

**Document Version**

Final published version

**Licence**

CC BY

**Citation (APA)**

Schrills, T., Kahr, P., Langer, M., Kaur, H., & Gadiraju, U. (2026). AI CHAOS! 1st Workshop on the Challenges for Human Oversight of AI Systems. In *Companion Proceedings of the 31st International Conference on Intelligent User Interfaces, IUI 2026 Companion* (pp. 237-239). (International Conference on Intelligent User Interfaces, Proceedings IUI). Association for Computing Machinery (ACM). <https://doi.org/10.1145/3742414.3794953>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.  
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

**Sharing and reuse**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

# AI CHAOS! 1st Workshop on the Challenges for Human Oversight of AI Systems

Tim Schrills  
Institute for Human-Centered  
Interactive Systems  
University of Lübeck  
Lübeck, Germany  
tim.schrills@uni-luebeck.de

Patricia Kahr  
Department of Informatics  
University of Zurich  
Zurich, Switzerland  
patricia.kahr@uzh.ch

Markus Langer  
Department of Psychology  
University of Freiburg  
Freiburg im Breisgau, Germany  
markus.langer@psychologie.uni-  
freiburg.de

Harmanpreet Kaur  
University of Minnesota  
Minneapolis, Minnesota, USA  
harmank@umn.edu

Ujwal Gadiraju  
Web Information Systems  
Delft University of Technology  
Delft, Netherlands  
u.k.gadiraju@tudelft.nl

## Abstract

As AI systems are increasingly adopted in high-stakes domains such as healthcare, autonomous driving, and criminal justice, their failures may threaten human safety and rights. *Human oversight* of AI systems is therefore critically important, as a potential safeguard to prevent harmful consequences in high-risk AI applications. Although regulations like the European AI Act mandate human oversight for high-risk AI, we lack methodologies and conceptual clarity to implement it effectively. Independent of policy and regulation, poorly designed oversight can create dangerous illusions of safety while obscuring accountability. This interdisciplinary workshop aims to bring together researchers from various disciplines, including AI, HCI, psychology, law, and policy, to address this critical gap. We will explore the following questions – How can we design AI systems that enable meaningful human oversight? What methods effectively communicate system states and risks to human overseers? How do we ensure scalable and effective interventions? Through papers, talks, and interactive group discussions, participants will identify oversight challenges, examine stakeholder roles, discuss supporting tools, methods, regulatory frameworks, and establish a collaborative research agenda. Our central goal is to further a roadmap that enables effective human oversight for the responsible deployment of AI in society.

## CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Information systems**; • **Computing methodologies** → *Artificial intelligence*; • **Applied computing** → *Law, social and behavioral sciences*;

## Keywords

human oversight of AI, explainable AI, human-AI interaction, human-centered AI, risk, safety

## ACM Reference Format:

Tim Schrills, Patricia Kahr, Markus Langer, Harmanpreet Kaur, and Ujwal Gadiraju. 2026. AI CHAOS! 1st Workshop on the Challenges for Human Oversight of AI Systems. In *Companion Proceedings of the 31st International Conference on Intelligent User Interfaces (IUI Companion '26)*, March 23–26, 2026, Paphos, Cyprus. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3742414.3794953>

## 1 Workshop Overview and Relevance

In recent years, we have witnessed the increasing deployment of AI-powered applications in high-stakes domains, such as medical diagnosis and treatment recommendations, autonomous driving systems, predictive policing and recidivism assessment, and automated hiring processes [1, 6]. Failures in these contexts can directly threaten human safety and fundamental human rights. The opacity and complexity of many AI systems further compound these risks. Stakeholders often cannot understand, intervene in, or oversee automated decisions that significantly impact human lives (i.e., through monitoring or detecting failures or harmful outcomes).

Echoing this growing call for *human oversight*, emerging ethical guidelines and legislation (such as the European AI Act [5]) around the world are mandating human oversight of AI-powered systems in high-risk contexts. Human oversight can be broadly defined as a multi-stakeholder overseeing of an AI-based system throughout its lifecycle [10]. Although human oversight of AI systems is a potentially effective strategy to mitigate the risks of AI bias and errors, it demands that humans oversee and monitor AI systems effectively – recognizing critical moments, and execute timely interventions at scale [9]. Once AI regulations, demanding human oversight, are in force, system designers and developers must at least address two critical questions related to intelligent user interfaces. (1) *How can we design interfaces that effectively communicate AI reasoning and outputs to human overseers?*, and (2) *How can we design interfaces that enable overseers to control, monitor, and intervene in AI-based systems effectively?*



