Trust and Transparency in XAI for Workplace Automation: Guiding Industry Decisions on Process Automation

NINA HUBIG, IT:U, Austria
DANIEL KOLB, Leibniz Supercomputing Centre, Germany
ROMEO KIENZLER, IBM Research, Switzerland

As organizations increasingly adopt artificial intelligence for workplace automation, critical questions emerge about *where* and *how* to automate processes responsibly. This half-day workshop addresses the intersection of Explainable AI (XAI), trust, and transparency in workplace automation contexts. We bring together researchers and industry practitioners to explore decision-making frameworks that guide automation adoption while maintaining human oversight, regulatory compliance, and worker dignity. The workshop features presentations of cutting-edge research papers, keynote insights from industry leaders, and interactive sessions where participants collaboratively develop practical frameworks for assessing automation readiness. Expected outcomes include a published proceedings of research contributions, actionable decision frameworks for industry professionals, and establishing a community of practice bridging academic XAI research with real-world workplace automation challenges. This initiative directly addresses CHIWork 2026's theme by examining the friction points and flow opportunities in AI-augmented work.

1 Motivation

The adoption of AI-driven automation in workplaces presents organizations with a paradox: while automation promises efficiency gains and productivity improvements, opaque "black box" AI systems create friction through diminished trust, regulatory challenges, and ethical concerns [2, 3]. The CHIWork 2026 theme of "*Flow – Fun – Friction*" captures this tension perfectly—automation can enhance work *flow* when properly implemented, yet introduces significant *friction* when stakeholders cannot understand, trust, or validate AI decisions [5, 11].

Recent regulatory frameworks, including the EU AI Act, mandate transparency and human oversight for high-risk AI systems in workplace contexts, particularly in recruitment, performance management, and task allocation [6]. Organizations face critical decisions about *which* processes to automate, *how* to maintain appropriate human oversight, and *when* automation may undermine rather than enhance work quality. However, research on XAI techniques and industry practices for automation decision-making remain largely disconnected [1, 7].

Issues to be Addressed: This workshop addresses three critical challenges:

(1) **The Decision Gap:** Industry practitioners lack structured frameworks for determining automation readiness and which processes benefit from AI augmentation versus full automation versus continued human control [2].
(2) **The Trust Deficit:** Workers and managers often distrust AI systems due to opacity in decision-making, lack of recourse mechanisms, and concerns about bias or fairness [10, 13].
(3) **The Implementation Divide:** XAI research advances rapidly in academic contexts, but translation to workplace settings, shaped by legacy systems, diverse stakeholder needs, and regulatory requirements, remains limited [11].

Concrete Objectives:

- **Knowledge Synthesis:** Present and discuss cutting-edge research on XAI, human-in-the-loop (HITL) approaches, bias detection, and trust mechanisms in workplace contexts [8, 12].

Authors' Contact Information: Nina Hubig, IT:U, Linz, Austria, nina.hubig@it-u.at; Daniel Kolb, Leibniz Supercomputing Centre, Munich, Germany, daniel.kolb@lrz.de; Romeo Kienzler, IBM Research, Zürich, Switzerland, romeo.kienzler1@ibm.com.

- **Framework Development:** Collaboratively develop practical decision-making frameworks that help organizations assess which processes are suitable for automation, what level of human oversight is appropriate, and how to implement explainability mechanisms.
- **Community Building:** Establish ongoing connections between academic researchers, industry practitioners, and policy stakeholders to foster continued knowledge exchange and collaborative research. This connects the *what* with the *how*.
- **Research Agenda Setting:** Identify critical gaps in current XAI research from an industry perspective and define priority areas for future HCI work on workplace automation.

Connection to CHIWork 2026 Theme: This workshop directly engages with the conference theme by examining how XAI can reduce *friction* in automated workplace systems, enable better *flow* through appropriate AI augmentation, and potentially restore *fun* to work by automating tedious tasks while preserving meaningful human engagement. By focusing on decision frameworks for *where* to automate, we address the fundamental challenge of navigating opportunities and challenges for the workforce of tomorrow.

2 Workshop Mode

Our workshop will operate as a **in person**, with primary emphasis on in-person participation.

Operational Format: Our workshop employs a mixed format combining formal presentations, interactive discussions, and hands-on collaborative activities. We emphasize *active participation* through small-group exercises, case study analysis, and framework co-creation. This approach aligns with adult learning principles and ensures industry practitioners gain actionable insights alongside exposure to research advancements [9].

