions of AI literacy and organisational access to training opportunities play in this context.

Factor analyses reveal four distinct AI use cases among journalists: (a) text processing, (b) audiovisual content creation, (c) investigative work, and (d) management and marketing-related applications. Although women journalists report higher levels of perceived AI literacy than men, our findings mostly reveal no overall relationship between gender and AI use. In one notable exception, women are even more likely than men to employ AI tools for marketing-related purposes. These differences are not mediated by perceptions of disparities in AI-related training opportunities or AI literacy, suggesting that women's engagement with AI may reflect strategic adaptation within specific professional niches rather than differences in perceived skills or access.

Taken together, our findings indicate that AI's impact on journalism is not monolithic. While AI may indeed empower women in some areas, such empowerment tends to occur at the margins of the journalistic field—particularly in roles oriented toward communication management and audience engagement rather than core reporting or investigative work. The study thus highlights how the integration of AI simultaneously opens and constrains opportunities for journalists, underscoring the need to situate technological adoption within broader gendered structures of professional practice.

REFERENCES

Acker, J. (1990). Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations. Gender and Society, 4(2), 139–158.

Borchardt, A., Mulhall, E., Bremme, K., López Garrido, B., & Johanny, Y. (2024). Leading Newsrooms in the Age of Generative AI. European Broadcasting Union. https://www.ebu.ch/news/2025/04/ebu-news-report-focuses-on-leading-newsrooms-in-the-age-of-generative-ai

De Vuyst, S., & Raeymaeckers, K. (2019). Is Journalism Gender E-Qual?: A study of the gendered accumulation and evaluation of digital capital in journalism. Digital Journalism, 7(5), 554–570. https://doi.org/10.1080/21670811.2017.1369357

Fürst, S., Vogler, D., Pfeuti, S., Ryffel, Q., Lombardi, D., Blassnig, S., & Porlezza, C. (2025). Einsatz und Auswirkungen von Künstlicher Intelligenz im Journalismus – Befunde einer schweizweiten Befragung von Medienschaffenden (Vol. 3/2025). Forschungszentrum Öffentlichkeit und Gesellschaft (fög) der Universität Zürich (UZH).

Holman, L., & Perreault, G. P. (2023). Diffusion of innovations in digital journalism: Technology, roles, and gender in modern newsrooms. Journalism, 24(5), 938–957. https://doi.org/10.1177/14648849211073441

Lauerer, C., & Grünewald, M. (2025). WJS3 Methodology. In Journalism Under Duress—Worlds of Journalism Study Report (Wave 3: 2021–2025) (pp. 8–10). Worlds of Journalism.

Schaetz, N., & Schjøtt, A. (2025). AI Hype and its Function: An Ethnographic Study of the Local News AI Initiative of the Associated Press. Digital Journalism, 1–18. https://doi.org/10.1080/21670811.2024.2443163

Thurman, N., Thäsler-Kordonouri, S., & Fletcher, R. (2025). AI adoption by UK journalists and their newsrooms: Surveying applications, approaches, and attitudes. Reuters Institute for the Study of Journalism. https://doi.org/10.60625/RISJ-EA11-Q402

Session 2

Contesting Openness in AI Critiquing the Participatory Potential of Open Source AI

Rainer Rehak \\ Weizenbaum Institute \\ Berlin \\ Germany
André Ullrich \\ Weizenbaum Institute \\ Berlin \\ Germany
Gergana Vladova \\ Humboldt-Universität zu Berlin \\ Germany

KEYWORDS
artificial intelligence; participatory AI; open source; openness; small AI; social sustainability; open-washing

ABSTRACT
Recent research highlights the potential benefits of artificial intelligence (AI) applications for promoting social sustainability, including democratization, equity, and participation. However, many contemporary AI systems are closed, proprietary, and highly centralized. They do not grant access to training data, limit transparency and opportunities for community-led modifications, and hinder the (re)weighting of parameters or post-deployment adjustments—just to name a few concerns. "Open source AI" systems, on the other hand, are supposedly participatory by design due to their openness. Drawing on the widely adopted definition of openness provided by the Open Source Initiative (OSI), derived from the free and open source software (FOSS) domain, this conceptual paper argues that applying this notion of openness to AI systems constitutes merely a discursive move. The transfer of the FOSS-based definition of openness to AI is often intended to transfer the positive and participatory "aura" associated with FOSS. However, an analysis of the differing socio-technical characteristics underlying software and AI reveals significant discrepancies in their structural and normative conditions. Failing to account for these differences risks the "open-washing" of

otherwise non-participatory AI systems. Conversely, recognizing these distinctions opens up alternative pathways for FOSS-like participatory practices regarding AI systems. We conclude with a constructive, two-pronged proposal: First, we call for an extension of the current, predominantly technical notion of openness in AI towards a definition rooted in participatory principles, emphasizing a process-oriented understanding of design and operation. Secondly, we advocate for a stronger focus on "small AI"-systems that can be operated and understood by individuals and small organizations. Such lightweight systems, we argue, allow for transparency, reusability, experimentation, and extensibility and hence enable self-determination and community empowerment through AI technologies—precisely the kind of socio-technical practices associated with the original concept of “openness”. However, this understanding applies to a much smaller subset of AI systems. Our work contributes to the academic and societal discourse on sustainable and participatory AI by drawing from critical data and algorithm studies, science and technology studies, computer science, and participatory research.

This talk is based on Rehak, Rainer; Ullrich, André; Vladova, Gergana. (2025) Contesting Openness in AI. Reflections on the transformative value and participatory potential of open source AI. In: Proceedings of the 2025 ACM International Conference on Information Technology for Social Good (GoodIT '25).

Session 2

The Household Impact of Generative AI: Evidence from Internet Browsing Behavior

Gregor Schubert \\ University of California Los Angeles \\ United States
Miao Ben Zhang \\ University of Southern California \\ Los Angeles \\ United States
Michael Blank \\ Stanford University \\ United States

KEYWORDS
household finance; generative AI; internet browsing; technology adoption; inequality; chatbots

ABSTRACT
This paper studies the impact of generative AI on U.S. households using detailed internet browsing data from a large sample of home devices during 2021–2024. Our analysis of households' adoption of ChatGPT reveals several new empirical findings: First, we show that private household adoption of generative AI has been rapid but uneven, creating a substantial "generative AI divide" among households, as high-income and younger households adopt generative AI faster than low-income and senior households. Second, we develop a new measure of household exposure to generative AI technology shocks based on based on households' pre-ChatGPT 2021 browsing patterns and show that it predicts households' later GenAI adoption. Third, using pre-adoption exposure as an instrument, we show that adopting ChatGPT causes an increase in households' leisure browsing on home devices, while leaving time spent on productive digital tasks unchanged. Fourth, we provide empirical evidence that households tend to use ChatGPT in the context of productive non-market online tasks, such as education or job search, rather than for digital leisure activities. Together, these findings suggest that generative AI increases households' leisure time by making productive digital activities more efficient. Using a quantitative model, we show that these estimates suggest a sizable household productivity gain from generative AI.

Generative AI and Society:
What is at Stake?

Session 3: New Skills and Capacities at Work

Session 3

GenAI: Changes in Experiential Knowledge and Workplace Learning in Organizations

Annika Becker \\ University of Duisburg-Essen \\ Germany
Frank Kleemann \\ University of Duisburg-Essen \\ Germany

KEYWORDS
digitized knowledge work; experiential knowledge; workplace learning

ABSTRACT
The presentation deals with changes in digitized knowledge work due to the implementation of generative AI in work processes. The study is based on 32 narration-based qualitative interviews with employees in various positions at two companies in the fields of corporate real estate and IT consulting. Employees in both companies use large language models (LLMs) in their work practices. The presentation explores the qualities of experiential knowledge and associated workplace learning that are necessary to make (productive) use of generative AI in organizations. Our empirical findings are limited to LLMs, as the most prominent form of GenAI tools currently used. However, our contribution intends to initiate a discussion if and to what extent our findings on the relevance of experiential knowledge and the organizational design of learning processes of LLMs in organizations can be generalized to other generative AI models.

The presentation of our empirical findings will, in a first step, provide an overview how LLMs are applied in the two companies and what the interests of employees and organizations are. presented. It shows that the integration of AI tools into the work process is primarily efficiency oriented. In a second step, we focus on requirements to integrate LLMs in work practice. Our main finding is that of central importance for successfully implementing an AI tool are knowledge how to use its capabilities, continuous critical reflection of the tool's output, and an iterative (and "interactive") processing of the tool's outputs (i.e., an [imagined] "dialogue", with the tool). Thus, the use of AI tools in work processes requires workers to draw on meta-knowledge about what generative AI is capable to produce, whether the provided content is accurate, and where its limitations lie. Such meta-knowledge is continuously augmented "on the job" in the use of AI tools; which points to the third step of our analysis: the changes in learning processes brought about by the use, and particularly by the continuous development of generative AI tools. This leads to a particular interweaving of formal and informal learning practices in organizations. Individual experimentation and engagement with specific digital tools (which are of interest to subjects in specific situations) and learning by doing are followed by the mutual transfer of practical knowledge, acquired skills, and experience – both informally and in collective, formally organized sessions. Integration of knowledge and skills is not top-down but is co-designed by employees based on their own needs. The processes of knowledge transfer are not specifically coordinated but should be

viewed as contingent – depending on the needs and interests of employees and on economic considerations of the organization.

Session 3

From Augmentation to Agency: Transformational Skills in Human–GenAI Knowledge Work

Kendra Pöhlmann \\ University of Augsburg \\ Germany

KEYWORDS
generative AI; knowledge work; human–machine interaction; transformational skills; sustainability; twin transformation

ABSTRACT
The rapid diffusion of generative AI (GenAI) is profoundly reshaping knowledge work. Beyond questions of productivity and task substitution, GenAI reconfigures how knowledge is created, evaluated, and enacted in everyday work practices. While current debates often focus on efficiency gains or technological capabilities, less attention has been paid to the human capacities required to engage productively, critically, and sustainably with GenAI systems. This paper argues that the emerging transformation of knowledge work cannot be adequately understood without considering transformational skills as a central mediating factor in human–machine interaction.

Building on research at the intersection of AI & sustainability and organizational psychology, the paper conceptualizes transformational skills as a bundle of dispositional human capacities—including emotional self-regulation, reflexivity, ambiguity tolerance, ethical judgment, and sense-making—that enable actors to navigate complexity, uncertainty, and socio-technical change. These skills shape how knowledge workers appropriate GenAI tools, interpret their outputs, and integrate them into decision-making and creative processes. Rather than treating GenAI as a neutral productivity tool, the paper frames human–GenAI interaction as a dynamic co-regulative process in which experiential knowledge and subjective judgment remain decisive.

Empirically and conceptually, the contribution addresses three core dimensions of the knowledge-work stream outlined in the conference call. First, it examines shifting boundaries between substitution and augmentation. While GenAI can automate certain cognitive tasks, its effective use increasingly depends on human capacities to contextualize, question, and ethically assess AI-generated outputs. Second, it discusses winners and losers of the transformation of knowledge work. The paper argues that disparities will not solely emerge along lines of technical expertise, but also along access to and cultivation of transformational skills, potentially reinforcing existing inequalities in organizations and labor markets. Third,

it highlights the growing relevance of experiential knowledge in working with AI, emphasizing how tacit learning, reflective practices, and emotional regulation influence trust, overreliance, and critical distance in human–GenAI collaboration.

From a sustainability perspective, the paper situates these dynamics within the broader framework of the "twin transformation" of digitalization and sustainability. It cautions against purely efficiency-driven narratives of GenAI adoption that risk rebound effects, intensified workloads, and erosion of professional autonomy. Instead, it proposes a human-centered and sustainability-oriented perspective on knowledge work, in which transformational skills function as leverage points for aligning technological innovation with long-term social and organizational resilience.

The paper contributes to current debates on GenAI and society by shifting the analytical focus from technological capabilities to human agency in knowledge work. It offers a conceptual framework for understanding how new forms of human–machine interaction are shaped by skills beyond technical literacy and discusses implications for organizational design, skill development, and governance of GenAI in knowledge-intensive settings.

