How do Levels of Automation in AI-Assisted Decision-Making Influence Cognitive Engagement?

Linus Holmberg[1,2,*,†], Maria Riveiro[1] and Tom Ziemke[2]

[1]School of Engineering, Jönköping University, 553 18 Jönköping, Sweden
[2]Department of Computer and Information Science, Linköping University, 581 83 Linköping, Sweden

Abstract
This paper explores how Levels of Automation (LoA) in AI-assisted decision-making (AIDM) affect cognitive engagement. We argue that the human decision-making process is often unintentionally automated rather than supported in AIDM, and that this has consequences for cognitive engagement. When automation replaces rather than augments human analysis, the role of the decision-maker risks shifting to that of a passive supervisor, potentially reducing reflection, learning, and critical scrutiny. Inspired by the LoA framework, we discuss how different system designs influence the distribution of cognitive work between humans and machines across some decision-making stages. We highlight how some forms of automation may invite disengagement, while others preserve or even foster cognitive engagement. Our goal is to promote a more intentional design of AIDM tools; tools that not only deliver outcomes, but also sustain meaningful human engagement.

Keywords
Human-AI Interaction, Human-Computer Interaction, Levels of Automation, Decision-making, Cognitive Engagement

1. Introduction

Artificial intelligence (AI) systems are increasingly used in decision-making across domains. While these systems often enhance efficiency, they risk reducing cognitive engagement by automating aspects of the decision-making process. The AI community initially believed explanations would help decision-makers rely appropriately on AI, leading to human-AI complementary performance [1, 2], but empirical results have been mixed and context-dependent [2, 3, 4, 5, 6, 7, 8].

We argue that the design of AI-assisted decision-making (AIDM) tools influences whether they *support* or *automate* aspects of the human decision-making process. Automating entails "replacing" humans to some degree, whereas supporting entails enhancing them. This distinction matters because automation changes the human's role [9, 10] and can reduce cognitive engagement [11]. Cognitive engagement is important when we want humans to analyze information actively rather than passively [5, 6]. For instance, for learning, reasoning, or critical scrutiny. This paper adopts the Levels of Automation (LoA) framework to explore how different AIDM pipelines influence the human role and cognitive engagement.

This paper aims to elucidate how the implementation of AIDM tools shapes the human decision-maker's role and cognitive engagement. In particular, we discuss how automation at different stages of the decision-making process, such as analysis or decision selection, can either diminish or support user involvement. We also connect our discussion to the emerging concept of *frictional AI* [12], which challenges the assumption that efficiency and seamlessness are always desirable. Instead, this perspective explores how deliberate forms of friction, such as prompting reflection or lowering automation, can help preserve or restore cognitive engagement in AIDM. We return to this in Section 3.

HHAI-WS 2025: Workshops at the Fourth International Conference on Hybrid Human-Artificial Intelligence (HHAI), June 9–13, 2025, Pisa, Italy
*Corresponding author.
†This author conceptualized the initial idea and wrote the first draft of this paper.
linus.holmberg@ju.se (L. Holmberg); maria.riveiro@ju.se (M. Riveiro); tom.ziemke@liu.se (T. Ziemke)
0000-0002-0020-1756 (L. Holmberg); 0000-0003-2900-9335 (M. Riveiro); 0000-0001-6883-2450 (T. Ziemke)

