Human Agency and Skill in AI-Supported Work: Countering Cognitive Offloading

Caterina Fregosi
University of Milano-Bicocca
Milan, Italy
caterina.fregosi@unimib.it

Simon Fischer
Donders Institute for Brain, Cognition, and Behaviour
Nijmegen, Netherlands
simon.fischer@donders.ru.nl

Chiara Natali
Vrije Universiteit Amsterdam
Amsterdam, Netherlands
c.natali@vu.nl

Hanna Schraffenberger
iHub, Radboud University
Nijmegen, Netherlands
hanna.schraffenberger@ru.nl

Federico Cabitza
University of Milano-Bicocca
Milano, Italy
Digital Health and Wellbeing Center,
Fondazione Bruno Kessler (FBK)
Trento, Italy
federico.cabitza@unimib.it

Abstract

This workshop responds to the observation that, in work contexts, AI systems may promote unintended socio-cognitive consequences, including automation bias, overreliance, reduced critical engagement and sense of agency, and gradual deskilling. It reframes cognitive offloading as a design challenge and proposes a longitudinal perspective on AI-supported work across three time horizons (short, medium, and long term) and three dimensions: performance, agency, and skills. Through structured collaborative activities, participants will identify benefits, risks, protective factors, design strategies, and evaluation measures. The workshop aims to develop a shared vocabulary and a capability-aware research agenda that foregrounds human agency and skill sustainability in AI design.

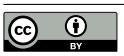
CCS Concepts

• **Human-centered computing** → **HCI theory, concepts and models**; **HCI design and evaluation methods**; *Collaborative and social computing theory, concepts and paradigms.*

Keywords

Human-AI Interaction, Decision Support Systems, Frictional AI

ACM Reference Format:
Caterina Fregosi, Simon Fischer, Chiara Natali, Hanna Schraffenberger, and Federico Cabitza. 2026. Human Agency and Skill in AI-Supported Work: Countering Cognitive Offloading. In *Adjunct Proceedings of the 5th Annual Symposium on Human-Computer Interaction for Work (CHIWORK Adjunct '26), June 22–25, 2026, Linz, Austria.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3805029.3818273



CHIWORK Adjunct '26, Linz, Austria

ACM ISBN 979-8-4007-2598-2/26/06
https://doi.org/10.1145/3805029.3818273

1 Motivation

In work contexts, AI-based decision support systems promise to enhance performance by improving accuracy, reducing workload, and enhancing professional reasoning processes. The dominant narrative on AI-supported work has focused on augmentation: the idea that AI amplifies human capabilities and enables better decisions.

However, growing empirical evidence suggests that AI integration may also produce unintended socio-cognitive consequences. These include automation bias [6, 7], inappropriate reliance [1, 15], reduced critical engagement and agency [8], and—over time deskilling or inhibition of upskilling [12]. When professionals rely repeatedly on algorithmic recommendations, they may progressively offload cognitive tasks that were previously central to their expertise [5]. While such cognitive offloading can increase short-term efficiency, its long-term effects on agency, skill development, professional judgment, and epistemic responsibility remain insufficiently understood [13]. Importantly, these long-term effects are inherently difficult to assess empirically, as they would require longitudinal designs capable of tracking cognitive and professional changes over extended periods of real-world AI use, an approach rarely implemented in current research.

In Human-Computer Interaction literature, most AI systems evaluations still prioritize performance metrics such as accuracy and seamless interaction. Much less attention has been paid to how interaction design choices, such as friction elements, explanation formats, or collaboration protocols, shape users' sense of agency, responsibility, and skill retention.

Recent work in human–AI collaboration has proposed alternative interaction paradigms that promote reflective engagement rather than passive acceptance [3]. These include frictional AI [11], judicial AI [2, 4], dissenting or contrastive explanations [14], evaluative AI [10], and reflection machine [9].

However, these approaches remain fragmented across research communities, and there is no shared framework for evaluating their impact on human agency and skill sustainability.
This workshop addresses cognitive offloading as a design problem.