Session 3

Strategic Simplicity Gets the Most: Evaluating Prompting Techniques in AI-Assisted Fact-Checking

Laurence Dierickx \\ University of Bergen \\ Norway
Andreas L. Opdahl \\ University of Bergen \\ Norway
Gustav Linden \\ University of Bergen \\ Norway

KEYWORDS
human-machine interaction; prompting; large language models; fact-checking

ABSTRACT
Despite their well-documented limitations in accuracy, consistency, and contextual reasoning, large language models are increasingly used in fact checking and journalistic verification workflows. While they can assist with tasks such as summarisation, headline drafting, and claim evaluation, they also generate irrelevant additions and fabricated content, which undermines their reliability in information sensitive environments. Computer science research has attempted to mitigate these risks through advanced prompt engineering techniques, yet these methods are often technically complex, inconsistent across models, and inaccessible to non-expert users. Although prompting research has grown significantly, particularly in efforts to identify strategies suited to non-expert use, there remains limited systematic and empirical evidence comparing the effectiveness of prompting techniques across multiple models and tasks. This study addresses this gap by investigating which prompting strategies best support non expert users in fact checking with large language models and whether simple prompts can improve reliability, stability, and task performance.

The study employs a comparative experimental design to evaluate seven prompting techniques selected for their accessibility and practical relevance to non-expert users. These techniques were applied to three core fact checking tasks: headline generation, content summarisation, and verdict formulation. Fifteen published fact checks were selected to represent diverse topics and verdict types, simulating realistic verification conditions. Each fact check was processed using all seven prompting techniques across the three tasks and four widely accessible large language models, producing a manually curated dataset of two thousand five hundred and twenty outputs.

A two-stage prompting design was used. In the first stage, manually created prompts were applied to each task. In the second stage, prompts were refined through meta prompting, whereby a large language model is instructed to improve the clarity and structure of the original instructions while preserving the task's intent. Comparing these stages enabled systematic evaluation of the effect of prompt optimisation on performance across techniques, tasks, and models.

Evaluation combined automated natural language processing metrics with human judgement to capture both technical performance and epistemic quality. Automated metrics assessed the degree to which model outputs aligned with reference texts in terms of meaning, wording, structure, and accuracy. Human evaluation examined whether each task was completed correctly, whether content was factually accurate, and whether summaries and headlines were clear, concise, and consistent. All outputs were assessed twice to ensure reliability.

The results indicate that effective prompting depends less on complexity and more on clear wording and precise task definition. Simple vanilla prompts consistently produced the most accurate, concise, and stable outputs across tasks and models. More complex techniques occasionally improved performance but frequently reduced stability or introduced unnecessary verbosity. By providing systematic empirical evidence in a field where comparative studies remain scarce, this study frames prompting as a form of linguistic engineering in which semantic structure mediates human intent and machine interpretation. It therefore offers practical guidance for high stakes, information sensitive domains where non expert users must interact with AI systems.

Session 3

Dead Labor Learns to Reason: GenAI and the Manufacture of Knowledge

Peter Schulz \\ Friedrich-Schiller-Universität Jena \\ Germany

KEYWORDS
GenAI; knowledge work; scientific knowledge; sociology of science; critical sociology

ABSTRACT
The diffusion of generative AI into research has reignited long-standing debates about the nature of scientific knowledge, expertise, and discovery. Much of the current discussion oscillates between utopian visions of unleashed innovativeness and dystopian concerns about deskilling and automation. This paper argues that both positions miss an important point and the integration of GenAI into research practices should rather be understood as part of a historically continuous reconfiguration of the relation of current human labor and materialized past labor.

The theoretical core of the paper is a re-reading of Karin Knorr Cetina's early work on the manufacture of scientific knowledge in light of GenAI and its goal is to link debates from the philosophy of science (Kuhn, M. Polanyi) with Marxian approaches to automation and the distinction between living and dead labor.

Knorr Cetina's understanding of scientific knowledge as the product of knowledge work provides a fruitful starting point for a sociological discussion of GenAI. In her account, scientific

knowledge does not derive its valuation primarily from correspondence with an external reality, but from its pragmatic success within the scientific field. Therefore, the question of epistemic validity shifts toward the reproduction of explicit and implicit standards of representing scientific knowledge. At the same time, she emphasizes that scientific practice relies on situated capacities that cannot be reduced to formal rule-following: she names the ability to apprehend situations holistically and to operate across multiple levels simultaneously. Such capacities resonate strongly with Michael Polanyi's notion of personal knowledge, which he grounds in embodied skills of pattern recognition and stochastic inference. The central role of humans in this process is questioned because both, the reproduction of explicit and implicit standards as well as pattern recognition and stochastically drawn conclusions are the core functional principles of GenAI.

But Knorr Cetina conceptualizes scientific knowledge production as a chain of selections and translations in which materialized results of prior knowledge work like laboratory equipment and computer programs become the objects of further selection and transformation. This allows the analysis of relations between living knowledge work and its materializations within the scientific process, and GenAI can be understood as a particularly dense form of materialized "dead labor": it incorporates prior selections into algorithmic form and, in doing so, further reshapes the composition of scientific knowledge work towards the dominance of dead labor.

This reformulates the question of AI and scientific knowledge work: How does GenAI redistributes cognitive labor within established research practices? Drawing on Kuhn's notion of normal science, as well as Toulmin's and Polanyi's emphasis on the creative role of individual judgment, one might argue that a widespread use of GenAI further constrains scientific knowledge within established paradigms. But from Knorr Cetina's perspective that scientific innovation is the result of incremental transformations emerging from discursive interaction within scientific fields, GenAI rather intervenes in these interactions by automating selected aspects of judgment and cognition and therefore alters the conditions under which scientific knowledge is produced without fundamentally transforming the logic of scientific progress itself.

Generative AI and Society:
What is at Stake?

Session 4: Generative AI in Knowledge Work

Session 4

Agentic AI and Journalistic Knowledge Making in PSM Newsrooms – Who Orchestrates Whom?

Pauline Reitzer \\ University of Vienna \\ Austria

KEYWORDS
agentic AI; journalistic practice; knowledge-making; communicative AI

ABSTRACT
European PSM newsrooms rejected AI as a research tool, citing concerns about bias, hallucinations, and the use of probabilistic systems as risky for journalistic practice. Yet, within the short period of the last year, AI "agents" for research purposes were becoming central to PSM newsroom innovation, reflecting a broader transformation in how journalistic knowledge work is organised. This rapid adoption responds to mounting challenges: the proliferation of synthetic content, the persistence of fake news, and the sheer volume of information now available make traditional research increasingly complex and time-consuming. In this context, agentic AI emerges as a logical solution, capable of navigating vast, fragmented, and machine-readable data landscapes. These developments situate journalism within what could be termed after Knoblauch (2017) the era of intra-activity, in which machines do not merely assist humans, but increasingly communicate and operate with other machines, reshaping knowledge-making in ways that carry direct implications for democratic public life.

In my project I examine how communicative AI reshapes journalistic practices when newsrooms introduce "AI agents" as research tools. As journalists will increasingly collaborate with agentic AI in the research phase of news production, the paper asks: How does agentic AI affect knowledge-making processes in the PSM newsroom?

The paper draws on ethnographic observations from an ongoing case study at a European public service media organisation and follows the introduction of an AI research agent throughout its "beta" phase. Methodologically, the AI tool itself is approached as an ethnographic field site. The analysis combines an examination of its technical affordances over time (prompt-based walk-through), as well as it explores its "black box" through semi-structured execution logs. These materials are analysed with regard to the communicative patterns through which the system structures and anticipates journalistic knowledge-making practices.

Preliminary findings connect transforming journalistic research practices to broader transformations in the ecosystem of information retrieval and synthesis and underline the need for further research on agentic systems in knowledge-making processes in the news industry and across domains.

Session 4

GenAI at Work: Promise, Practice, and Impact

Florian Butollo \\ Goethe University Frankfurt \\ Germany
Christine Gerber \\ WZB Berlin Social Science Center \\ Berlin \\ Germany
Ann-Kathrin Katzinski \\ WZB Berlin Social Science Center \\ Berlin \\ Germany
Marlene Kulla \\WZB Berlin Social Science Center \\ Berlin \\ Germany

KEYWORDS
generative AI; work; programming; journalism; productivity; automation; augmentation

ABSTRACT
By processing significant shares of the world's knowledge, generative AI (genAI) is expected to reduce the effort to produce text, images and other output that resemble the ones by humans. In public discourse, genAI is associated with productivity hikes in knowledge work and a renewed threat of a substitution of work in areas that were thought to be immune to automation.

In our contribution we present empirical findings on the adoption of genAI "in the wild" (Sun et al. 2024). We shed light on concrete introduction processes, use-cases and consequences for workers in two distinct fields of knowledge work: journalism and IT programming. Our empirical findings are based on 32 expert interviews and seven qualitative case studies (43 interviews) in journalism and programming (plus one group meeting and observation) as well as a netnography in programming. They highlight three major points.

First, even though productivity hikes could be proven in experimental settings and analyses of single tasks, the real impact on overall individual and organizational productivity is less clear. The usage of genAI within the complexity of real-life work situations requires additional work by individuals and organizations to adopt genAI. This includes in particular the search for relevant use cases and objectives, the choice of appropriate tools, the adoption of necessary skills through experimentation, the contextualization of outputs and possibly also the transformation of work organization.

Second, genAI can be used to raise the productivity of single tasks and thereby substitute work, but it is often also used in a complementary fashion, by which tasks can be fulfilled that were out of reach for human actors so far. Such cases of an "augmentation" of work depend on the abilities of human workers to operate genAI systems at the frontier of new possibilities. GenAI results decisively depend on human work capacity.

Third, interaction with genAI systems results in new forms of human-machine-interaction that involve the ability to delegate tasks to AI systems and to engage in co-creation between humans and genAI systems. Humans working with genAI systems show different types of "boundary work" (Langley et al. 2019) that redefines their professional roles, in relation to genAI applications, but also in comparison with genAI-using colleagues whose tasks, skills,

and responsibilities may have shifted since the introduction of genAI. Issues that affect work quality (e.g. autonomy, hierarchy, work intensity) are negotiated within such dynamic relationships.

Overall, our findings suggest that genAI is not simply entering knowledge work as a tool of efficiency, but as a contested technology that oscillates between expectations and workplace practices, between persistent professional cores and the reconfiguring of work. Rather than delivering the expected productivity gains, its effects emerge only through situated human labour, organizational adaptation, and professional boundary work. By tracing these dynamics in journalism and programming, our paper addresses a central question of our time: not only if generative AI will transform knowledge work, but under what conditions, for whom and with what consequences.

REFERENCES

Langley, A., Lindberg, K., Mørk, B. E., Nicolini, D., Raviola, E., & Walter, L. (2019). Boundary work among groups, occupations, and organizations: From cartography to process. Academy of management annals, 13(2), 704-736.

Sun, Y., Jang, E., Ma, F., & Wang, T. (2024). Generative AI in the wild: Prospects, challenges, and strategies. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, 1-16.

Session 4

Beyond Efficiency: Effects of Generative AI on Productivity in High-Skilled Knowledge Work

Leon Hellbach \\ Weizenbaum Institute \\ Berlin \\ Germany
Philip Wotschack \\ Weizenbaum Institute \\ Berlin \\ Germany

KEYWORDS
generative AI; productivity; knowledge work; white collar work; quality of work; innovation; efficiency; output

ABSTRACT
Estimates of productivity effects due to generative AI adoption have been provided by a number of surveys from management and economics literature, pointing at considerable time savings (Necula et al. 2024). However, their findings remain limited. On the one hand, they often suggest universal effects of generative AI by neglecting differences in the organizational and work-place context. As studies from organizational and labor sociology have shown, the implementation and utilization of new technologies is strongly shaped by organizational practices, institutional structures, and organizational actors (Joyce et al. 2023). Moreover, productivity effects of generative AI are often reduced to efficiency and time savings. Literature on productivity in knowledge work shows that it consists of more than just output-oriented dimensions (Ramírez & Nembhard 2004).

Our paper sheds light on these issues by studying the implementation of a genAI system in a large German industrial company, asking: (1) How does the use of genAI in high-skilled knowledge work change work tasks, workflows, and work results? (2) How do these effects relate to classic notions of productivity defined by input-output relations? To what extent do they go beyond?

First, we carefully reconstruct effects of genAI adoption at the workplace within the organizational context, building on the distinction between automation and augmentation effects (Zhou et al. 2021). Second, we investigate to what extent genAI effects translate into productivity gains in classical terms of input-output relations, and within concepts that go beyond these terms.

The empirical base of our study is an in-depth pre-post company-level case study combining qualitative and quantitative data. Before and after the introduction of a large language model (LLM) in a large German industrial company, members of the management, works council, their lawyer, the responsible trade union, a technical advisory service and 17 pilot users have been interviewed. Moreover, a standardized longitudinal survey among 64 pilot users in the company has been conducted. Most of the interviewees are high-skilled employees with an academic agree, often working in leading functions. To cover a broad variety of white-collar

occupations and tasks, pilot users from 11 departments have been selected. The longitudinal design gives us the opportunity to observe effects of AI adoption over time.