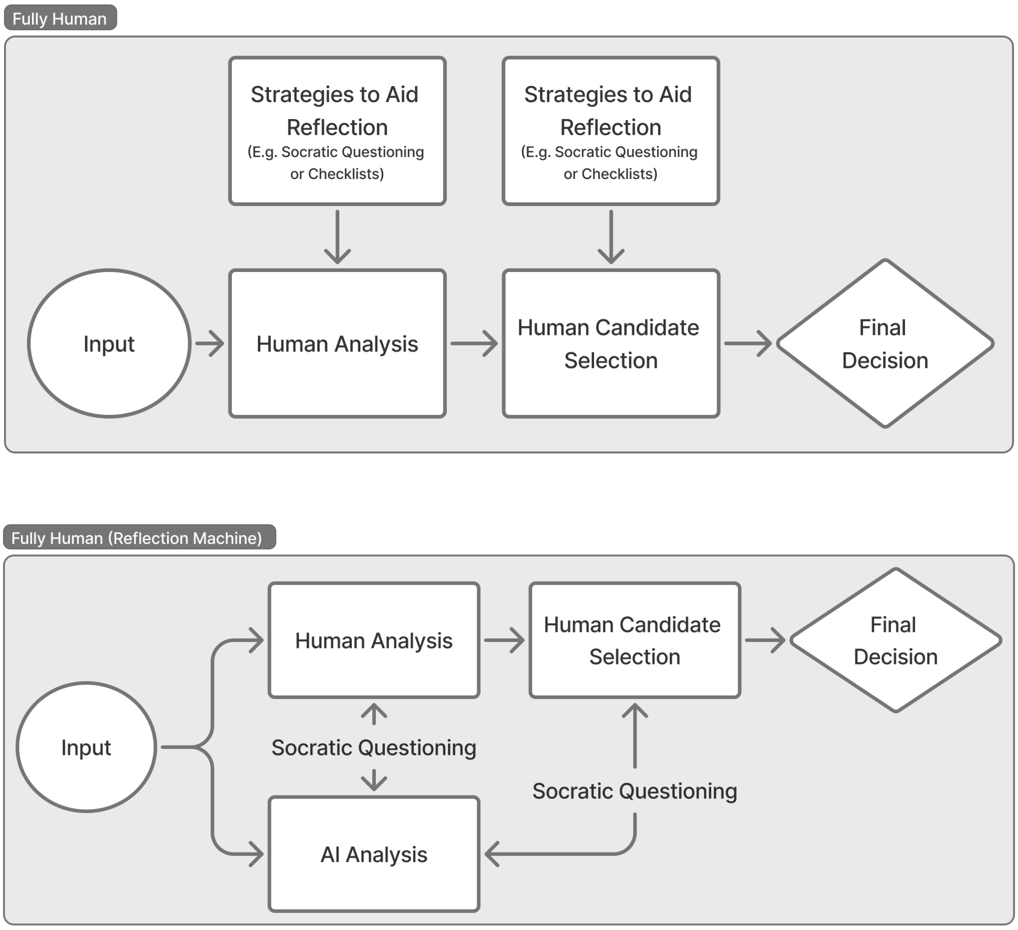

Figure 1: Upper: A fully human decision-making process supported without automation. Lower: The flow of a reflection machine [13], where Socratic questions are generated based on data insights and the decision-maker's analysis with the aim of prompting reflection.

2. Levels of Automation in AIDM

We employ the Levels of Automation (LoA) framework to describe varying degrees of automation of the human decision-making process. LoA places processes on a spectrum ranging from manual to fully automated [14, 15, 16, 17].

Decision-making involves multiple stages, including information acquisition, analysis, decision selection, and action implementation. Each stage can have a different LoA [15]. For instance, information gathering can be fully automated while analysis remains manual. This paper focuses on the analysis and decision-selection stages. However, the LoA analysis could be extended to include all stages. The following tiers (inspired by [15, 14]) are not meant as distinct levels, but as examples across a spectrum.

Fully Human (No/Low LoA): The human controls the decision-making process completely. Analog strategies (e.g., checklists, guided reflection) can support decision-making without automation [18, 19] (Fig.1, upper). AI can also support reasoning without providing analysis or recommendations. For instance, the Reflection Machine [13] uses (Socratic) questions to encourage reflection, forcing users to stay cognitively engaged throughout the decision-making process (Fig.1, lower).

AI for Insight (Moderate LoA): The decision-maker remains the primary driver in the decision-making process, while AI can assist by automating aspects of the analysis or offering hypotheses. One example of this is Evaluative AI [20, 21], where the system presents (automated) evidence for and against different hypotheses *during* the analysis, without explicitly recommending anything. However, this evidence is generated independently from the human analysis (Fig.2), i.e., the human decision-maker is typically not a part of generating the evidence for and against hypotheses.

Human-Supervised AI (High LoA): The system's analysis and decision selection is automated and separate from the human's, shifting the human to a supervisory role (Fig.3), which may lower cognitive engagement [9, 11]. To re-engage decision-makers, strategies like requiring a judgment before showing the AI output have been proposed [5]. However, even if decision-makers get re-engaged, they must still make sense of the AI output post-hoc using backward reasoning [22, 23, 24]. Backward reasoning can introduce bias in decision-making processes [24].