Rather than asking whether AI improves performance, we ask: Can we design AI systems that preserve or even strengthen human agency and skills while still improving task outcomes?
The aims of this workshop are threefold:

- **Shared Vocabulary** – To articulate a shared conceptual vocabulary around cognitive offloading, agency, deskilling, and skill sustainability in AI-supported work.
- **Design Exploration** – To examine how specific interaction design strategies influence human expertise development, maintenance, and erosion, and how they can preserve or enhance users' sense of agency in AI-supported work.
- **Research Agenda Building** – To identify empirical methods, metrics, and longitudinal study designs capable of assessing both short-term and long-term agency and skills shifts.

Ultimately, our goal is to advance a research agenda that considers human agency and skill as primary design goals, shifting the discourse from performance optimization toward capability-aware design.

2 Workshop Mode

The workshop will be organized as an on-site, half-day event and primary in-person format, allowing for limited hybrid participation of presenters who may be unable to attend in person.

The workshop is designed to be highly interactive and inclusive. All activities will be structured so that participants can engage while seated; Individual reflection and written contributions will be integrated alongside verbal discussions to accommodate different communication preferences and needs. Moreover, three short breaks have been included as part of the workshop programme to lower focus fatigue for all participants.

A detailed agenda will be shared with accepted participants in advance, and we will actively invite feedback to adapt the session to participants' accessibility requirements.

3 Workshop Activities

The activities of this half-day workshop are planned to cover a total of 3.5 hours, with extra leeway to ensure a smooth organisation. Activities will combine short presentations with structured interactive sessions and moments of reflection. All activities will be designed to accommodate different participation styles (verbal and written contributions) and to avoid requiring specific physical abilities.

- **Introduction (20 minutes)**. The organizers will introduce the workshop theme and objectives.
- **Position Paper Presentations (40 minutes)**. Participants will deliver short (5 minute) lightning presentations of their accepted position papers. Contributions will be clustered into thematic groups to guide subsequent discussions.
- **Break (10 minutes)**
- **Interactive Activity, Part 1 – Longitudinal Analysis of AI-Supported Work (60 minutes)** Participants will collaboratively analyze a shared case of AI-supported work (e.g., AI-assisted professional decision-making). They will be divided into three teams, each focusing on a distinct dimension: *Team Performance*, *Team Agency* and *Team Skills*. Each team will examine the case across three temporal horizons: (1) Short-term (0–6 months), (2) Medium-term (1–3 years) and (3) Long-term (5+ years). For each time horizon, teams will identify: *Expected benefits*, *Potential risks*, *Protective factors*, *Design strategies to mitigate negative effects*, and *Possible empirical measures to assess their focal dimension*. Team Performance will focus on accuracy, efficiency, appropriate reliance and error dynamics. Team Agency will examine decision ownership, responsibility, and trust calibration. Team Skills will analyze expertise development, retention, and possible deskilling effects. The workshop organisers moderate and stimulate to the conversation.
- **Break (10 minutes)**
- **Interactive Activity, Part 2 – Presentations (40 minutes)**. Each team will present the outcome of the interactive activities: key risks, protective design strategies, and methodological challenges. The plenary discussion will reflect on potential tensions or alignments among performance optimization, agency preservation, and skill sustainability.
- **Break (5 minutes)**
- **Closing (20 minutes).** The workshop will conclude with a plenary discussion aimed at identifying shared research challenges and opportunities for future collaboration.