While our survey gives us a quantitative overview of general trends within the company, our in-depth qualitative approach makes it possible to carefully explore work related changes and effects due to AI adoption.

Our study indicates substantial efficiency gains among workers by automation and augmentation of tasks. Vital augmentative effects of genAI beyond efficiency gains lie (1) in the improvement of quality and (2) the innovation of new outputs and workflows, showcasing that effects of genAI go beyond narrow conceptions of productivity measured by input-output relations. They overlook qualitative changes and innovations, which are shaped by social practices within the organization and are essential to the very foundation of high-skilled knowledge work on the one hand, and to the competitiveness and rising demands of dynamic markets on the other hand.

REFERENCES

Necula, Sabina-Cristiana; Fotache, Doina; Rieder, Emanuel (2024): Assessing the Impact of Artificial Intelligence Tools on Employee Productivity: Insights from a Comprehensive Survey Analysis. In: Electronics 13 (18), S. 3758. https://www.doi.org/10.3390/electronics13183758

Ramírez, Yuri W.; Nembhard, David A. (2004): Measuring knowledge worker productivity. In: Journal of Intellectual Capital 5 (4), S. 602–628. https://www.doi.org/10.1108/14691930410567040

Zhou, Lina; Paul, Souren; Demirkan, Haluk; Yuan, Lingyao; Spohrer, Jim; Zhou, Michelle; Basu, Julie (2021): Intelligence Augmentation: Towards Building Human- machine Symbiotic Relationship. In: THCI 13 (2), S. 243–264. https://www.doi.org/10.17705/1thci.00149

Session 4

Managing Emergent Work Practices: HR Initiatives and Generative AI Use

Sonja Koehne \\ Humboldt Institute for Internet and Society \\ Berlin \\ Germany

KEYWORDS
generative artificial intelligence; human resource management; organizational control

ABSTRACT
As generative artificial intelligence (GenAI) diffuses organizations, managing its introduction and use poses distinct challenges to organizations. While discriminative AI systems can make classifications and predictions, GenAI systems use training data to produce seemingly new content (Feuerriegel et al., 2024; Jebara, 2004). Existing research suggests that GenAI is different from discriminative AI because it is more readily available, more broadly applicable, and can be used more exploratively by workers (Baygi & Huysman, 2025; Krakowski, 2025). These features have important consequences for organizations. For example, the public availability of GenAI through tools such as ChatGPT enables workers to use GenAI tools not authorized by the company for work purposes.

When new technologies are introduced to organizations, HR departments are typically tasked with developing training programs, drafting guidelines, and managing organizational change (Budhwar et al., 2023; Nyberg et al., 2025). These initiatives often assume shared use cases among workers, partially aligned learning trajectories, and relatively stable technologies. For example, a new enterprise resource planning software is typically introduced with standardized training modules, predefined workflows, and clearly specified user roles. In the case of GenAI, however, prior research suggests that its use is highly personalized and context-specific (Huysman, 2025; Krakowski, 2025), which gives rise to emergent and individualized work practices. Such developments challenge established HR approaches and call for new modes of engagement or what Nyberg et al. (2025, p. 11) refer to as an "experimentation imperative with GenAI." Against this backdrop, I ask the following research question: How do HR initiatives attempt to maintain organizational control over emergent and individualized work practices?

To answer this question, I draw on 15 preliminary interviews with HR practitioners involved in diverse roles and initiatives related to GenAI implementation. My preliminary analysis suggests that while the use of GenAI challenges organizational control, HR teams respond with initiatives that tolerate ambiguity and iteratively limit risk. I organize the preliminary findings in three themes. First, GenAI use leads to an erosion of organizational control because work practices become dynamic and unscripted. These emergent practices cannot be easily stabilized or monitored, resulting in governance challenges for HR that stem from individualized use rather than worker resistance. Second, HR practitioners engage in initiatives that tolerate ambiguity and aim less at formalizing behavior and more at encouraging

use and making individualized practices organizationally acceptable, such as "promptathons." Third, HR practitioners engage in initiatives that iteratively limit risk and make emergent and individualized practices safe, such as by providing rapid access to safe and approved GenAI technologies that are regularly updated and by continuously revising policies and rules of use alongside everyday experimentation.

Session 4

Expectations of Digitalization: Explaining AI Adoption in Organizations

Niklas Ullrich \\ Zeppelin University \\ Friedrichshafen \\ Germany
Florian Muhle \\ Zeppelin University \\ Friedrichshafen \\ Germany

KEYWORDS
artificial intelligence; digitalization; organizations; Systems Theory, Mixed-Methods Research

ABSTRACT
Abstract. Artificial Intelligence (AI) has become a focal point of public debate, oscillating between promises of unprecedented progress and fears of disruption. In particular, communicative AI applications—such as ChatGPT, Gemini, Claude, Copilot, and DeepSeek—have gained widespread attention and impressed consumers and investors alike. While their technological performance can either be framed as impressive or disappointing, AI has become an unavoidable reference point for organizational communication. The use of AI in organizations across multiple societal spheres is increasing, and the topic is deemed highly relevant for the future. This development is especially pronounced in the economic system, with recent studies revealing that CEOs globally are most concerned with transforming their businesses quickly enough to keep up with technology, including AI. This is impressive, as many furthermore report, they have not yet realized financial returns from their AI endeavors. Nonetheless, AI is pursued in large corporations, small- and medium-sized enterprises, and by freelancers alike.

Against this backdrop, our paper asks how organizational decisions regarding the implementation of communicative AI can be explained. The question arises because neither the technological capabilities of AI applications nor the established explanations from Science and Technology Studies (STS) adequately capture the dynamics observed in organizations. While imaginaries, narratives, myths, or visions help to account for how shared meaning around AI is culturally produced, these approaches insufficiently explain how organizations translate such meaning into concrete decisions. We argue instead that a systems approach provides a more precise account of AI adoption as an organizational phenomenon and the underlying (societal) processes.

From a systems-theoretical perspective, organizations are "decision machines" that continuously transform uncertainty into decidable communication. At their core lies membership, which is stabilized by formalized expectations. Deciders in organizations are confronted with expectations and reflexive expectations and also occupy boundary positions, which come with additional responsibilities and the ability to introduce information into the system. The deciders are confronted with expectations and reflexive expectations to digitalize—to implement AI. These expectations to digitalize lead to corresponding decisions. With decisions being the primary form of communication of organizations. These decisions then "travel" through the respective organization, leading to recursive follow-up decisions that refine and edit the initial decision to implement AI.

We demonstrate this process and explain how organizational decisions regarding the implementation of communicative AI come about through a case study of a large industrial corporation that introduced Microsoft Copilot company-wide. Our study follows a mixed-methods approach, combining quantitative surveys and qualitative research methods, primarily interviews, observations, and document analysis. Our material outlines the process: how AI emerged as a critical issue, how expectations and reflexive expectations structured initial and subsequent decisions, and how those decisions recursively shaped subsequent implementation steps.

The paper contributes a systems-theoretical explanation of (AI) adoption as expectation-driven. The perspective clarifies how AI arises as a topic in organizations and why and how adoption follows even under conditions of technological ambiguity. It extends sociological research on digitalization, organizational change, and emergent technologies. The paper explores the intersection of interpersonal and AI-based bereavement support. Drawing on empirical data from discourse analysis and ethnographic fieldwork, the paper analyses the relationship between conventional bereavement support and tech solutions, focusing on how AI developers envision good bereavement support, build on or neglect established communicative practices, and how conventional services adapt to transforming grieving practices. Developers of AI grief applications promise users a 24/7 companion who is always there to support them throughout their grieving journey. While bereavement support has often been considered a practice immune to automation, artificial intelligence is increasingly discussed in the professional field and is imagined transforming working practices, communication, and companion relations. Hence, it is crucial to understand how this technological shift both draws from and influences the profession, and how it potentially changes the companionship of bereaved individuals. The field of death, dying, and bereavement has become increasingly professionalized over the past decades. Following Pfadenhauer's (2003) notion of professionalism as performative, we will first outline how the profession of bereavement support can be characterized by its communicative patterns. Drawing on discourse analysis (cf. Keller 2013), we trace how professional competence in bereavement support first and foremost entails knowing how to communicate with the bereaved. As generative artificial intelligence is primarily so successful because it can produce artificial communication (cf. Esposito 2017), we argue that it is precisely the capability to mimic professional communicative patterns, that enables AI developers to appropriate professional fields through communicative AI (Hepp et al. 2023). Drawing on ethnographic fieldwork, we discuss how professions distinguish their practices through redefined conceptualization of communicative

competence. As artificial intelligence already capacitates various bereavement support skills, the profession engages in boundary-making processes that are collective negotiations reacting to narratives and experiences, fears and hopes of the replacement, modification or enhancement of a professional practice. Therefore, we aim to outline the interrelated processes of demarcation and appropriation in bereavement support. By analyzing the design and turn takes of AI grief companions, we reconstruct the inscribed notions of good and appropriate communication and companionship. Through a sociology of knowledge perspective, we analyze how communicative AI draws from and reshapes communicative competence in the field of bereavement support and refigures the relationships between bereaved and companion.

REFERENCES

Esposito, E. (2017). Artificial Communication? The Production of Contingency by Algorithms. Zeitschrift Für Soziologie, 46(4), 249–265. https://doi.org/10.1515/zfsoz-2017-1014

Hepp, A., Loosen, W., Dreyer, S., Jarke, J., Kannengießer, S., Katzenbach, C., Malaka, R., Pfadenhauer, M., Puschmann, C., & Schulz, W. (2023). ChatGPT, LAMDA, and the hype around Communicative AI: The Automation of Communication as a field of research in media and communication studies. Human-Machine Communication, 6, 41–63. https://doi.org/10.30658/hmc.6.4

Keller, R. (2013). Zur Praxis der Wissenssoziologischen Diskursanalyse. In R. Keller & I. Truschkat (Eds.), Methodologie und Praxis der Wissenssoziologischen Diskursanalyse (pp. 27–68). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-93340-5_2

Pfadenhauer, M. (2003). Professionalität. Eine wissenssoziologische Rekonstruktion institutionalisierter Kompetenzdarstellungskompetenz. (1st ed.). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-663-11163-4

Generative AI and Society:
What is at Stake?

Session 5: Generative AI in Politics

Session 5

Between Human Judgment and Algorithmic Vision: Epistemic Power in AI-Assisted Medicine

Michael Heinlein \\ ISF Munich—Institute for Social Science Research \\ Germany
Judith Neumer \\ ISF Munich—Institute for Social Science Research \\ Germany

KEYWORDS
algorithmic automation; artificial intelligence; clinical gaze; epistemic power; machine vision; medical imaging; professional vision; visual practice

ABSTRACT
The talk examines how Artificial Intelligence (AI) reshapes professional work by intervening in the very act of seeing. Drawing on ethnographic observations and interviews in German gastroenterology departments, it analyses the situated use of AI-assisted polyp detection systems in colonoscopy to show how algorithmic pattern recognition reorganises the clinical gaze, and with it the conditions of professional authority and expertise. Although AI-assisted polyp detection is not a generative system in a narrow sense, it exemplifies how data-driven visual AI reconfigures professional perception in ways analogous to generative systems: both redistribute epistemic authority, reorganise attention and redraw the boundaries between human judgement and machine output. The paper primarily contributes to the stream on epistemic power and the transformation of expertise in AI-mediated knowledge production.

In colonoscopy, AI systems conduct real-time visual analysis by marking suspected polyps with bounding boxes that may guide, interrupt or contradict the physician's gaze. These operations turn the automation of perception into a site of epistemic power struggles over what counts as legitimate clinical knowledge. Rather than simply augmenting professional competence, AI introduces new asymmetries between embodied judgement and algorithmic indication.

Based on our fieldwork, we identify three recurrent modes of practice—technology-driven, experience-driven and interrupted—that illuminate how clinicians negotiate, resist and recalibrate algorithmic knowledge production in situ. These modes show that epistemic power in AI-assisted work is enacted through continuous attention management and alignment: physicians must decide when to trust, follow, question, or override algorithmic cues in their clinical decision-making, thereby performing an ongoing, embodied negotiation of expertise, integrity and legitimacy under algorithmic conditions. Clinicians do not merely react to AI outputs but actively interpret professional expectations about what the system does, should do and when intervention is warranted.

Conceptually, the analysis draws on practice- and subject-theoretical perspectives from the sociology of work to foreground the sensory, bodily and cognitive dimensions of professional vision. Algorithmic systems formalise perceptual routines and standardise diagnostic

attention by externalising elements of professional judgement into machine vision. This does not straightforwardly deskill physicians; instead, it recomposes the labour process. Expertise is no longer secured solely through knowledge and skill but increasingly depends on the capacity to integrate and govern a sociotechnical assemblage that renders professional perception both visible and (partially) manageable.