Fully Automated (Max LoA): The system operates independently, making decisions with minimal human intervention (Fig.4). Human cognitive engagement is thus (intended to be) low or absent by design. These systems remove the human decision-maker from the picture. Indeed, these systems typically need to be overseen. But not at every distinct decision.

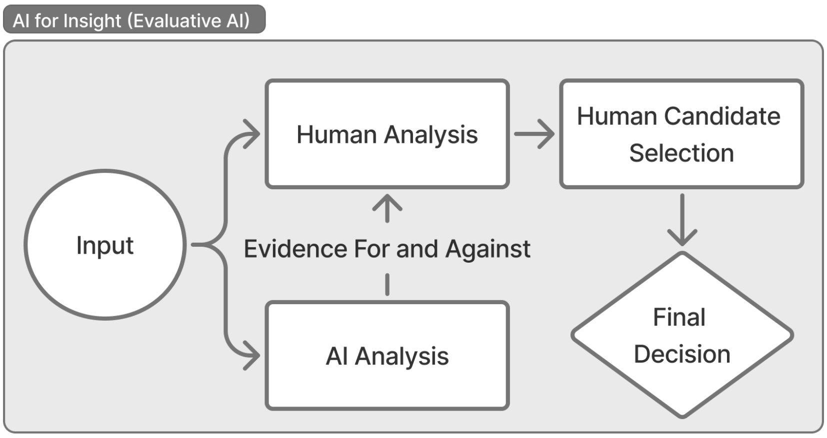

Figure 2: The decision-making flow using the evaluative AI paradigm [20]. The AI-generated evidence is fed to humans during the analysis. However, humans are not a part of this evidence generation.

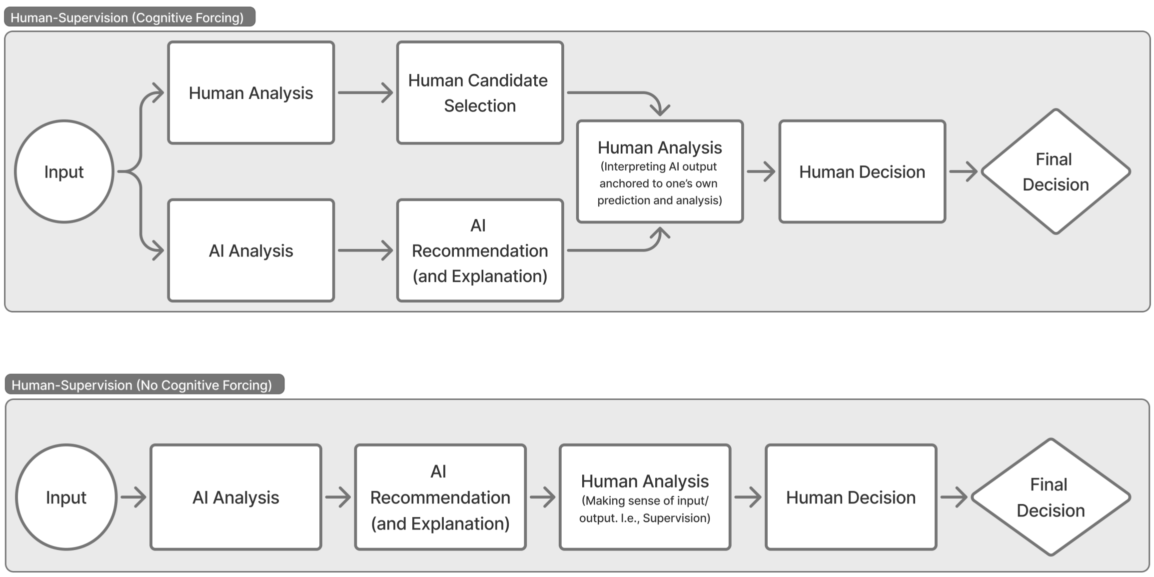

Figure 3: Upper: The decision-making process in most recommendation-driven systems. The decision-maker's starting point is the system output. Lower: A cognitive forcing attempt where the user is forced to make their own prediction *before* seeing the AI output [5]. The analysis processes are separate, and the human is anchored in their own prediction when making sense of the AI output.