Participant Engagement: We anticipate 10-20 participants with diverse backgrounds, such as Academic Researchers, Industry Practictioners, Policy Stakeholders and Graduate Students. Participants will submit position papers, case studies, or research papers (2-4 pages) in advance. Accepted submissions will be compiled into a pre-workshop proceedings distributed to all attendees, enabling productive, grounded discussions.

Accessibility: We commit to making the workshop accessible by providing materials in advance for screen reader compatibility, ensuring visual presentations include alt-text descriptions, offering captions for video content, and accommodating diverse participation needs as they arise.

3 Workshop Activities

The half-day workshop runs from 9:00 AM to 1:00 PM on Monday, June 22, 2026. The schedule balances formal research presentations with interactive, participatory activities designed to bridge research and practice.

4 Call for Participation

Join us for a workshop bridging XAI research and workplace automation practice!

As organizations navigate increasing automation of workplace processes, critical questions emerge: *Which processes should be automated? How do we maintain trust and transparency? What human oversight is appropriate?* This workshop brings together researchers, industry practitioners, policy stakeholders, and students to explore Explainable AI (XAI) as a foundation for responsible workplace automation decisions.

We invite submissions addressing: XAI techniques and methods applicable to workplace contexts; Trust, transparency, and accountability in AI-driven automation; Human-in-the-loop (HITL) approaches and mixed-initiative

Table 1. Proposed Structure of the Workshop.

Time	Activity
09:00 - 09:15	**Welcome & Introduction.** The organizers introduce workshop goals, themes, expected outcomes, and schedule. Participants then briefly introduce themselves (name, affiliation, interest in topic).
09:15 - 10:00	**Keynote: "From Black Boxes to Glass Boxes: Industry Perspectives on XAI Implementation".** An invited speaker from industry discusses real-world challenges and successes implementing XAI in workplace automation. This includes a 15 minute Q&A with workshop participants.
10:00 - 10:30	**Coffee Break & Networking.** Informal networking opportunity and display of posters of selected submissions for asynchronous viewing.
10:30 - 11:30	**Research Paper Presentation.** Presentations of accepted research papers, followed by a moderated discussion synthesizing themes across presentations.
11:30 - 12:15	**Interactive Activity: "The Automation Decision Matrix".** Splitting of participants divided into small mixed groups (researchers + practitioners). Each group receives a realistic workplace automation scenario and collaboratively assesses, e.g., what level of automation is appropriate, what explainability mechanisms are needed, how human oversight should be structured, and what trust and ethical considerations are applicable. Groups use a shared decision matrix framework (provided by organizers) to document their analysis. Digital collaboration tools enable hybrid participation.
12:15 - 12:45	**Case Study Presentations & Framework Discussion.** Groups present their case study analyses, engage in facilitated discussion identifying common patterns, decision criteria, and framework components, and begin synthesizing a generalizable "XAI Automation Readiness Framework" based on collective insights.
12:45 - 01:00	**Synthesis & Next Steps.** The organizers summarize key themes and insights from the day and discuss publication plans for workshop proceedings with the participants. The workshop wraps up by establishing an ongoing community of practice.

systems; Bias detection, fairness mechanisms, and ethical AI frameworks; Case studies of XAI implementation in manufacturing, services, healthcare, or other sectors; Regulatory compliance perspectives (EU AI Act, GDPR, sectoral regulations); Decision-making frameworks for automation adoption; Worker perspectives on AI transparency and trust; Evaluation methods for XAI effectiveness in applied settings.

Submission Types: We welcome position papers, research papers, and industry case studies (2-4 pages, ACM single-column format, excluding references). Submissions will be reviewed by organizers for relevance and quality. All accepted submissions will be included in the workshop proceedings and presented during the workshop.

5 Plans to Publish Workshop Proceedings

We are committed to ensuring wide dissemination of workshop contributions and outcomes. Accepted submissions will be compiled into a **workshop proceedings** published through **CEUR-WS.org** (CEUR Workshop Proceedings), a well-established open-access publication venue for workshop papers indexed by major academic databases including DBLP, Google Scholar, and Scopus [4]. Our envisioned **Publication Timeline** is:

- **Pre-workshop:** Authors of accepted submissions will receive feedback and revision suggestions from organizers. Camera-ready versions will be due one week before the workshop (June 15, 2026).
- **Post-workshop:** Organizers will compile proceedings including workshop overview, all accepted papers, and a synthesis of key insights from the interactive sessions.

- **Publication:** Submit proceedings to CEUR-WS.org by July 15, 2026, with expected publication by August 2026.