The 'algorithmic turn' in medicine thus reconfigures medical work at technical and normative levels: it reframes the locus of knowledge production, redistributes epistemic power and reshapes accountability in hybrid constellations of human and machine agency. In this light, the AI-assisted clinical gaze appears less as augmented stability than as a site of contested coordination. Understanding these dynamics requires an empirical focus on the labour process itself, where ascribed roles and expectations of AI, human expertise and sociotechnical assemblages materialise in situated modes of performance: a contingent alignment of algorithmic vision and professional practice.

Session 5

Repositioning the Sword of Damocles: China's State-led and Development-First AI Governance Pivot

**Boli Yang ** The Chinese University of Hong Kong \\ China

KEYWORDS
AI governance; China; development first; authoritarianism; state-led

ABSTRACT
This article examines the pivot in China's approach to AI governance, tracing its evolution from initial regulatory alignment with the EU's governance-first model toward a state-led, development-priority framework. Analysing legislative and policy evolution between 2017 and 2025, the study attributes this shift to structural incompatibilities with the EU's governance logic, intense technological competition, containment pressures from the US, and pressing domestic economic challenges. Rather than merely converging with the US market-driven model, China has crafted a hybrid "state-led and controlled development" model that integrates selective deregulation and innovation incentives with socialist core values and national security imperatives. Furthermore, the article reveals how this model operates through dual mechanisms: proactive state steering of resource allocation in prioritised AI sectors, and restrictive controls on content, algorithms, and applications deemed threatening to social stability or party leadership. The research contributes to comparative AI governance literature by conceptualising China's adaptive, authoritarian agility and analysing the tensions it creates between innovation promotion and political control. Besides, it reflects on

the implications of the deregulatory trends in the EU, the US, and China for global norm-setting. It identifies fragile yet meaningful common ground—shared ethical principles such as safety, fairness, and human well-being—that may offer a foundation for minimal cooperative engagement amid deepening geopolitical fragmentation. Finally, this study concludes that China's AI governance leverages technological development to achieve strategic autonomy, while raising questions about the sustainability of its dual emphasis on accelerated growth and entrenched political control.

Session 5

The Sound of Silencing: Identities and Ideologies in Commercial Text-To-Speech

Alice Ross \\ University of Edinburgh \\ United Kingdom
Nina Markl \\ University of Essex \\ Colchester \\ United Kingdom
Catherine Lai \\ University of Edinburgh \\ United Kingdom
Lauren Hall-Lew \\ University of Edinburgh \\ United Kingdom

KEYWORDS
voice user interfaces; speech synthesis; large speech models; voice AI; natural language processing; diversity and inclusion

ABSTRACT
Contemporary text-to-speech (TTS, also called Voice AI) technology allows the synthesis of speech that is frequently described as highly 'natural' and, in some contexts, indistinguishable from human speech. Voice interfaces using such synthesised speech have been adopted in a wide range of contexts. Recognising that listeners are likely to hear human-like voices as belonging to different demographic/social groups, and that these social judgments exist within ideological frameworks, we design and carry out a novel experiment that draws upon sociolinguistic theory to investigate these issues in a leading commercial TTS system.

We used the TTS v2 model and natural language prompting 'Voice Design' system offered by ElevenLabs, a popular commercial platform with over 5 million registered users as of 2025, to produce a set of synthesised voices given prompts that contain no demographic information. The prompts combine the template 'a voice that sounds [adjective]' with a balanced set of adjectives selected from existing language attitudes literature to convey positive personality traits associated either with status (competent, confident, educated, intelligent, professional) or solidarity and social desirability (considerate, friendly, kind, polite, warm). Analysing evaluations of these voices by 60 listeners (US and Canadian English speakers), we find the set lacking in diversity overall: 93% of the generated voices were predominantly perceived as white, 86.6% perceived as men's voices, and 60% perceived as US-accented.

We conclude that this TTS model is disproportionately likely to reproduce white, male, US- or UK-accented speech when prompted to convey competence and other positive traits. Further, the output reveals apparently systematic associations between certain adjectives and demographic groups: 'kind' with (white, American and British) women; 'educated' and 'polite' with white British men. We argue that this skewed representation of (synthetic) speakers and their traits is not an isolated or neutral quirk. It arises from the collection, labelling and use of data in the system, and it exists in the context of an era characterised by increasing dissemination of dangerous white supremacist, anti-migrant, and misogynist ideas.

Our findings can be understood in terms of 'overrepresentation' of white, male, US- and UK-accented speech, or as 'erasure' of all the voices that are not reproduced: considering the space of human(-like) voices, a subset of unnaturally fluent, Mainstream American-accented, white, and predominantly male voices are becoming normalised. If the use of TTS systems like this one continues to increase, particularly in education and entertainment, it seems inevitable that the overall set of voices an individual is exposed to will become homogenised in future. In addition to Voice AI systems' destructive impact on voice acting and voiceover work, this new normal would influence the cognitive development and language attitudes of future generations, particularly towards already marginalised varieties and disfluent speakers. Bias in natural language processing and other machine learning models is not under-researched, but work on the topic can be deepened and strengthened with analyses informed by social awareness of how biases and stereotypes contribute to material inequities and harms.

Session 5

Beyond the Principle-Practice Gap: How Digital AI Ethics Tools Fall Short in the Age of GenAI

Lisa Koeritz \\ Universität Tübingen \\ Germany
Sonja Pfisterer \\ Universität Tübingen \\ Germany

KEYWORDS
AI ethics; operationalization; tools

ABSTRACT
AI Ethics Tools have traditionally been positioned as instruments aiming to translate abstract ethical principles into actionable guidance during AI development, closing the principle-to-practice gap. Yet, as Generative AI (GenAI) has changed the landscape of ethical AI development, little systematic research has examined what new requirements arise for those tools when situated in a GenAI dominated landscape. Based on a content analysis informed by grounded theory of publicly available interactive digital AI ethics tools (open-source repositories, commercial platforms, academic toolkits) (N=30) designed to directly help

practitioners translate abstract ethical principles into practical recommendations during AI development and published between 2018 and 2024, we examine how these tools approach ethical integration and interaction with GenAI-specific challenges. We identify and discuss four interrelated challenges:

First, we find that the tools regularly impose predefined metrics (e.g., fairness scores), rigid evaluation frameworks, and prescriptive outputs (e.g., "good/bad" labels, recommendations), thereby constituting a form of epistemic power: tools shape what counts as "ethical," "fair," or "responsible" by embedding implicit assumptions about values, risks, and expertise.

Second, our analysis also reveals that many tools were designed prior to widespread GenAI adoption, assuming a relatively static, component-based AI system, making them ill-suited to handle the inherent training data opacity of GenAI-based systems. Tools that depend on inaccessible training data cannot substantiate their ethical claims, rendering those unreliable in practice.

Third, GenAI models import assumptions from their training data and providers, which are then reproduced through user interfaces or perpetuated by being embedded in systems, often without critical reflection. As a result, they inherit and propagate risks and assumptions without the possibility of contestation. Tools assuming access to or relying on information unavailable in GenAI systems, normalize provider-centric knowledge production.

Fourth, GenAI's unpredictability (hallucinations, prompt dependency) intensifies accountability demands, yet tools ignore this. Instead of addressing why ethical failures occur, they often provide uncontextualized output, implicitly demanding users' technical, ethical, and critical literacy to map guidance to their own context.

Fundamentally, we find current AI-Ethics-Tools do not adequately address GenAI-specific challenges. With these findings we argue that the traditional tooling approach to the principle–practice gap is no longer sufficient in a GenAI era: With GenAI shaping the AI development landscape, the challenges of integrating ethical principles in technology development with and of AI systems itself have changed and complexity has increased even further.

Our paper hence argues for a rethinking of tool design, not only to include GenAI-specific concerns (e.g., hallucination, provenance, prompt dependency), but to explicitly reflect the very own epistemic assumptions and power dynamics integrated into digital AI ethics tools themselves. In a context of increasingly proprietary and complex GenAI systems, they otherwise risk reproducing and deepening existing power imbalances between technology providers and diverse stakeholders in knowledge production.

Session 5

Epistemic Authority in the Governance of Generative AI's Societal Harms: Discourses of Participation and Systemic Risk

Julian A. Morgan \\ Humboldt-Universität zu Berlin \\ Germany
Vladimir Apraxine \\ KU Leuven \\ Belgium

ABSTRACT
Our paper centers on the question of the regulation of the societal harms of generative AI (GenAI) and departs from the traditional conception of AI harms as caused by defects in AI systems or their use for specific high-risk purposes. Instead, it focuses on how the growing adoption of GenAI leads to "societal harms" (Smuha, 2021) that apply pressures on fundamental fabrics of society, such as the erosion of public trust, political polarization, and the exacerbation of socio-economic inequalities (Taeihagh, 2025).

Starting from an understanding of the governance of societal harms as contingent upon "risk determinations" (Beck, 1992), where the allocation of the epistemic authority to make these determinations can result from technical factors as well as legal-institutional structures (Pouillot, 2021; Bodó, Weigl & Araujo, 2025), and in line with a reading of GenAI as a "socio-technical system" (Hedfeld, 2026), our paper aims to situate the relationship between legal frameworks, policy discourse, and the distribution of epistemic authority in GenAI governance. In this light, it argues that the regulation of the societal harms of GenAI should depart from technocratic compliance mechanisms and foster a participatory approach that redistributes epistemic authority to a wider network of stakeholders (Noorman & Swierstra, 2023; Kaminski & Malgieri, 2024). In particular, citizens and civil society organizations (CSOs) should be considered vital actors in the identification and mitigation of such harms, as their expertise and experience are essential in the determination of the contextual elements of the societal impacts of GenAI (Pi & Proctor, 2025).

The paper starts by conducting a legal analysis of AI regulation in the EU, which reveals structural limitations for the empowerment of citizens' and CSOs' epistemic authority in the governance of GenAI's societal harms. Whether at the level of redress mechanisms, CSO inclusion, or systemic risk mitigation, the European legal framework has been critiqued for allowing the centralization of epistemic authority over the governance of societal harms of GenAI to AI firms themselves (Lazaro Cabrera & Maier, 2026; Wachter, 2024). These limitations reveal that although the AI Act formally acknowledges the existence of GenAI's societal harms and the need for participatory governance in furthering public values, it tends to consolidate the primacy of corporate epistemic frameworks (Kieslich, Helberger & Diakopoulos, 2026; Griffin, 2024).

To better understand how participatory approaches and public values are subsumed by corporate epistemic authority, the paper further draws on a (critical) discourse analysis (Howarth, 2010; Mouffe & Laclau, 2014) of recent EU policy documents to examine how GenAI governance is narrativized and articulated through institutional practices, regulatory discourses, ethical frameworks, and ideological framings. Indeed, despite rhetorically gesturing towards public values, democratic governance, fundamental rights, and a broad understanding of harm, European legal frameworks have been shown to be increasingly (co)determined by AI firms at all stages of the law-making and oversight processes (Khanal, Zhang & Taeihagh, 2025). The analysis identifies three dominant and interrelated strands of discourse (formations) that provide the structure for the rearticulation of public values with corporate interests: A "digital sovereignty" discourse (Grohmann & Costa Barbosa, 2026), a "trustworthy AI" discourse (Laux, Wachter, & Mittelstadt, 2022; Stamboliev & Christiaens, 2025), and a discourse of innovation and "regulatory simplification" (Larsson, Hildén & Söderlund, 2026). The contextual analysis of these three strands reveals how the problematization of GenAI's societal harms in EU policy discourse is increasingly structured by the re-articulation of public values with (antagonistic) corporate imperatives and interests. Across these strands, the paper offers a descriptive account of how this allows for the reallocation of epistemic authority away from public institutions and normative deliberation toward corporate and technical expertise. These dynamics raise critical questions about democratic participation, accountability, and the future role of legal institutions and frameworks in governing complex sociotechnical systems. By foregrounding the epistemic authority over governance of GenAI's societal harms and the discursive process of its distribution as central analytical objects, the paper contributes to broader debates on participative governance and the realization of public values in the algorithmic society.

REFERENCES

Beck, U. (1992). Risk Society. SAGE.

