# Can Less Be More? Understanding the Relation of Expertise and Automation Patterns in Medical XAI

Anton de Vries[1,*], Thomas Franke[1] and Tim Schrills[1]

[1]University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany

### Abstract

Human-centered design of intelligent DSS in medicine could be achieved through lower automation levels in the stage of decision-making. However, reduced automation may affect users differently based on expertise. The objective of our research was to explore how expertise and the use of various automation patterns in XAI systems for medical diagnosis affect user performance and UX. We conducted a between-subjects experiment ($N = 21$) with medical novices and experts. Participants interacted with two DSS differing in automation levels. Performance (accuracy, confidence) and automation-related UX (Subjective information processing awareness, perceived trustworthiness, diagnosticity) were assessed. Results suggest that only novices tend to over-rely on lower automation. In contrast, experts did not exhibit this tendency, but their subjective information processing awareness increased. Our findings indicate that reducing automation levels to enhance human-AI integration may not consistently improve performance or UX. Designers should consider user expertise and context when developing DSS with lower automation levels.

### Keywords

Medical Decision Support System, Human-Centered AI, Explainable AI, Evaluative AI, Trustworthy AI, Human-Automation interaction, Levels of Automation



## 1. Introduction

In medicine, explainable artificial intelligence (XAI) is used to improve the use of clinical decision support systems (DSS) in diagnostic tasks. However, in highly automated XAI, explanations may have an increasingly persuasive effect rather than optimally fostering user understanding of system recommendations [1]. That is, XAI explanations potentially recommend and defend one hypothesis, thereby limiting the extent to which users consider alternative hypotheses. To avoid this restriction to one hypothesis, recent XAI systems have started to reintegrate users into the decision-making process by, for example, enabling them to evaluate and test multiple hypotheses with the system. By refraining from a distinct recommendation, designers reduce the degree of automation found in the XAI system [2].

However, user characteristics, along with users' contexts, may influence which degree of automation is most suitable for them. Choosing an automation pattern (i.e., which information processing stage is automated to which degree), may depend on both, user and task characteristics. For example, users' levels of expertise in the underlying decision-making domain may play a key role in determining the appropriate degree of automation for a DSS.

The influence of expertise on automation-related UX and task performance has been the subject of many studies. For example, results indicate that expertise affects situation awareness (SA) [3] and the ability to identify errors within the system [4]. Additionally, in the context of XAI, expertise has been shown to influence the usefulness of explanation types (e.g., novices struggle with visual explanations) [5]. However, studies often avoid examining these effects in the context of varying automation patterns.

*Joint Proceedings of the ACM IUI Workshops 2025, March 24-27, 2025, Cagliari, Italy*

*Corresponding author.

†These authors contributed equally.

a.devries@uni-luebeck.de (A. de Vries); thomas.franke@uni-luebeck.de (T. Franke); tim.schrills@uni-luebeck.de (T. Schrills)

0000-0002-7211-3771 (T. Franke); 0000-0001-7685-1598 (T. Schrills)

The objective of our research was to empirically investigate how users' expertise in a medical decision (i.e., diagnosing forearm fractures in children) influences the utilization of different automation patterns in regard to users' task performance and automation-related UX. To this end, we developed two interfaces for an AI-based DSS with varying patterns of automation (i.e., Differing levels of automation at the decision selection stage, see [6]). The L4-System automated this stage up to Level 4 [6], providing a single recommendation for diagnosis and treatment. In contrast, the L3-System utilized Level 3 automation, allowing users to test and compare multiple hypotheses. To evaluate the effects of both systems on performance (decision accuracy, decision accuracy in wrong cases, and decision confidence) and automation-related user experience (UX) (Subjective information processing awareness, perceived trustworthiness, and perceived diagnosticity) malfunctions were deliberately integrated into each system for testing with both novices and experts.

All in all, the present work contributes to the human-centered development of artificial intelligence (AI) systems and especially the optimal integration of human and AI. Our experimental study highlights the importance of user expertise when determining the appropriate automation pattern in DSS. Simply involving users in the decision-making process by limiting automation to information analysis does not consistently enhance performance or UX.

## 2. Background

High degrees of automation in the decision-making stages (see [6]) can negatively impact performance [7] and user experience [8]. That is, higher levels of trust may lead to over-reliance: for example, Klingbeil et al. [9], demonstrated that users may tend to follow recommendations even when these contradict available information and their own assessments. XAI aims to enable users to assess system results by making system outputs more intelligible [10]. XAI can tackle the problem of opaque systems, with which users' affordance to engage is lower, which in turn promotes over-reliance [11]. However, Bansal et al. [1] demonstrated that explanations may convince users to comply with the system's reasoning even if it is incorrect.