This work is licensed under a Creative Commons Attribution 4.0 International License.  
*IUI Companion '26, Paphos, Cyprus*  
© 2026 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-1985-1/26/03  
<https://doi.org/10.1145/3742414.3794953>

Currently, the discourse on human oversight lacks conceptual clarity and methodologies to understand and implement effective human oversight in AI-powered systems [10]. Poorly implemented oversight mechanisms can create a false sense of safety, leading stakeholders to believe that systems are properly monitored when they are not [2–4], and can obscure accountability and delay the identification of system problems [7, 8]. Therefore, effective human oversight requires not only technical innovation but also a deep understanding of human behavior and the affordances of socio-technical systems. At a recent Dagstuhl seminar\*, researchers urged interdisciplinary discussions in addressing these fundamental challenges. Building on this, our workshop aims to bridge the communities and create interdisciplinary collaboration among researchers across fields such as artificial intelligence, formal methods, human-computer interaction, psychology, law, and policy.

The workshop will solicit papers, host talks, and bring together workshop attendees for group discussions to share research and perspectives, and develop a future research agenda for effective human oversight of AI systems. Our focus will include, but not be limited to, the following topics:

- Identifying current challenges in human oversight of AI systems
- Understanding the role of various stakeholders in human oversight
- Developing and aligning theoretical models of human oversight and frameworks for its evaluation
- Developing methods, tools, processes, and resources to support stakeholders in overseeing AI systems
- Discussing regulatory measures, public policies, or industry standards around human oversight
- Sharing perspectives about human oversight, identifying and addressing open challenges
- Identifying the user interface-related opportunities and challenges
- Exploring the tradeoff between effective human oversight for risk mitigation and challenges of human autonomy, development, and self-efficacy

## 2 Workshop Organizers

The following workshop organizers bring together complementary disciplinary expertise, represent different academic institutions, and are at varying stages of their careers. This diversity ensures a broad set of perspectives on human oversight of AI systems.

*Dr. Tim Schrills* is a postdoctoral researcher at the University of Lübeck. He studies Human-Centered AI with a focus on how human and artificial information processing can be integrated. His experimental work investigates how automation levels and explanations affect human oversight and control in AI-enabled decision support. His research has appeared in venues such as TiiS, Behaviour & Information Technology, and CHI, and he co-organized the STEP-HAI workshop at IUI 2025.

*Dr. Patricia Kahr* is a postdoctoral researcher at the University of Zurich. She explores how people collaborate with AI systems, aiming to shape human-AI interactions beyond efficiency by emphasizing human agency, control, and meaningful work, thereby

enabling appropriate and beneficial AI use. Her work, contributing to HAI practices and methods, has been published in IUI, CHI, and TiiS. She also co-organized the STEP-HAI workshop at IUI 2025.

*Dr. Markus Langer* is a full professor of Work and Organizational Psychology at the University of Freiburg. His interdisciplinary research at the intersection of psychology, computer science, and law focuses on human oversight, explainable AI, and human-centered design of AI-based systems. He co-organized the Dagstuhl seminar on “Challenges of Human Oversight” in 2025 and the AIsoLA track on “Responsible and Trusted AI”.

*Dr. Harmanpreet Kaur* is an Assistant Professor at the University of Minnesota. She works on human-centered AI, explainability, and hybrid intelligence systems, applying these to domains such as exploratory data analysis, workplace well-being, and knowledge search. Her research combines the critical evaluation of existing systems with the design of new, human-centered approaches. She has organized events on these topics at CHI, CSCW, HCOMP, FAccT, and KDD.

*Dr. Ujwal Gadiraju* is an Associate Professor at Delft University of Technology. His research advances human-centered AI through computational techniques and systems that enhance human experiences, align with human values, and promote inclusivity and appropriate reliance. He has extensive community experience, co-leading workshops and tutorials on HCI and AI at CHI, CSCW, HCOMP, UMAP, IUI, and related venues.

## 3 Workshop Activities

Our workshop structure is aimed at building community, exchanging knowledge, and co-creating new research agendas for this emerging topic. To do so, we plan for the following activities throughout our **full-day workshop**:

*Lightning Introduction:* At the start of the workshop, and to serve as an icebreaker, participants respond to a few questions, introducing themselves and their interests—(a) Who are you? (b) What is one fun thing about you that we probably do not know? (c) What would you like to get out of this workshop?

*Invited Talks:* These aim to provide diverse perspectives on human oversight challenges from leading experts in all disciplines. The program will feature one or more invited talks to present a grounding, inspire discussions, encouraging cross-disciplinary dialogue and identifying opportunities for collaboration between technical and human-centered approaches to AI oversight.