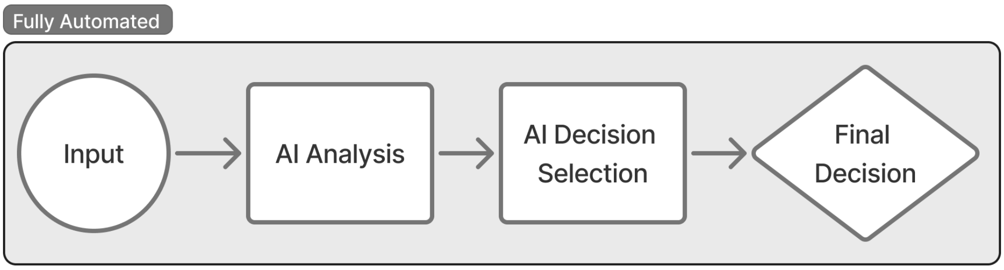

Figure 4: This figure illustrates a fully automated decision-making process without human involvement.

3. Cognitive Engagement and Frictional Interventions Across LoA

By exploring how frictional AI and LoA relate conceptually, we hope to help identify where and when frictional interventions may be appropriate within the decision-making process. By mapping specific stages, such as information analysis or decision selection, to varying degrees of automation, it is potentially possible to identify opportunities for introducing friction in a more targeted way. For instance, some frictional strategies may be particularly valuable in moderate-to-high LoA settings, where the system mostly automates the analytical tasks and the user risks becoming disengaged. In contrast, systems operating at low LoA may already sustain engagement, but can still benefit from scaffolding mechanisms (e.g., [13, 25]). This mapping supports a nuanced understanding of cognitive engagement as something that can be proactively designed for, not merely recovered after the fact, depending on how automation is distributed across the decision-making pipeline.

Many AIDM systems follow a "recommend and defend" approach [20], where AI provides a recommendation and an explanation. Studies show this approach can lead to over-reliance, where users are convinced by the explanations [5, 6, 20, 26]. One critical issue is that recommendation-driven AIDM often separates human analysis from AI analysis, forcing users to interpret system-generated recommendations and explanations post-hoc (Fig. 3) [22, 23], potentially reducing decision-makers to passive supervisors [20, 26, 13], with low engagement. To address this, several studies have proposed cognitive forcing strategies [5, 6, 27, 23], such as requiring users to make an initial judgment before seeing the AI's recommendation, as a way to increase engagement and reduce anchoring effects. However, even interventions like this must contend with the fact that the system's analysis and output remain separate from the user's reasoning process (Fig. 3), requiring the decision-maker to make sense of the output post-hoc [22, 23, 24].

Higher LoA makes it more difficult to maintain cognitive engagement [9, 11]. Thus, when the purpose is to *support* rather than *automate* human decision-making, one could keep the LoA low and potentially preserve user engagement more naturally. This way, intentionally lowering LoA at strategic places in the decision-making process could be seen as introducing friction. Indeed, low LoA does not guarantee an engaged decision-maker. However, decision-makers can still benefit from supportive scaffolding, such as prompts for reflection, as exemplified in the Reflection Machine [13].

Taken together, cognitive engagement in AIDM systems is not only a matter of avoiding high automation but of deliberately designing for the right kind of friction at the right point in the process. Friction should not be seen as a usability flaw, but as a design resource [28], especially in systems where automation risks cognitive disengagement. By deliberately introducing moments that slow down interaction, lowering automation, prompting reflection, or challenging user assumptions, friction in AIDM design can potentially support cognitive engagement across varying levels of automation. Rather than opposing automation, friction and automation can be seen as co-determining forces that shape cognitive engagement in AIDM.

4. Concluding Remarks

Designing AIDM systems requires careful consideration of whether their purpose is to automate or support human decision-making. When the goal is to *support* human decision-making, designers should be careful when introducing automation in ways that risk reducing cognitive engagement. Lower LoA can help maintain cognitive engagement throughout the decision-making process by keeping users actively involved in the decision-making process.

However, we do not argue that all AIDM tools should have low LoA. The appropriate LoA depends on the context, stakes, and goals of the system. In low-stakes or time-critical environments, high automation with minimal human involvement may be both acceptable and beneficial.