Bodó, B., Weigl, L., & Araujo, T. (2025). Governance by trust mediators in the digital society: The redistribution of risk and vulnerability. Journal of Trust Research, 16(1), 6-30. https://doi.org/10.1080/21515581.2025.2571505

Griffin, R. (2024). Codes of Conduct in EU Digital Regulation and AI Policy: The Potential and Risks of Soft Law Tools. In K. Prifti, E. Demir, J. Krämer, K. Heine, & E. Stamhuis (Eds.), Digital Governance (Information Technology and Law Series, Vol. 39). T.M.C. Asser Press, The Hague. https://doi.org/10.1007/978-94-6265-639-0_12

Grohmann, R., & Costa Barbosa, A. (2026). Sovereignty-as-a-service: How big tech companies co-opt and redefine digital sovereignty. Media, Culture & Society, 48(2), 416–424. https://doi.org/10.1177/01634437251395003

Hedfeld, P. (2026). AI as a socio-technical actor: Rethinking definitions for ethics and governance. AI and Ethics, 6(2), 254. https://doi.org/10.1007/s43681-026-01123-1

Howarth, D. (2010). Power, Discourse, and Policy: Articulating a Hegemony Approach to Critical Policy Studies. Critical Policy Studies, 3(3–4), 309–335. https://doi.org/10.1080/19460171003619725

Kaminski, M., & Malgieri, G. (2025). Impacted Stakeholder Participation in AI and Data Governance. Yale Journal of Law & Technology, 27, 247. https://scholar.law.colorado.edu/faculty-articles/1757

Khanal, S., Zhang, H., & Taeihagh, A. (2025). Why and how is the power of Big Tech increasing in the policy process? The case of generative AI. Policy and Society, 44(1), 52–70. https://doi.org/10.1093/polsoc/puae012

Kieslich, K., Helberger, N., & Diakopoulos, N. (2026). Scenario-based sociotechnical envisioning (SSE): An approach to enhance systemic risk assessments. AI and Ethics, 6(3), 255. https://doi.org/10.1007/s43681-026-01084-5

Larsson, S., Hildén, J., & Söderlund, K. (2026). Implications of regulating a moving target: Between fixity and flexibility in the EU AI Act. Law, Innovation and Technology, 0(0), 1–24. https://doi.org/10.1080/17579961.2026.2633682

Laux, J., Wachter, S., & Mittelstadt, B. (2022). Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and the Acceptability of Risk Social Science Research Network. https://doi.org/10.2139/ssrn.4230294

Lazaro Cabrera, L., & Maier, M. (2026). Potential Avenues for Redress for AI-related Harms: A Visual Explanation. Center for Democracy & Technology. https://cdt.org/insights/a-visual-explanation-of-potential-avenues-for-redress-for-ai-related-harms-under-the-ai-act/

Mouffe, C., & Laclau, E. (2014). Hegemony and Socialist Strategy: Towards a Radical Democratic Politics (3rd ed.). Verso.

Noorman, M., & Swierstra, T. (2023). Democratizing AI from a Sociotechnical Perspective. Minds & Machines, 33, 563–586. https://doi.org/10.1007/s11023-023-09651-z

Pi, Y., & Proctor, M. (2025). Toward empowering AI governance with redress mechanisms. Cambridge Forum on AI: Law and Governance, 1, e24. https://doi.org/10.1017/cfl.2025.9

Pouliot, V. (2021). Global governance in the age of epistemic authority. International Theory, 13(1), 144–156. https://doi.org/10.1017/S1752971920000433

Smuha, N. (2021). Beyond the individual: governing AI's societal harm. Internet Policy Review, 10(3). https://doi.org/10.14763/2021.3.1574

Stamboliev, E., and Christiaens, T. (2025). How Empty Is Trustworthy AI? A Discourse Analysis of the Ethics Guidelines of Trustworthy AI. Critical Policy Studies, 19, 39. https://doi.org/10.1080/19460171.2024.2315431

Taeihagh, A. (2025). Governance of generative AI. Policy and Society, 44(1), 1–10. https://doi.org/10.1093/polsoc/puaf001

Wachter, S. (2024). Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond. Social Science Research Network https://doi.org/10.2139/ssrn.4924553

Generative AI and Society:
What is at Stake?

Session 6: Theorizing GenAI in Knowledge Production

Session 6

Making Choices Rational – The Unquestioned Premises on Decision-Making in Modern AI Systems

Uli Meyer \\ Johannes Kepler Universität Linz \\ Austria
René Werner \\ Johannes Kepler Universität Linz \\ Austria

KEYWORDS
organizational decision-making; artificial intelligence; organizational theory; performativity

ABSTRACT
In our talk, we address a central question to understand current AI models such as Generative AI: What kind of unquestioned premises are ingrained in the currently spreading modern AI systems? Our central claim is that there exists an elective affinity between decision-making models in the fields of Artificial Intelligence (AI) and Organizational Theory. Based on Herbert Simon's work on both AI systems and organizational decision-making we examine how models of decision-making are based on very specific societal notions about what constitutes decisions. We examine the properties of these societal ideas and identify six key characteristics, emphasizing rational calculations based on a logic of consequences. These specific notions of decision-making converge again in the phenomenon of AI-based algorithmic decision-making in organizations, as we demonstrate using examples from descriptions and advertisements of such systems, the current literature on their use, and empirical research concerning organizational practices.

And while current AI-systems including Generative AI are based on a sub-symbolic paradigm and Simon's work was focused on symbolic AI systems, we claim that these identified notions of decision-making still persist in contemporary AI. This is highly relevant to the question of expertise, responsibility and epistemic power raised in the Call for Papers, since we argue that these unquestioned premises also filter what kind of information are deemed to be relevant and essential to inform decision-making processes. Our talk therefore addresses the discussion on Generative AI by studying the performative effect of these unquestioned premises on decision-making on the design and use of contemporary AI systems.

Session 6

Recursive Escalation: How Generative AI Increases Digital Society's Demand for Reflexivity

Carsten Ochs \\ University of Kassel \\ Germany
Jonathan Kropf \\ University of Kassel \\ Germany
Markus Uhlmann \\ University of Kassel \\ Germany
Klara-Aylin Wenten \\ University of Kassel \\ Germany

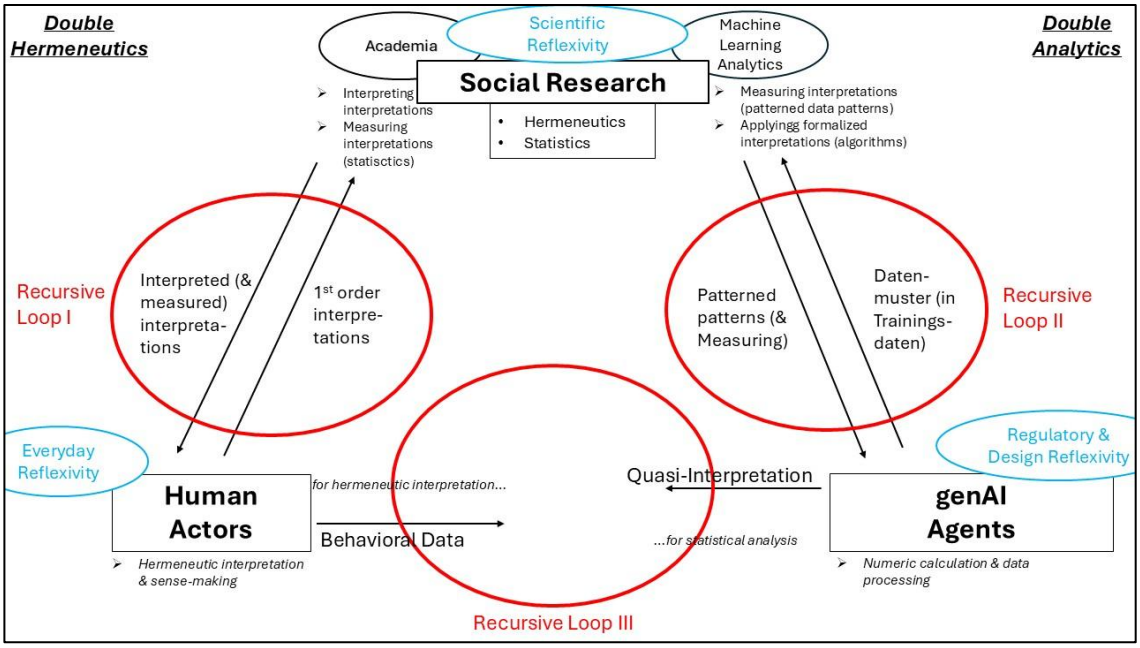
KEYWORDS
AI and meaning-making; sociodigital recursion; changes in knowledge production; reflexivity

ABSTRACT
Sociology and its object, "society," have always stood in a reflexive relationship to each other (Garfinkel 1984; Giddens 1984; Schütz 2010). According to Giddens (1995) actors, in their hermeneutic, everyday practices, operate with interpretations of the social that are in turn interpreted by sociology (Schütz' "second-order constructions"). The permeability of the practice / social analysis boundary enables the reinsertion of scientifically interpreted interpretations into everyday interpretive practices. This gives rise to a recursive loop between practice and social science that Giddens (1995: 338) called "double hermeneutics": interpretations and interpreted interpretations potentially enter into a relationship of mutual stabilization and/or amplification.

Taking double hermeneutics as a point of departure, our contribution advances the thesis that, first through digitization and, more importantly, through the integration of interaction-capable generative AI (GenAI) into meaning-based everyday practices, (a) a multiplication of recursive loops occurs and, consequently, (b) societal demands for reflexivity increase. As regards digitization, Marres (2017: 129) has argued that digital infrastructures simultaneously exhibit a social (as facilitators of sociality) and epistemic character (as data analysts). In digital infrastructures, social science is conducted beyond academia; here, double hermeneutics emerge in the form of the inscription of social-scientific models (e.g., Merton's insights into citation practices) into the functioning of infrastructures (search algorithms), which in turn feeds back into practice (ibid.: 67).

With the integration of machine learning into digital infrastructures, a relationship of mutual stabilization and amplification beyond hermeneutics arises. Statistical patterns identified in training data guide system operations that participate in the enactment of everyday practices, insofar as these patterns exert structuring effects in the form of probabilistic predictions ("patterned patterns") (Pasquinelli 2017; Mühlhoff 2021; Schulz-Schaeffer 2025). E.g., behavioral patterns structure users' news selection ("newsfeeds"), which in turn affects their usage behavior and, consequently, the data generated about that behavior. We refer to this recursive loop between data collection and AI-based pattern analysis as "double analytics."

With genAI, the recursive loops of double hermeneutics and double analytics are brought into a novel relationship: the double analytics of machine learning underlies the operating mode of technical agents that introduce quasi-interpretations into interactions with humans in an autonomous manner (Schulz-Schaeffer 2025: 9); although the statistical generation of these quasi-interpretations remains largely "devoid of meaning" (Burrell 2016). Nevertheless, quasi-interpretations are hermeneutically processed by human interaction partners and thus seep into the sphere of double hermeneutics ("artificial communication," Esposito 2022; "data reflectivity/reflexivity," Mahnke et al. 2024; "artificial sociality," Depounti/Natale 2025), thereby potentially triggering recursive escalations (e.g., genAI's quasi-interpretations generated with statistical bias are translated into human sexist interpretations, which subsequently serve as data material for further analytics). The relating of "double analytics–double hermeneutics" thus constitutes another recursive loop:



When it comes to double hermeneutics "recursion" means that the realm of practices (i.e. interpretation) comprises precisely those notions (i.e. interpreted interpretations) that sociological interpretation helped to create, resulting in recursive stabilization (e.g., sociological interpretation helps to create the concept of sovereignty, the latter informs political practices and is in turn to be identified by sociological analysis etc.). Recursive escalation induced by genAI, on the other hand, may take three different forms: a.) discrimination, b.) sycophancy or c.) hallucinatory takeover. In all three cases there is a mutual reinforcement of data patterns identified in training data, their translation into operational quasi rules (i.e. the weights that shape how gen AI agents' operations are performed), and the re-entering of operations' outcome into the social realm. This may result in discriminatory AI based decision making, in the enforcement of pathological behavior, or in disinformation. Our presentation will ultimately close by pointing out the increased need for reflexivity in the realm of everyday practices (Lamla 2021) as well as in terms of regulation (Beckers/Teubner 2024) and data analysis (Ho et al. 2025).

Session 6

Generative AI and Society: The Emergence of a New Media Regime

Nadja Schaetz \\ University of Hamburg \\ Germany
**Emilija Gagrčin ** University of Bremen \\ Germany

KEYWORDS
media; democracy; technology; knowledge production; epistemic authority; reflexivity

ABSTRACT
Ongoing public and legal disputes surrounding generative AI point to a deeper transformation in political communication. Rather than reflecting a temporary regulatory gap, these conflicts signal a constitutive moment (Starr, 2004) marked by a misalignment between established institutional norms and emerging communicative practices. This misalignment unsettles foundational assumptions, as established conceptualisations of creativity, originality, integrity, and legitimacy of knowledge no longer map neatly onto new forms of mediated production. What is at stake is not only how generative AI should be governed, but also who counts as a meaningful participant in public communication and which forms of knowledge are afforded credibility and authority.