Apparently, XAI does not guarantee more cognitive engagement. For example, Zhang et al. [12] found that highly automated 'recommend-and-defend' systems may invoke a sense of competition in clinicians, leading to disengagement and abandonment of the system. 'Recommend-and-defend,' is a term from recent XAI literature, referring to Level 4 automation systems (See [6]) that recommend a single option and provide explanations to justify it [2]. In contrast, approaching automated DSS by improving all stages prior to the decision selection could be described as Evaluative-AI (EAI). Introduced by Miller [2], EAI focuses on human hypothesis evaluation by providing users with feedback regarding their specified hypotheses. On a conceptual level, EAI systems exhibit lower levels of automation in the decision selection stage, which may have a positive impact on users' situation awareness [7], decrease over-reliance and reintegrate users into decision-making [2].

### 2.1. Expertise as a major user characteristic in DSS

In addition, user characteristics (i.e. facets of user diversity [13]) may influence how explanations affect over-reliance [14]. One key characteristic that may exhibit this moderating role is user experience. In terms of explanations, experts benefit from different types of explanations (e.g., visual explanations) differently than novices [5]. Goddard et al. [15] conducted a systematic review, finding that both experts and novices are susceptible to automation bias (i.e., users over-rely on recommendations from automated DSS), with novices being more strongly affected. In response, mitigation strategies might be adopted to users' expertise. Kupfer et al. [16] investigated strategies to reduce automation bias in AI-based personnel selection. They found that awareness of potential system errors increases the depth of review and reduces automation bias, leading to better decision-making, which applies to both experts and novices, with novices benefiting particularly from detailed instructions. All in all, previous research demonstrates that both experts and novices are prone to over-reliance on AI-based DSS, with novices being more strongly affected.

### 2.2. Behavior and Experience connected to over-reliance in DSS

To effectively capture the effects of high automation levels in DSS, it is crucial to examine user performance, particularly in scenarios where the AI system provides incorrect recommendations. As demonstrated by the study on trust and reliance in clinical DSS [17], eliciting user performance offers valuable insights into the extent of over-reliance and the ability to override faulty system outputs. In addition to performance measures, research on automation in DSS needs to incorporate psychological variables such as user confidence. Studies like those by Guo et al. [18] highlight that AI systems can induce varying levels of confidence, which influence decision-making.

Furthermore, user experience metrics such as subjective information processing awareness (SIPA) may affect how users interact with AI-systems [19]. The work of Vasconcelos et al. [20] underscores that while explanations can foster trust, they may also inadvertently lead to over-reliance if users perceive the system as infallible. Finally, incorporating diagnosticity into research, defined as the system's ability to provide information that aids users in distinguishing between hypotheses, may connect research on DSS automation and human abductive reasoning. Shin et al. [21] argue that high diagnosticity enhances decision-making by guiding users through complex reasoning processes.

## 3. Present Research

The objective of the present research was to better understand how expertise influences both the user's task performance and automation-related UX when interacting with different automation patterns in XAI systems for diagnostic decision-making. For this study we distinguished between task performance (decision accuracy, decision accuracy in wrong cases, confidence) and automation related UX (SIPA, trust and diagnosticity) as variables. We decided to incorporate a medical decision task that, in fact, is often carried out by insufficiently specialized health care professionals: underarm fractures in children. Fractures are among the most common injuries in children, with 10–25% of all pediatric injuries in Germany involving bones [22]. However, diagnosing and treating fractures in children is often challenging, particularly in rural areas where specialized knowledge may be lacking, leading to potential misdiagnoses, inappropriate treatments, or delays due to referrals. That is, in the present research, we focused on AI support in the correct detection and treatment selection of fractures in children. The assumption that novices and experts react differently to degrees of automation initially leads to assumptions regarding task performance when using an AI system:

**H1** Reducing the level of automation in the decision stage impacts experts' and novices' task performance differently.

Performance-related sub-hypotheses were defined based on **H1**: **H1.A** (decision accuracy), **H1.B** (accuracy in incorrect cases), and **H1.C** (decision confidence). System experience may vary due to factors like better understanding from prior knowledge or lack of critical evaluation due to limited expertise, influencing automation-related UX. Similarly, UX-related sub-hypotheses were defined for **H2**: **H2.A** (SIPA), **H2.B** (perceived trustworthiness), and **H2.C** (perceived diagnosticity).

**H2** Reducing the level of automation in the decision stage impacts experts' and novices' automation-related user experience.