Still, this paper has argued that too often, the human decision-making process is unintentionally automated when the intention is to support it. Design choices that shift analysis away from the user can reduce opportunities for reflection and understanding. In particular, we emphasize that fostering cognitive engagement may not only require reducing automation but also deliberately introducing friction, moments that encourage users to slow down, reflect, or reassess, and lowering LoA could be a part of that. Friction, in this sense, is not a flaw in the interface but a tool for keeping human decision-makers cognitively engaged in increasingly automated systems. We hope to inspire researchers and designers to reflect on how and where interventions can be introduced to sustain cognitive engagement in AIDM.

Acknowledgments

We gratefully acknowledge the grant from the Swedish Research Council project XPECT (VR 2022-03180).

Declaration on Generative AI

The authors used ChatGPT-4o and Grammarly for grammar and spelling checks. All content was reviewed, edited and the authors take full responsibility for the content.

References

[1] Y. Zhang, Q. Vera Liao, R. K. Bellamy, Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making, FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (2020) 295–305. URL: https://dl.acm.org/doi/10.1145/3351095.3372852. doi:10.1145/3351095.3372852.

[2] V. Chen, Q. V. Liao, J. Wortman Vaughan, G. Bansal, Understanding the Role of Human Intuition on Reliance in Human-AI Decision-Making with Explanations, Proceedings of the ACM on Human-Computer Interaction 7 (2023). URL: https://dl.acm.org/doi/10.1145/3610219. doi:10.1145/3610219.

[3] G. Bansal, B. Nushi, E. Kamar, W. S. Lasecki, D. S. Weld, E. Horvitz, Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance, in: The Seventh AAAI Conference on Human Computation and Crowdsourcing, 2019. URL: www.aaai.org.

[4] G. Bansal, R. Fok, M. T. Ribeiro, T. Wu, J. Zhou, E. Kamar, D. S. Weld, B. Nushi, Does the whole exceed its parts? The effect of ai explanations on complementary team performance, Conference on Human Factors in Computing Systems - Proceedings (2021) 16. URL: https://dl.acm.org/doi/10.1145/3411764.3445717. doi:10.1145/3411764.3445717.

[5] Z. Buçinca, M. B. Malaya, K. Z. Gajos, To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making, Proc. ACM Hum.-Comput. Interact 5 (2021) 21. URL: https://doi.org/10.1145/3449287. doi:10.1145/3449287.