To make sense of this moment, we draw on Delli Carpini and Williams' (2011) concept of media regimes, understood as relatively stable configurations of actors, institutions, processes, and norms that shape communicative practices in democratic life. Media regime change occurs during constitutive moments, when economic, political, cultural, or technological shifts open the field for contestation over ownership, regulation, and legitimacy.

Building on this framework, this paper theorizes the emergence of a new and qualitatively distinct historical arrangement: the generative media regime. Unlike previous regimes, which were defined by the mediation, remixing, or circulation of information, this regime is characterized by the introduction of synthetic production. By acting as "media within media" (Hepp et al., 2023), generative AI transforms not only the technological infrastructures of political communication but also its institutional and epistemic foundations. Drawing on the notion of hybrid figurations (Hepp et al., 2023, pp. 50–51) and scholarship that conceptualizes GenAI-based media as communicative participants rather than mere intermediaries (Guzman & Lewis, 2020), we situate generative AI within evolving configurations of human and machine communication.

Responding to concerns that recent scholarship on generative AI and democracy often treats AI as historically unprecedented and analytically isolated from media and communication research (Karppinen et al., 2025), the paper offers a corrective. Extending the media regimes framework through literature on knowledge orders (Neuberger et al., 2023; Bartsch et al., 2025), we conceptualize generative AI as part of a longer trajectory of mediated democracy.

In doing so, the paper illustrates how the dual shaping of generative AI and society is reconfiguring established relationships among news media, technology, politics, and citizenship, thereby unsettling the foundations of epistemic judgment and authority in contemporary democracies.

Generative AI and Society:
What is at Stake?

Session 7: Generative AI in Science

Session 7

Scientific Rigour at Stake: The Effects of GenAI on HCI Research

Lea Stöter \\ University of Kassel \\ Germany
Konstantin Lackner \\ University of Kassel \\ Germany

KEYWORDS
generative AI; discursive malleability; HCI; interaction paradigms; research practices

ABSTRACT
With the recent surge of Generative AI (GenAI) use in scientific research, concerns have emerged regarding the robustness of peer-review processes, the submission and publication of pseudoscientific work, and the increasing influence of industry-driven narratives at the expense of scientific rigour (Ahmed et al. 2023, Andrews et al. 2024, Messeri and Crockett 2024).

We argue that this may in part be due to the discursive malleability of the technology (Siuda et al. 2023) brought on by the hype-fueled narrative and lack of transparency coming out of the AI industry (e.g. Narayanan and Kapoor 2024) as well as the systems' opaque architecture. Similar to how machine learning research was rendered vulnerable to pseudoscientific approaches through the hype surrounding their technology (Andrews et al. 2024), industry's functional over-promise on GenAI and the anthropomorphised design of the tools is doing the same now. The language used in scientific research on and including GenAI can be seen as site through which this malleability is enacted.

Drawing on Brock's (2018) Critical Technocultural Discourse Analysis, we explore the connections between the intent-based outcome specification paradigm (Nielsen 2023) of GenAI tools and the language used in research studies on or including the models. The corpus is organised in two parts: First, we investigate the industry narrative surrounding GenAI tools using the corporate websites of the five most-commonly used chatbot applications (ChatGPT by Open AI, Claude by Anthropic, Gemini by Google, Grok by xAI, and Copilot by Microsoft). The listed chatbots additionally make up the focus of the discursive interface analysis.

Second, we analyse research articles in the Proceedings of the CHI Conference on Human Factors in Computer Systems published between 2023 and 2025. We scan the abstracts of research articles from sessions mentioning AI in their title for GenAI or LLM usage before further examining the full article.

Investigating both the technological artifact – in this case the interface – and the communication about the artifact allows us to make connections between the semiotic and material, in order to demonstrate the interplay between form, function and belief (Brock 2018) shaping the discursive malleability of GenAI.

Finally, we conclude that these dynamics have significant implications for the integrity of scientific knowledge production in HCI as well as academic research as a whole, particularly given the disproportionate influence of industry actors in shaping contemporary technological discourses.

Session 7

Beyond Tool Use: The WISE Framework for Researchers' Generative AI Literacy in the Age of Digital Intelligence

Wenjuan Gao \\ Beihang University \\ Beijing \\ China

KEYWORDS
genAI literacy; researchers' competencies; WISE framework; higher education; research practice

ABSTRACT
The rapid advancement of Generative Artificial Intelligence (GenAI) is profoundly reshaping higher education across teaching, learning, research, and academic governance. As one of the earliest and most deeply affected communities, researchers have increasingly adopted conversational AI tools for literature review, code generation, academic writing, and research design. While these practices substantially enhance the efficiency of knowledge production, they also introduce emerging challenges related to academic integrity, cognitive dependency, and ethical responsibility. Notably, leading universities worldwide have experienced a marked shift in their stance toward GenAI. For example, several institutions within the UK Russell Group—including the University of Oxford, the University of Cambridge, and Imperial College London—implemented comprehensive bans on ChatGPT in March 2023 due to concerns over academic misconduct, but subsequently issued joint guidelines in July 2023 permitting the responsible use of GenAI tools under conditions that safeguard academic integrity, educational equity, and the development of AI literacy. This transition from restriction to regulation suggests that the central issue surrounding GenAI is no longer whether it should be used, but how it can be used rationally, ethically, and effectively.

Against this backdrop, this study addresses a critical question in the era of digital intelligence: what capabilities become essential for researchers in an academic environment characterized by the widespread use of GenAI? Drawing on a qualitative research design, the study conducts a systematic text analysis of GenAI usage guidelines from QS Top 100 universities, complemented by in-depth interviews with postgraduate students (master's and doctoral) and faculties. Based on the integrated analysis, the study conceptualizes graduate researchers' GenAI literacy as comprising four core competencies—the ability to wield,

integrate, scrutinize, and envision GenAI in research practice—and synthesizes these competencies into an integrative WISE framework.

Within this framework, the ability to wield refers to researchers' capacity to select GenAI tools aligned with specific learning and research needs, design effective prompts for human–AI interaction, and critically optimize and integrate AI-generated outputs to enable efficient human–machine collaboration. The ability to integrate emphasizes not only an understanding of GenAI's fundamental principles, technical paradigms, language generation logic, and developmental trajectories, but also the capacity to embed these technological capabilities within disciplinary knowledge systems, thereby expanding disciplinary boundaries and reconstructing personal knowledge frameworks. The ability to scrutinize highlights the importance of maintaining critical judgment toward AI-generated content, including evaluating its accuracy, bias, logical coherence, and risks of homogenization, while clearly delineating boundaries of use, upholding academic ethics, safeguarding human agency, and avoiding overreliance on GenAI. Finally, the ability to envision foregrounds a future-oriented and leadership-driven perspective, whereby researchers leverage GenAI to drive research innovation, transcend traditional disciplinary boundaries, address complex real-world problems, adapt to rapid technological iteration, and ultimately transition from passive users to proactive leaders of AI-enabled research.

By articulating the WISE framework as an ability-oriented model of GenAI literacy, this study contributes a structured conceptualization of the competencies required of graduate researchers in the digital intelligence era, and offers a reflective lens on the evolving relationship between artificial intelligence and research practice.

Session 7

Automated Governance in Science? The Impact of Generative AI on Epistemic Authority and Responsibility

Linda Nierling \\ Karlsruhe Institute of Technology \\ Germany
Angelina Sophie Dähms \\ Karlsruhe Institute of Technology \\ Germany
Dana Mahr \\ Karlsruhe Institute of Technology \\ Germany

KEYWORDS
science governance; responsible AI; qualitative expert study

ABSTRACT
Digital transformation processes are not only prevalent in society but are also occurring and affecting the science system. The ongoing debate within science assesses the use, impact and consequences of generative AI (genAI) for knowledge production (Messeri & Crocket 2024), implications for the overall system (Fecher et al. 2023, Watermeyer et al. 2024) as well as how work practices in science are affected (Parisi & Sutton 2024, Kenney & Lincoln 2025). The stakes of these developments are manifold with regard to efficacy, epistemic changes as well as power relations, but also affect the redistribution of epistemic authority and responsibility. Thus, this contribution focuses on perspectives, strategies and actions actors from the science system have and analyses how the current transformation of the science system is perceived and shaped by them.

It draws on the project "Leibniz WissenschaftsCampus – Digital Transformation of Research" (DiTraRe). It presents empirical findings from a qualitative expert interview study with ten international science governance actors (including funding bodies, publishing entities, and research associations) conducted from April to June 2025. Results show that actors of the science system have to deal with challenges on three levels. At the level of individual scientists, e.g. increased efficacy for writing and idea generation in science, up- and deskilling processes in academic work leads to modified modes of knowledge creation. On an organizational level, actors have to develop strategies to cope with specific risks of genAI, e.g. epistemic risks or risks of misuse. Lastly all actors have to deal with developing guiding principles which cover different disciplines, research traditions while lacking regulatory power on the one hand and granularity for practical use on the other.

Based on the empirical findings, the contribution argues that a responsible AI use in science needs a context-specific approach, which takes different roles of actors of the science system into account. Here, "automated governance" describes the partial delegation of evaluative, decision-making or steering processes (such as peer review, quality assessment or funding decisions) to AI-based technical infrastructures. It highlights the specific role actors can play in the ongoing digital transformation in science. Funding bodies have to provide overarching

frameworks serving as orientation for scientists from different disciplines. Publishing entities have to deal with the increasing number of publications and the need to develop human-in the loop approaches for peer-review and quality assurance practices. Finally, research associations need to keep their research sovereignty, while guiding researchers by guidelines and tools.

The paper argues that emerging forms of "automated governance" risk shifting core scientific negotiation processes (such as peer review, quality assessment or policy advice) from social and institutional arenas to technical infrastructures, thereby fundamentally challenging established notions of epistemic authority and responsibility in science.

Session 7

Open Science Practices and Epistemic Diversity in the Age of Artificial Intelligence

Angelie Kraft \\ Weizenbaum Institute \\ Berlin \\ Germany
Jochen Knaus \\ Weizenbaum Institute \\ Berlin \\ Germany
Sonja Schimmler \\ Weizenbaum Institute \\ Berlin \\ Germany

KEYWORDS
epistemic diversity; epistemic marginalization; open science; generative AI

ABSTRACT
Artificial intelligence (AI), especially large language models (LLMs), are changing the playing field for Open Science. Hosseini et al. (2025) map out how extractive data practices for the sourcing of AI training data, as well as the uptake of AI systems to access scientific knowledge are challenging the existing ideals of Open Science by transforming previously open knowledge into implicit and closed model knowledge. Moreover, while LLM-based chatbots and agents may provide simplified access to complex scientific content, they are also prone to factual inaccuracy, bias, sycophancy and "hallucinations". Generated summaries often do not accurately reproduce original content and fabricate false (cf., Magesh et al.). This renders AI systems an unreliable source of knowledge for researchers and can yield negative effects for researchers who remain un- or falsely attributed, which may discourage them from making their scientific output openly available.

Our work extends on this discussion by referencing a body of research that suggests that LLMs may contribute to an exacerbation of hermeneutic injustice, i.e., an exclusion of marginalized communities from the collective pooling of knowledge (Kraft & Soulier, 2024; Fricker, 2007). For instance, LLMs obscure sources in languages that are commonly underrepresented in mainstream language technology (Kay et al., 2024) and they amplify the marginalization of non-dominant knowledges (as measured along societal dimensions such

as gender, ethnicity, geographic location, etc.) by reproducing "information that is statistically dominant because it is statistically dominant" (Mollema, 2025, p. 6).

On the one hand, Open Science can function as a remedy to these concerns: the properly documented sharing of datasets, models, and methods enable transparency and reproducibility (Solaiman, 2023; Gebru et al., 2021, Mitchell et al., 2019). These are important conditions for more responsible AI use as they allow for, e.g., rigorous evaluation and informed use. On the other hand, scholars like Leonelli (2022) point out that established Open Science practices are not equipped to deal with questions of epistemic diversity. She argues that universalist conceptualizations of Open Science fail to account for the diversity of scientific practices and the local conditions of scientific communities. For instance, to fully leverage openly available data utilizing LLMs might require certain human and computational resources that are not available to the same degree in every setting. Hence, not all scientific communities benefit equally from the sharing of data, which ultimately discourages scientists located in lower-resource settings from participating. We argue that modern LLMs as well as the aggressive data practices of corporations behind the technology amplify these inequalities further. The surge of AI once more challenges universalist takes on Open Science and feeds into a paradox where openness might simultaneously help and harm.