4 Call for Participation

When AI becomes a routine collaborator at work, what happens to human expertise and agency? As algorithms suggest diagnoses, draft legal arguments, grade assignments, or review code, professions are not only supported, they are reshaped. Skills are redistributed, judgments recalibrated, and responsibility renegotiated. While short-term performance improvements are frequently documented, much less is known about the long-term effects of AI support on human agency and skill development. Does cognitive offloading free professionals to focus on higher-level reasoning or does it gradually erode core competencies? This workshop invites researchers and practitioners to critically examine the sustainability of human expertise in AI-supported work. We welcome perspectives from HCI, AI, cognitive science, philosophy, organizational studies, healthcare, law, education, and other domains where AI is transforming professional practice. Topics of interest include (but are not limited to): cognitive offloading, responsibility, deskilling and upskilling, human-AI collaboration protocols, longitudinal evaluation methods, and metrics for agency and skill sustainability. Participants are invited to submit a 2–4 page position paper (excluding references) outlining empirical, theoretical, conceptual, or methodological contributions. Proceedings are non-archival: accepted submissions will be made available on the workshop website and on a dedicated Github repository. By bringing together diverse perspectives, the workshop aims to develop a research agenda for human-AI collaboration.

5 Workshop Proceedings

Proceedings are non-archival. Contributions, as well as collaborative workshop outputs, will be made available on the workshop website (https://agency-and-skill.github.io).

6 Organizers

Caterina Fregosi. Caterina Fregosi is a third-year PhD candidate in Informatics at the University of Milano-Bicocca, with a background in Cognitive Sciences. Her research investigates human–AI interaction protocols in high-stakes domains, with a particular focus on explainable AI and its impact on user agency, reliance, and cognitive biases. Her work has received the Best Paper Award at xAI 2024 and has been presented at leading venues (IUI'25, CSCW'25, AAAI'26). During her PhD, she has been a visiting researcher at Technological University Dublin, Eindhoven University of Technology, and NOVA School of Science and Technology in Lisbon.

Simon Fischer. Simon Fischer is a PhD candidate at the Donders Institute for Brain, Cognition, and Behaviour. His research is motivated by legal requirements, such as the European AI Act, which require human oversight and calibration of reliance on decision-support systems. To this end, he investigates the usefulness of data-driven questions during human-AI decision-making to create points of friction and promote cognitive engagement and critical thinking of the decision-maker. He presented his work at leading venues, such as AIES'25, IUI'25, HHAI'25, and FAccTRec'24.

Chiara Natali, PhD. Natali recently obtained a PhD in Computer Science from the University of Milano-Bicocca, Italy. Her research focuses on AI over-reliance and automation bias in high-stakes work settings (clinical diagnosis), proposing friction-in-design strategies to sustain human reasoning and professional skill acquisition. Her work has received Best Paper awards (xAI24; ECSCW 2025) and the Best Doctoral Consortium Paper award (CHItaly 2023). She has organised multiple international workshops and tutorials on agency-supporting and skill-supporting strategies in human-AI interaction and their effects on AI reliance and authority (HHAI'23; INTERACT'23; HHAI'24; HAI'24; HHAI'25; AVI'26). Previously, she was a Visiting Research Fellow at IDSIA (DTI, SUPSI - A.Y. 2024/2025) through the Swiss Government Excellence Scholarship (ESKAS 2024.0002).

Hanna Schraffenberger. Dr. Hanna Schraffenberger is an assistant professor at Radboud University's Digital Security group and is affiliated with Radboud's interdisciplinary research hub on digitalization and society (iHub), where she is co-leading the digital influence group (DIG). Schraffenberger teaches courses about human-centered AI and design and encourages students to consider the societal dimensions of emerging technologies. Her research focuses on digital sovereignty and human autonomy in the digital sphere. She has acted as the editor-in-chief of the AR[t] magazine on Augmented Reality, art, and technology, and various of her projects have been exhibited, e.g., at the InScience film festival and in the Van Gogh Museum. Schraffenberger is a proud member of the IPN EDI (equity, diversity, inclusion) working group, which aims to improve equity, diversity, and inclusion in the ICT landscape.