[6] K. Z. Gajos, L. Mamykina, Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning, International Conference on Intelligent User Interfaces, Proceedings IUI 22 (2022) 794–806. URL: https://doi.org/10.1145/3490099.3511138. doi:10.1145/3490099.3511138.
[7] A. Papenmeier, G. Englebienne, C. Seifert, How model accuracy and explanation fidelity influence user trust, arXiv (2019). URL: https://arxiv.org/abs/1907.12652v1.
[8] Q. V. Liao, K. R. Varshney, Human-Centered Explainable AI (XAI): From Algorithms to User Experiences, arXiv (2021) 1–18. URL: https://arxiv.org/abs/2110.10790v5. doi:10.48550/ARXIV.2110.10790.
[9] L. Bainbridge, Ironies of Automation, in: Analysis, design and evaluation of man–machine systems, Elsevier, 1983, pp. 129–135. doi:10.1016/B978-0-08-029348-6.50026-9.
[10] M. R. Endsley, Ironies of artificial intelligence, Ergonomics 66 (2023) 1656–1668. doi:10.1080/00140139.2023.2243404.
[11] M. R. Endsley, From Here to Autonomy: Lessons Learned from Human-Automation Research, Human Factors 59 (2017) 5–27. doi:https://doi.org/10.1177/0018720816681350.
[12] F. Cabitza, C. Natali, L. Famiglini, A. Campagner, V. Caccavella, E. Gallazzi, Never tell me the odds: Investigating pro-hoc explanations in medical decision making, Artificial Intelligence in Medicine 150 (2024) 102819. URL: https://www.sciencedirect.com/science/article/pii/S0933365724000617. doi:10.1016/J.ARTMED.2024.102819.
[13] P. Haselager, H. Schraffenberger, S. Thill, S. Fischer, P. Lanillos, S. v. d. Groes, M. v. Hooff, Reflection Machines: Supporting Effective Human Oversight Over Medical Decision Support Systems, Cambridge Quarterly of Healthcare Ethics (2023) 1–10. doi:10.1017/S0963180122000718.
[14] M. R. Endsley, D. B. Kaber, Level of automation effects on performance, situation awareness and workload in a dynamic control task, Ergonomics 42 (1999) 462–492. doi:10.1080/001401399185595.
[15] R. Parasuraman, T. B. Sheridan, C. D. Wickens, A model for types and levels of human interaction with automation, IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans. 30 (2000) 286–297. doi:10.1109/3468.844354.
[16] T. B. Sheridan, The System Perspective on Human Factors in Aviation, in: Human Factors in Aviation, 2 ed., Academic Press, 2010, pp. 23–63. doi:10.1016/B978-0-12-374518-7.00002-X.
[17] T. B. Sheridan, W. L. Verplank, T. L. Brooks, Human/computer control of undersea teleoperators, NASA. Ames Res. Center The 14th Ann. Conf. on Manual Control (1978).
[18] K. A. Lambe, G. O'Reilly, B. D. Kelly, S. Curristan, Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review, BMJ Quality & Safety 25 (2016) 808–820. URL: https://qualitysafety.bmj.com/content/25/10/808https://qualitysafety.bmj.com/content/25/10/808.abstract. doi:10.1136/bmjqs-2015-004417.
[19] S. Prakash, R. M. Sladek, L. Schuwirth, Interventions to improve diagnostic decision making: A systematic review and meta-analysis on reflective strategies, Medical Teacher 41 (2019) 517–524. URL: https://www.tandfonline.com/doi/abs/10.1080/0142159X.2018.1497786. doi:10.1080/0142159X.2018.1497786.
[20] T. Miller, Explainable AI is Dead, Long Live Explainable AI! Hypothesis-driven Decision Support using Evaluative AI, ACM International Conference Proceeding Series (2023) 333–342. URL: https://dl.acm.org/doi/10.1145/3593013.3594001. doi:10.1145/3593013.3594001.
[21] T. Le, T. Miller, L. Sonenberg, R. Singh, Towards the New XAI: A Hypothesis-Driven Approach to Decision Support Using Evidence, arXiv (2024). URL: https://arxiv.org/abs/2402.01292v3.
[22] L. Reicherts, Z. T. Zhang, E. von Oswald, Y. Liu, Y. Rogers, M. Hassib, AI, Help Me Think—but for Myself: Assisting People in Complex Decision-Making by Providing Different Kinds of Cognitive Support, Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (2025) 1–19. URL: https://dl.acm.org/doi/pdf/10.1145/3706598.3713295. doi:10.1145/3706598.3713295.
[23] Z. T. Zhang, S. S. Feger, L. Dullenkopf, R. Liao, L. Süsslin, Y. Liu, A. Butz, Beyond Recommendations: From Backward to Forward AI Support of Pilots' Decision-Making Process, Proceedings of the ACM on Human-Computer Interaction 8 (2024). URL: https://dl.acm.org/doi/pdf/10.1145/3687024. doi:https://doi.org/10.1145/3687024.

[24] D. Wang, Q. Yang, A. Abdul, B. Y. Lim, Designing theory-driven user-centric explainable AI, in: Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, Glasgow, 2019, pp. 1–15. URL: https://dl.acm.org/doi/10.1145/3290605.3300831. doi:https://doi.org/10.1145/3290605.3300831.
[25] B. Kim, R. Khanna, O. Koyejo, Examples are not Enough, Learn to Criticize! Criticism for Interpretability, in: 30th Conference on Neural Information Processing Systems (NIPS 2016), 2016.
[26] S. Passi, M. Vorvoreanu, Overreliance on AI: Literature review, Technical Report, Microsoft (AETHER), 2022.
[27] C. Rastogi, Y. Zhang, D. Wei, K. R. Varshney, A. Dhurandhar, R. Tomsett, Deciding Fast and Slow: The Role of Cognitive Biases in AI-assisted Decision-making, Proceedings of the ACM on Human-Computer Interaction 6 (2022) 83. URL: https://dl.acm.org/doi/10.1145/3512930. doi:10.1145/3512930.
[28] S. Inman, D. Ribes, "Beautiful seams": Strategic revelations and concealments, Conference on Human Factors in Computing Systems - Proceedings (2019). URL: https://dl.acm.org/doi/pdf/10.1145/3290605.3300508. doi:10.1145/3290605.3300508;CSUBTYPE:STRING:CONFERENCE.