Our work builds on the ongoing debate regarding the ways in which AI counteracts the achievements of Open Science and connects it to concerns about epistemic diversity and marginalization. It furthermore provides a brief outlook regarding the consequences we may draw from these tensions, and directions for best practice recommendations we may formulate to support researchers to navigate in this new environment.

REFERENCES

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92.

Fricker, M. (2007). Epistemic Injustice: Power and the Ethics of Knowing. Oxford University Press.

Hosseini, M., Horbach, S. P. J. M., Holmes, K., & Ross-Hellauer, T. (2025). Open Science at the generative AI turn: An exploratory analysis of challenges and opportunities. Quantitative Science Studies, 6, 22–45. https://doi.org/10.1162/qss_a_00337

Kay, J., Kasirzadeh, A., & Mohamed, S. (2024). Epistemic Injustice in Generative AI. In Proc. AIES 2024, 7, 684–697. https://doi.org/10.1609/aies.v7i1.31671

Kraft, A. and Soulier, E. (2024). Knowledge-Enhanced Language Models Are Not Bias-Proof: Situated Knowledge and Epistemic Injustice in AI. In Proc. FAccT 2024, 1433–1445. https://doi.org/10.1145/3630106.3658981

Leonelli, S. (2022). Open Science and Epistemic Diversity: Friends or Foes? Philosophy of Science, 89(5), 991–1001. https://doi.org/10.1017/psa.2022.45

Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., and Ho, D.E. (2025). Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools. Journal of Empirical Legal Studies, 22 (2): 216–242. https://doi.org/10.1111/jels.12413

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019). Model cards for model reporting. In Proc. FAT* 2019, 220–229. https://doi.org/10.1145/3287560.3287596

Mollema, W. J. T. (2025). A Taxonomy of Epistemic Injustice in the Context of AI and the Case for Generative Hermeneutical Erasure. AI Ethics, 5, 5535–5555. https://doi.org/10.1007/s43681-025-00801-w

Solaiman, I. (2023). The gradient of generative AI release: Methods and considerations. In Proc. FAccT 2023, 111–122. https://doi.org/10.1145/3593013.3593981

Session 7

ICLR vs. LLMs: Investigating the Role of Generative AI in the Peer Review Process of ICLR 2026

Juni Schindler \\ University of Zurich \\ Switzerland

KEYWORDS
generative AI; peer review; ICLR 2026; epistemic power

ABSTRACT
The peer-review process for the 19,500 papers submitted to this year's International Conference on Learning Representations (ICLR 2026) faced unprecedented challenges: a leak in the conference's peer review system OpenReview exposing reviewer identities coincided with a surge of peer reviews generated by large language models (LLMs), with estimates suggesting 21% were fully LLM-generated [1]. While LLMs were permitted for light editing of reviews at ICLR 2026, fully LLM-generated reviews containing false claims or hallucinated references were considered code of ethics violations [2]. The conference organisers promised a crackdown on LLM-generated reviews, but the controversy left many researchers frustrated and questioning the future of peer review.

There is, of course, a degree of irony in this story, but it also brings to the fore epistemic ruptures that threaten peer review as an important pillar of academic knowledge production. How does knowledge production change if AI is increasingly replacing the epistemic gatekeeping of traditional peer review? What will be the role of human researchers in an ever-accelerating feedback loop between LLM-generated code, manuscripts, and now also peer reviews? We need to address these questions to prevent a monopoly of epistemic power in

the hands of the few multi-billion-dollar tech companies running the AI infrastructure behind these transformations.

In this contribution, I investigate the role of generative AI, specifically LLMs, in academic knowledge production using the peer review process at ICLR 2026 as a case study. Analysing over 75,000 reviews retrieved via the OpenReview API, I compare content, linguistic markers, and reviewer scores between LLM-generated and human-written reviews. The quantitative findings are enriched through qualitative interviews detailing the experiences of conference participants with generative AI in their dual roles as authors and reviewers.

By interpreting the results through a critical lens informed by software studies and human-computer interactions, I respond to the challenges identified above and reflect on pathways toward accountability in AI-powered review processes [3]. The insights gained about the epistemic power of generative AI in this contribution extend beyond academia, offering crucial insights for sustaining democratic access to knowledge production amidst the rise of AI technologies.

REFERENCES

[1] Naddaf, M. Major AI conference flooded with peer reviews written fully by AI. Nature (2025). https://www.doi.org/10.1038/d41586-025-03506-6

[2] ICLR 2026 Program Chairs. ICLR 2026 Response to LLM-Generated Papers and Reviews – ICLR Blog. https://blog.iclr.cc/2025/11/19/iclr-2026-response-to-llm-generated-papers-and-reviews/ (2025).

[3] Klumbytė, G., Piehl, H. & Draude, C. Towards Feminist Intersectional XAI: From Explainability to Response-Ability. Workshop: Human-Centered Explainable AI (HCXAI), Conference on Human Factors in Computing Systems CHI '23 (2023).

Generative AI and Society:
What is at Stake?

Session 8: Generative AI in Creative Work

Session 8

From DIY to AI: Independent Musicians, Artificial Intelligence, and the Reconfiguration of Cultural Work in Brazil

Sebastian Piraces \\ DigiLabour \\ Toronto \\ Canada

KEYWORDS
cultural work; independent music; artificial intelligence; generative AI; platformization; global south; Brazil

ABSTRACT
Independent music scenes are historically intertwined with do-it-yourself (DIY) practices as strategies of resistance and/or construction of alternatives to the dynamics of the dominant cultural industries and their concentration of power (Hesmondhalgh 2013), employing their creativity not only towards artistic work but also developing and experimenting alternate modes of production, distribution, and sustainability (Oliveira 2023; Bennet 2018) whether it be for ideological motives (as opposition to the cultural industries) or the possibility to insert oneself into the cultural market/scene (for economic sustainability). Over the last two decades, however, these practices have increasingly been shaped by digital platforms, streaming services, and algorithmic systems that reorganize visibility, value, and creative autonomy (Jones 2021; Oliver 2024). The recent and accelerated proliferation of generative artificial intelligence (GenAI) for public use marks a new phase in this transformation, which leads us to question if and how new technology impacts independent musicians' careers. While GenAI promises increased productivity (McKinsey 2023), its adoption also raises concerns regarding new forms of dependency on Big Tech platforms, intensified data extraction, and the reconfiguration/precarization of creative labor. This paper investigates how independent musicians in Brazil—operating as solo entrepreneurs or small collectives—are incorporating GenAI into their creative and organizational routines, situating these practices within a longer historical process of digital transformation in the music industry. It examines how AI tools are integrated into artistic creation, communication and promotional strategies, while navigating streaming platforms that govern visibility, monetization, and legitimacy. The Brazilian context offers a relevant Global South perspective, where digital cultural labor unfolds within historical asymmetrical economic precarity, and patterns of dependency rooted in the coloniality of power (Quijano 2000).

Methodologically, this study follows a neo-materialist qualitative framework (Lemos 2020) and conducts interviews with contemporary independent musicians active in the Brazilian musical ecosystem as groundwork material, which is then analyzed through a critical theory framework engaging debates on surveillance capitalism (Zuboff 2019) and the transformation of cultural work (Nath 2025).

Through the interviews, the paper argues that GenAI simultaneously enables a form of "democratization" of cultural production—by lowering costs, accelerating workflows, and expanding access to creative tools—while deepening structural dependencies on platforms. Rather than simply augmenting creative autonomy, the integration of AI into creative work contributes to the intensification of self-management, the externalization of risks and the normalization of precarious cultural labor (Fisher 2011), blurring the lines between empowerment and exploitation and producing ambivalent subjectivities among independent artists while also incorporating new career-related responsibilities (Frenneaux 2023).

By framing GenAI as a new chapter in the longer histories of cultural and labor control rather than framing it as a radical rupture, this paper positions AI as a continuation—and intensification—of platform-based governance over creative work. It aims to contribute to current debates on cultural workers and GenAI, in a historical and Majority World perspective.

Session 8

All Remains the Same While Everything Changes?! Generative AI, Creativity, and Professional Identity in Advertising Agencies

Angela Graf \\ Bavarian Institute for Digital Transformation | bidt \\ Munich \\ Germany
Niina Zuber \\ Bavarian Institute for Digital Transformation | bidt \\ Munich \\ Germany

KEYWORDS
doing creativity; generative AI; practice theory; professional identity; advertising

ABSTRACT
Generative AI (GenAI) has rapidly entered creative industries, promising to inspire, accelerate, and partially automate creative work. At the same time, it has sparked concerns about the future of human creativity, the potential devaluation of creative labor, and the transformation of professional roles. These questions are particularly salient in the advertising industry, where creativity constitutes both a central business asset and a core element of professional identity. Despite the intensity of public debate, however, empirical insights into the impact of GenAI on creative work are limited. Against this background, we trace the implementation of GenAI across different stages of the creative process and analyze how this use shapes creative professionals' identity and their understanding of creativity.

Theoretically, we combine philosophical and sociological perspectives. From a philosophical viewpoint, we understand creativity as a human capacity closely tied to authorship and authenticity. Only by highlighting this connection, ideas or concepts such as novelty and innovation, ranging from recombination within existing conventions to more radical forms of creative rule-breaking, can be distinguished (Abel, 2006). From a sociological perspective,

and drawing on Bourdieusian practice theory, creativity is conceptualized as a socially embedded professional practice (Bourdieu, 1977, 1990, 1993; Krämer, 2014). In creative agencies, "doing creativity" unfolds collectively within a specific field, shaped by shared norms, evaluative criteria, and a professional habitus. This dual perspective allows us to analyze GenAI not only in terms of performance or output, but also in relation to underlying beliefs, dispositions, and symbolic boundaries.

Empirically, we draw on a single case study in a German creative advertising agency, employing a mixed-methods design, including time- and task-tracking data on GenAI usage as well as in-depth semi-structured interviews with strategists, copywriters, art directors, and creative directors (N=20).

Our analysis reveals that GenAI is used pervasively throughout the entire creative process and has become a regular and integral part of the workflow. It is applied in the context of desk research, ideation, conceptual development, and idea execution. Thereby, workflows and practices sometimes change fundamentally. However, simultaneously, the underlying self-conceptions, evaluative frameworks, and symbolic boundaries of creative professionalism remain largely preserved. Rather than being perceived as a disruptive force or competitor, generative AI is predominantly regarded as a tool for accelerating routines, broadening possibilities, and enhancing productivity, efficiency, and quantity without replacing the human core of creativity. Instead of threatening the creative professionals' identity, it reinforces and sharpens the contours of what is deemed 'truly creative', leading to a reassurance of the value of human creativity. Creatives continue to claim authorship, authenticity, and legitimate judgment as uniquely human faculties that they do not ascribe to machines.

REFERENCES

Abel, G. (Ed.). (2006). Kreativität. XX. Deutscher Kongreß für Philosophie. Felix Meiner. https://doi.org/10.28937/978-3-7873-2001-1

Bourdieu, P. (1977). Outline of a Theory of Practice. Cambridge University Press. Bourdieu, P. (1990). The Logic of Practice. Stanford University Press.

Bourdieu,P. (1993). European perspectives. The field of cultural production: Essays on art and literature (R. Johnson, Ed.). Columbia University Press.

Krämer, H. (2014). Die Praxis der Kreativität: Eine Ethnografie kreativer Arbeit. Transcript.

Session 8

Adapting to Generative AI in Creative Work: A Technological Frames Perspective on Creative Advertising

Georg von Richthofen \\ Humboldt Institute for Internet and Society \\ Berlin \\ Germany
Sonja Köhne \\ Humboldt Institute for Internet and Society \\ Berlin \\ Germany
Maja Golf-Papez \\ University of Sussex \\ Brighton \\ United Kingdom

KEYWORDS
creative work; creative advertising; generative AI; technological frames; netnography

ABSTRACT
Generative artificial intelligence (GenAI) is expected to transform creative work. Drawing on the concept of technological frames, three years of netnographic research in online advertising communities, and interviews with advertising creatives, we examine how creatives' interpretations of GenAI shape their responses. Our study identifies three interconnected domains through which creatives interpret GenAI—creative processes, jobs, and outputs—each characterized by distinct technological frames. We distinguish three adaptation strategies that creatives consider in response to emerging changes in the advertising industry: upskilling, reskilling, and deep skilling. Finally, we link technological frames to adaptation strategies and show how contextual career factors shape both perceived impact and adaptive response. The study offers implications for theory and practice, advancing our understanding of how creatives interpret and adapt to new technologies like GenAI.