Prof. Federico Cabitza. Federico Cabitza is an Associate Professor at the University of Milano-Bicocca, where he teaches Human-Computer Interaction and Decision Support Systems. He is the head of the MUDI Laboratory at the University of Milano-Bicocca, and also coordinator of the local node of the national "Informatics and Society" laboratory of the CINI consortium; since 2026, he is also Director of the Digital Health and Wellbeing Center at the Fondazione Bruno Kessler (FBK) of Trento (Italy). Since 2016, he has been collaborating with several hospitals, including the IRCCS Galeazzi Sant'Ambrogio Hospital in Milan (Italy). He is also a founding partner and scientific director of the university spin-off Red Open srl, which specializes in AI impact assessment.

Use of Generative AI Tools

Portions of this manuscript were edited for language clarity and grammar using ChatGPT (OpenAI). The tool was used to improve readability and did not contribute to the scientific content. All authors take full responsibility for the content of the manuscript.

References

[1] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction* 5, CSCW1 (2021), 1–21.

[2] Federico Cabitza, Lorenzo Famiglini, Caterina Fregosi, Samuele Pe, Enea Parimbelli, Giovanni Andrea La Maida, and Enrico Gallazzi. 2025. From oracular to judicial: enhancing clinical decision making through contrasting explanations and a novel interaction protocol. In *Proceedings of the 30th international conference on intelligent user interfaces*. 745–754.

[3] Simon WS Fischer, Hanna Schraffenberger, Serge Thill, and Pim Haselager. 2025. A taxonomy of questions for critical reflection in machine-assisted decision-making. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 8. 940–954.

[4] Caterina Fregosi. 2025. How to Explain in XAI?-Investigating Explanation Protocols in Decision Support Systems. (2025).

[5] Michael Gerlich. 2025. AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies* 15, 1 (2025), 6.

[6] Kate Goddard, Abdul Roudsari, and Jeremy C Wyatt. 2012. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association* 19, 1 (2012), 121–127.

[7] Kate Goddard, Abdul Roudsari, and Jeremy C Wyatt. 2014. Automation bias: empirical results assessing influencing factors. *International journal of medical informatics* 83, 5 (2014), 368–375.

[8] Jiajing Guo, Vikram Mohanty, Jorge H Piazentin Ono, Hongtao Hao, Liang Gou, and Liu Ren. 2024. Investigating interaction modes and user agency in human-llm collaboration for domain-specific data analysis. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–9.

[9] Pim Haselager, Hanna Schraffenberger, Serge Thill, Simon Fischer, Pablo Lanillos, Sebastiaan Van De Groes, and Miranda Van Hooff. 2024. Reflection machines: Supporting effective human oversight over medical decision support systems. *Cambridge Quarterly of Healthcare Ethics* 33, 3 (2024), 380–389.

[10] Tim Miller. 2023. Explainable ai is dead, long live explainable ai! hypothesis-driven decision support using evaluative ai. In *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency*. 333–342.

[11] Chiara Natali et al. 2023. Per aspera ad astra, or flourishing via friction: Stimulating cognitive activation by design through frictional decision support systems. In *CEUR workshop proceedings*, Vol. 3481. CEUR-WS, 15–19.

[12] Chiara Natali, Luca Marconi, Leslye Denisse Dias Duran, and Federico Cabitza. 2025. AI-induced deskilling in medicine: a mixed-method review and research agenda for healthcare and beyond. *Artificial Intelligence Review* 58, 11 (2025), 356.

[13] Jeeyun Oh, Soya Nah, and Zinan Darren Yang. 2025. How Autonomy of Artificial Intelligence Technology and User Agency Influence AI Perceptions and Attitudes: Applying the Theory of Psychological Reactance. *Journal of Broadcasting & Electronic Media* 69, 3 (2025), 161–182.

[14] Omer Reingold, Judy Hanwen Shen, and Aditi Talati. 2024. Dissenting explanations: Leveraging disagreement to reduce model overreliance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 21537–21544.

[15] Max Schemmer, Niklas Kuehl, Carina Benz, Andrea Bartos, and Gerhard Satzger. 2023. Appropriate reliance on AI advice: Conceptualization and the effect of explanations. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 410–422.