*Pecha-Kucha Session & Posters:* Authors of accepted papers can pitch their work through lightning talks, enabling attendees to find them at the subsequent sessions. Authors with accepted papers will present their work in a dedicated interactive poster session. This setting encourages discussions between authors and other attendees. The curated program based on the accepted submissions was organized into three coherent themes—(i) *Psychological Phenomena in Human Oversight*, (ii) *Distributed and Shared Human Oversight*, and (iii) *Governance and Design of Human Oversight*.

*Breakout Sessions:* These serve the goal to enable small-group discussions on specific topics or tasks. We aim to create space for

\*<https://www.dagstuhl.de/en/seminars/seminar-calendar/seminar-details/25272>

more active participation and diverse perspectives, which are then brought back into the main plenary.

*Panel Discussion with Invited Speakers:* Panelists will be asked to present informed by the discussions from the breakout sessions, the panel will revisit critical questions at the intersection of their disciplines and invite further engagement from attendees (e.g., fishbowl-style interaction).

### 3.1 Tangible Outcomes

The workshop information will be publicly available on the website,<sup>†</sup> including the workshop proposal, call for participation, accepted submissions, workshop program, participant information, and other relevant material. We plan to use this website to maintain papers accepted to the workshop, and thematically relevant papers from IUI and other venues. We aim to use this as an evolving repository that grows as we continue to run this workshop in the following years. Shortly after the workshop, we will publish a blog post highlighting the main findings from the breakout sessions.

### 3.2 PC Members

The following individuals acted on the program Committee, reviewing submissions and helping the organizers curate the workshop program: **Tim Miller** (University of Queensland, Australia), **Hanwei Zhang** (University of Saarland, Germany), **Niels van Berkel** (University of Aalborg, Denmark), **Martijn Willemsen** (Eindhoven University of Technology, The Netherlands), **Hanna Hauptmann** (Utrecht University, The Netherlands), **Agathe Balayn** (Microsoft Research, USA), **Margaret Burnett** (Oregon State University, USA), **Wen Duan** (Clemson University, USA), **Anna Maria Feit** (Saarland University, Germany).

### Acknowledgments

We thank Schloss Dagstuhl and the organizers of the Dagstuhl Seminar 25272 for providing the rich foreground from which this workshop emerged. Ujwal Gadiraju was partially supported by the TU Delft AI Initiative, the Model Driven Decisions Lab (*MoDDL*), the Robust LTP *GENIUS* Lab, and the *ProtectMe* Convergence Flagship. Markus Langer was partially funded by DFG grant 389792660 as part of TRR 248 CPEC – Center for Perspicuous Computing and by the Daimler and Benz Foundation as part of the project TITAN – Technologische Intelligenz zur Transformation, Automatisierung und Nutzerorientierung des Justizsystems (grant no. 45-06/24).

### References

- [1] Shreyan Biswas, Ji-Youn Jung, Abhishek Unnam, Kuldeep Yadav, Shreyansh Gupta, and Ujwal Gadiraju. 2024. “Hi. I’m Molly, Your Virtual Interviewer!” Exploring the impact of race and gender in AI-powered virtual interview experiences. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 12. 12–22.
- [2] Michael Chromik, Malin Eiband, Felicitas Buchner, Adrian Krüger, and Andreas Butz. 2021. I think i get your point, AI! the illusion of explanatory depth in explainable AI. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 307–317.
- [3] Ben Green. 2022. The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review* 45 (2022), 105681.
- [4] Johann Laux. 2024. Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act. *AI & society* 39, 6 (2024), 2853–2866.
- [5] Johann Laux, Sandra Wachter, and Brent Mittelstadt. 2024. Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance* 18, 1 (2024), 3–32.
- [6] Siddharth Mehrotra, Ujwal Gadiraju, Eva Bittner, Folkert van Delden, Catholijn M. Jonker, and Myrthe L. Tielman. 2025. “Even explanations will not help in trusting [this] fundamentally biased system”: A Predictive Policing Case-Study. In *Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*. 51–62.
- [7] Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 33–44.
- [8] Inioluwa Deborah Raji, Peggy Xu, Colleen Honigsberg, and Daniel Ho. 2022. Outsider oversight: Designing a third party audit ecosystem for ai governance. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. 557–571.
- [9] Ben Shneiderman. 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction* 36, 6 (2020), 495–504.
- [10] Sarah Sterz, Kevin Baum, Sebastian Biewer, Holger Hermanns, Anne Lauber-Rönsberg, Philip Meinel, and Markus Langer. 2024. On the quest for effectiveness in human oversight: Interdisciplinary perspectives. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 2495–2507.

<sup>†</sup><https://sites.google.com/view/aichaos/iui-2026>