Session 8

Win-win Exploitation and Creative Labor

Annette Zimmermann \\ University of Wisconsin-Madison \\ United States

KEYWORDS
generative AI art and creativity; philosophical theories of exploitation; exploitative optimization; epistemic exploitation; AI and labor

ABSTRACT
Contemporary discourse on the ethics of AI-powered art and creativity focuses on demonstrable harm: technological deskilling, job displacement, authorship controversies. These concerns share a common structure: identifying clear winners and losers where groups suffer concrete disadvantages.

However, this harm-centric framework cannot capture subtler ethical challenges in ostensibly mutually beneficial 'win-win' arrangements. Consider FN Meka, the AI-generated rapper created using voice synthesis trained on human vocal data. While the source artist received compensation, this exemplifies wrongful exploitation despite mutual benefit. His creative labor—vocal patterns, stylistic elements, expressive nuances—generated millions for technology companies and record labels, while he received a fraction of that value. Cases like this involve unfair advantage-taking as understood in the philosophical literature on exploitation. A popular, albeit not uncontroversial, philosophical position defines exploitation as one agent's disproportionate extraction of value from another's productive (or creative) labor. This problem has interesting, important implications beyond individual cases, thereby helping us adopt a more nuanced conceptualization of AI's structural societal impact, where Big Tech companies leverage concentrated wealth, power, and information resources to secure disproportionately favorable terms with creators. Companies may point to 'win-win' benefit distributions to justify such arrangements. This concentration simultaneously results from exploitation, and could enable further exploitative arrangements, possibly creating self-reinforcing inequality cycles if left unchecked.

In this paper, I build on and extend the sophisticated yet underexplored theoretical resources that contemporary analytic political philosophy can offer for analyzing the subtle problem of wrongful 'win-win exploitation' cases. Recognizing exploitation as a distinct ethical category provides crucial analytical tools for evaluating AI art's moral landscape, revealing that win-win scenarios may nonetheless involve (at least pro tanto) ethical wrongs requiring normative scrutiny beyond simple harm-based metrics. My argument does not imply that AI art and creativity are necessarily ethically doomed. It does imply that they require more nuanced, and possibly surprising, guardrails.

Generative AI and Society:
What is at Stake?

Session 9: Evaluating GenAI Knowledge Production

Session 9

AI Safety as a Space Between Fields

Anna Thieser \\ Columbia University \\ New York City \\ United States
Jack LaViolette \\ Columbia University \\ New York City \\ United States
Gil Eyal \\ Columbia University \\ New York City \\ United States

KEYWORDS
AI safety; professions; expertise; networks; fields

ABSTRACT
This paper examines the novel task of making artificial intelligences safe. What is the task of AI safety, and what is AI safety expertise? We pursue three interrelated studies. First, we analyze a curated sample of technical AI safety experiments, examining how they make speculative risks about rogue AI tangible through behavioral signatures, microcosms, existence proofs, and stress tests. Second, we map collaboration networks within a purposive corpus of AI safety publications, revealing how individual researchers with hybrid credentials stitch together academic and for-profit institutional worlds. Third, we examine job advertisements for AI safety positions to understand how organizations present themselves and describe candidates amid fundamental indeterminacies about their work. Rather than treating AI safety's contradictions and ambiguities as signs of immaturity or deception, we argue they constitute productive features of a space between fields – a trading zone of collaboration and capital exchange between universities, AI firms, non-profits and governments.

Session 9

The Knowledge Valued by AI Companies: An Analysis of Benchmarks Used to Advertise GenAI Models

Stefan Baack \\ Independent Researcher, Der Spiegel, Stanford University
Christo Buschek \\ Der Spiegel
Maty Bohacek \\ Stanford University \\ United States

KEYWORDS
GenAI benchmarking; knowledge production; AGI narratives; evaluation cultures

ABSTRACT
Benchmarks used to measure GenAI model performance have been widely criticized as unreliable indicators for real-world user experience (Weidinger et al. 2025), or even as fundamentally flawed (Raji et al. 2021). However, AI companies continue to highlight results on benchmarks prominently when they announce new models. The benchmarks included in those release announcements are carefully curated to advertise the model to prospective users and to distinguish it within an increasingly crowded market. They therefore encode the priorities and values of the respective AI companies: what roles and expectations they attach to their models, and what they optimized for. This not only shapes knowledge about models, but also expectations about their capability to support various kinds of knowledge production, especially since benchmarks framed to measure "knowledge and reasoning" are among the most prominently highlighted in these release announcements. A critical analysis of benchmarks used to advertise model capabilities is thus crucial for understanding how GenAI is transforming the way knowledge is produced.

We gathered 231 benchmarks highlighted across 138 model release announcements in 2025 from 11 major AI model builders (primarily the US and China). Because AI companies vary greatly in how they frame benchmarks, we developed a unified taxonomy based on what benchmark authors claim to measure, facilitating cross-company analysis. In addition, we studied the most popular benchmarks in-depth. Our analysis shows that very few benchmarks are used frequently, the majority (63%) are used only by a single model builder and almost half (41%) appear in only one release announcement. Taken together, benchmarks dedicated to coding and math are highlighted most often. However, the most frequently highlighted individual benchmark, GPQA Diamond, is from the second most popular benchmark category: “General knowledge application” benchmarks that evaluate performance on a broad range of subjects (while often also including math and coding).

Due to their popularity and the ambiguity of what they claim to measure, we analyzed the five most popular “General knowledge application” benchmarks in-depth. Despite ostensibly measuring general capabilities, they focus predominantly on science subjects (math,

physics, etc.) and share a vagueness about the object they seek to measure, as neither "knowledge" nor "reasoning" are explicitly defined. This vagueness appears to be at least partially due to the influence of Artificial General Intelligence (AGI) narratives. Two of the most popular general knowledge and reasoning benchmarks cite AGI literature and seek to measure progress towards particular "levels of AGI," and two others, including GPQA Diamond, are clearly informed by AGI narratives without explicitly citing AGI literature.

We suggest that benchmarks in model release announcements of major AI companies function as narrative devices meant to shape what is known about models according to the priorities of these companies, not as standardized measurements enabling cross-model comparison. Moreover, they indicate the type of knowledge AI companies seek to promote. The prominence of AGI narratives in popular General knowledge application benchmarks suggests a privileging of knowledge that fits this narrative: knowledge necessary to replace human labor, i.e. knowledge that is economically valuable.

REFERENCES

Raji, Inioluwa Deborah, Emily M. Bender, Amandalynne Paullada, Emily Denton, and Alex Hanna. 2021. “AI and the Everything in the Whole Wide World Benchmark.” arXiv:2111.15366. Preprint, arXiv, November 26. https://doi.org/10.48550/arXiv.2111.15366

Weidinger, Laura, Inioluwa Deborah Raji, Hanna Wallach, et al. 2025. “Toward an Evaluation Science for Generative AI Systems.” arXiv:2503.05336. Preprint, arXiv, March 13. https://doi.org/10.48550/arXiv.2503.05336

Session 9

From Statistical Property to Cognitive Capability: Testing Practices and Generalization Claims in Language Models

Susanne Förster \\ University of Siegen \\ Germany

KEYWORDS

benchmarking; testing infrastructures; epistemic power

ABSTRACT

Contemporary large language models are attributed with far-reaching cognitive capabilities and described as generalist systems with extensive world knowledge, able to potentially solve any prompt-based task (OpenAI 2023). This development rests on the establishment of the deep learning architectures in the 2010s, most prominently on the release of the Transformer (Vaswani et al. 2017) and its subsequent inclusion in all major model releases. These models are marketed as fundamentally open or foundational (Bommasani et al. 2021), yet

simultaneously characterized by their increasing opacity, resulting from a mathematical complexity caused by ever-larger training datasets and parameter counts (Burrell 2016) and by actively withholding architectural and training details (OpenAI 2023). The combination of proclaimed openness and opacity has transformed benchmark tests, which were previously primarily tools for quantification and evaluation (Raji et al. 2021), into key instruments through which claims about general capabilities of generative AI models are authorized.

Drawing on a close reading of key architectural papers and technical reports and building on STS and Critical Data and Algorithm Studies (Amoore et al., 2024, Campolo and Schwerzmann 2023), this paper traces the genealogy of contemporary capability claims through a sequence of deep learning models from Word2Vec (Mikolov et al. 2013) to GPT-3 (Brown et al. 2020). It examines how models, evaluation practices and interpretive frameworks co-evolved during a formative period roughly between 2013 and 2018, analyzing how benchmark performance across multiple tasks was gradually stabilized as evidence of general understanding and how this stabilization reconstituted what counts as evidence in AI research more broadly.

Rooted in classical statistics and foundational to machine and transfer learning, generalization originally described the capacity of a model to recognize learned patterns in unseen but structurally similar data. For nearly 40 years, evaluation practices shaped by the Common Task Framework (Liberman 2010, Donoho 2017) and popularized through the Netflix Prize (Bennett and Lanning 2007) and ImageNet challenge (Deng et al. 2009) framed evaluation as training and testing on narrowly defined tasks. Generalization was taken as a statistical property of models that were trained for one application, such as image recognition or recommender systems. This changed with models such as the Transformer, when systems began to be tested on more than one task, leading to a shift in defining generalization as general (language) understanding. The paper looks into how this transformation emerged and unfolded from early multi-task evaluations on related tasks to testing practices using a wide range of benchmarks to explore and communicate seemingly inherent, cognitive-like abilities of the models. High accuracy in these tests confers objectivity, authority and economic attractiveness to the models. In reconstructing this shift, the paper provides a critical account of how contemporary large-scale generative models acquired their current identity as general, capable and potentially dangerous systems.

REFERENCES

Bennett J, Lanning S (2007) The Netflix Prize. Proceedings of the KDD Cup Workshop 2007, Association for Computing Machinery, New York, 3–6.

Amoore L, Campolo A, Jacobsen B, Rella L (2024) A world model: On the political logics of generative AI. Political Geography 113:103134. https://doi.org/10.1016/j.polgeo.2024.103134

Bommasani R, Hudson DA, Adeli E, et al (2021) On the Opportunities and Risks of Foundation Models. http://arxiv.org/abs/2108.07258

Brown TB, Mann B, Ryder N, et al (2020) Language Models are Few-Shot Learners. http://arxiv.org/abs/2005.14165

Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society 3. https://doi.org/10.1177/2053951715622512

Campolo A, Schwerzmann K (2023) From rules to examples: Machine learning's type of authority. Big Data & Society 10. https://doi.org/10.1177/20539517231188725

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: A Large-Scale Hierarchical Image Database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. Institute of Electrical and Electronics Engineers, Piscataway, NJ, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Donoho D (2017) 50 Years of Data Science. Journal of Computational and Graphical Statistics 26:745–766. https://doi.org/10.1080/10618600.2017.1384734

Liberman M (2010) Obituary: Fred Jelinek. Computational Linguistics 36:595–599. https://doi.org/10.1162/coli_a_00032

Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient Estimation of Word Representations in Vector Space. http://arxiv.org/abs/1301.3781

OpenAI (2023) GPT-4 Technical Report. http://arxiv.org/abs/2303.08774

Raji ID, Bender EM, Paullada A, et al (2021) AI and the Everything in the Whole Wide World Benchmark. In: Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2). New York, NY, p 23

Vaswani A, Shazeer N, Parmar N, et al. (2017) Attention is all you need. arXiv:1706.03762 [cs]. Available at http://arxiv.org/abs/ 1706.03762

Session 9

Bye Bye Perspective API: Lessons for Measurement Infrastructure in NLP, CSS and LLM Evaluation

David Hartmann \\ Manuel Tonneau \\ Angelie Kraft \\ LK Seiling \\ Dimitri Staufer \\ Pieter Delobelle \\ Jan Fillies \\ Anna Ricarda Luther \\ Jan Batzner \\ Mareike Lisker

KEYWORDS
toxic language detection; evaluation infrastructure; epistemic dependency; reproducibility crisis; proprietary AI systems

ABSTRACT
The closure of Perspective API at the end of 2026 discards what has functioned as the de facto standard for automated toxicity measurement in NLP, CSS, and LLM evaluation research. We document the structural dependence that the communities built on this single proprietary tool and discuss how this dependence caused epistemic problems that have affected - and will likely continue to affect - collective research efforts. Perspective's model was periodically updated without versioning or disclosure, its annotation structure reflected a single corporate operationalisation of a contested concept, and its scores were used simultaneously as an evaluation target and an evaluation standard. Its closure leaves behind non-updatable benchmarks, irreproducible results, and ultimately a field at risk of perpetuating these issues by turning to closed-source LLMs. We use Perspective's announced termination as an opportunity to call for an independent, valid, adaptable, and reproducible toxicity and hate speech measurement infrastructure, with the technical and governance requirements outlined in this paper.