Additional Dissemination: We will also make the proceedings available on a dedicated workshop website and distribute through relevant professional networks (ACM SIGCHI community, XAI research mailing lists, industry practitioner networks). Participants interested in extending their workshop contributions into full research papers will be invited to collaborate on a potential special issue proposal to a journal such as *International Journal of Human-Computer Studies* or *AI & Society*, focusing on XAI for workplace automation.

6 Organizers

The workshop will be organized by a diverse team with complementary expertise spanning XAI research, workplace HCI, and industry practice. The organizing committee includes 3 members, who will collectively handle Submission Management, Logistics Coordination, Outreach, Facilitation, Publication and Post-workshop Engagement.

Nina Hubig (Main Contact) is an Assistant Professor of Explainable Artificial Intelligence at the Interdisciplinary Transformation University Austria (IT:U). Her research focuses on trust and transparency in XAI for high stake decisions spanning domains from industry, medicine and geo-environmental domain topics. Using interdisciplinary methods integrating deep learning, network science, and human centric methods she develops transparent and interpretable AI systems that make complex algorithmic decisions understandable in industrial contexts. Her background includes a Ph.D. from the Technical University of Munich and industry experience as a Data Scientist at BMW, where she worked on the process mining team applying machine learning to manufacturing and production optimization.

Daniel Kolb is a postdoctoral researcher at the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities. His research spans human–computer interaction, immersive technologies, and thanatosensitive systems, with a particular focus on Interactive Digital Testimonies and Embodied Conversational Agents. Using interdisciplinary, human-centered methods, he designs and empirically researches systems that preserve and convey interactions and conversations with contemporary witnesses for education and remembrance. His background addresses how users form, calibrate, and sustain trust in AI systems, particularly in socially and ethically sensitive contexts.

Romeo Kienzler works as AI Research Engineer at IBM Research Zurich with a career built at the intersection of scalable systems and machine learning, he is a recognized expert in deep learning, high-performance computing, and open-source ecosystems. He bridges the gap between massive AI models and human intuition. Through projects like TerraTorch and the IBM Geospatial Studio, he transforms “black-box” data into interactive, visual experiences. By integrating embedding space analysis, he gives users a direct look at the AI’s internal logic—making it easy to spot flaws in data, model architecture, or real-world application. For Romeo, this is the core of AI for Math and Science: building the transparency scientists need to turn complex numerical predictions into reliable, human-centered discoveries.

Acknowledgments

We thank the CHIWork 2026 organizing committee for the opportunity to propose this workshop. We also acknowledge the broader XAI research community and industry practitioners whose work motivates and informs this initiative. The authors wish to acknowledge the use of Claude in the writing of this paper. This tool was used to assist with improving the language in the paper. The paper remains an accurate representation of the authors’ underlying work and novel intellectual contributions

References

[1] Adadi, A., and Berrada, M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160.

[2] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (2020), 82–115.

[3] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., et al. Explainable explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (2020), 82–115.

[4] CEUR Workshop Proceedings. CEUR-WS.org: Free open-access publication service. http://ceur-ws.org/

[5] Doshi-Velez, F., and Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* (2017).

[6] European Commission. Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act). *Official Journal of the European Union* L 1689 (2024).

[7] Guidotti, R., Monreale, A., Ruggieri, S., et al. A survey of methods for explaining black box models. *ACM Computing Surveys* 51, 5 (2018), 1–42.

[8] Gunning, D., Stefik, M., Choi, J., et al. XAI—Explainable artificial intelligence. *Science Robotics* 4, 37 (2019).

[9] Kolb, D. A. *Experiential learning: Experience as the source of learning and development.* FT Press, 1984.

[10] Lipton, Z. C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16, 3 (2018), 31–57.

[11] Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38.

[12] Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., et al. Human-in-the-loop machine learning: A state of the art. *Artificial Intelligence Review* 56 (2023), 3005–3054.

[13] Ribeiro, M. T., Singh, S., and Guestrin, C. "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), 1135–1144.

[14] Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1, 5 (2019), 206–215.

[15] Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., and Vertesi, J. Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (2019), 59–68.

[16] Shin, D. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies* 146 (2021), 102551.

[17] Wang, D., Yang, Q., Abdul, A., and Lim, B. Y. Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), 1–15.

[18] Xu, F., Uszkoreit, H., Du, Y., et al. Explainable AI: A brief survey on history, research areas, approaches and challenges. In *Natural Language Processing and Chinese Computing* (2019), Springer, 563–574.

[19] Yang, Q., Steinfeld, A., Rosé, C., and Zimmerman, J. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), 1–13.

[20] Zhang, Y., Liao, Q. V., and Bellamy, R. K. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (2020), 295–305.