## 4. Method

### 4.1. Experimental Setup & Stimuli Design

In order to examine our hypotheses, we developed two DSS interfaces, employing varying levels of automation on the stage of decision selection: (A) Level 4 automation: Follows the recommend and defend approach, in which the system recommends one diagnosis and treatment for pediatric forearm fractures while defending its recommendation with explanations (See Figure 1). The system uses feature-importance explanations to communicate which features of the patient data contributed to

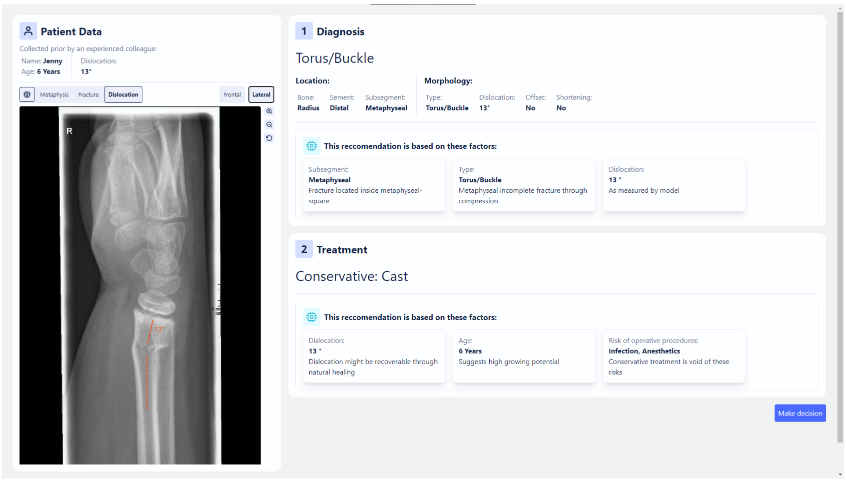
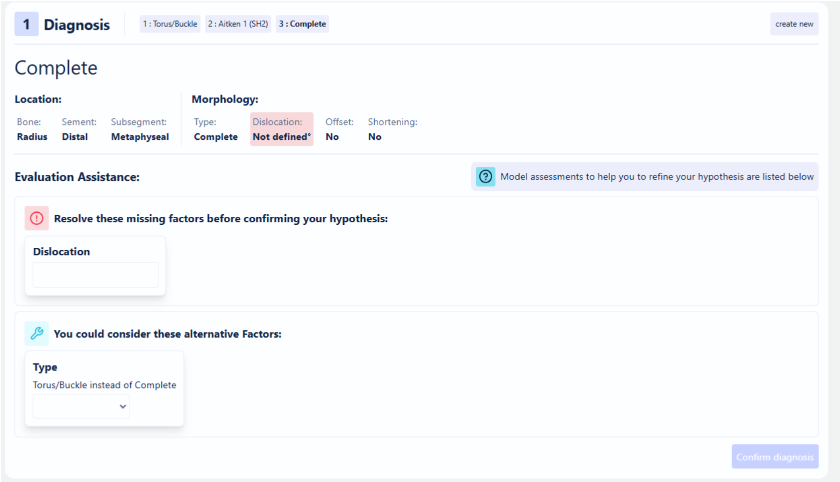



**Figure 1:** Prototypical L4 interface depicting patient data, diagnosis, and treatment recommendations. The interface includes the X-ray image, diagnosis explanation, and treatment explanation, with highlighted factors contributing to each recommendation.

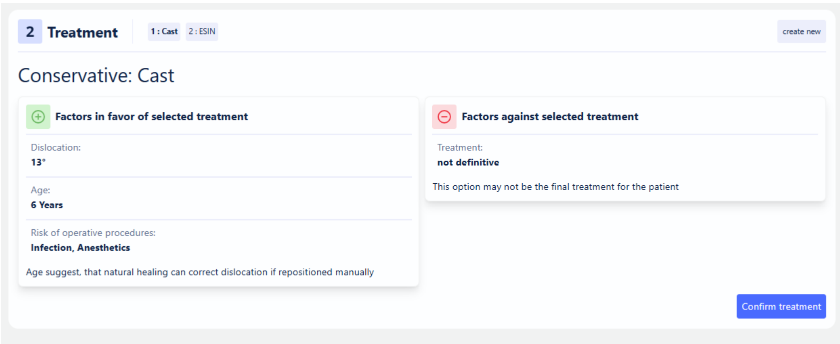
the calculation of the recommendation. (B) Level 3 automation: Offers feedback to user-generated hypotheses, and enables the comparison and exploration of multiple explanations for diagnosis and treatment selection (See Figure 2a and Figure 2b). Both interfaces were ultimately integrated into a web app that served as the study environment, guiding participants through the study. To standardize decision support, a scripted procedure replaced a genuine AI model for fracture detection or treatment selection. Four medical experts (physicians, pediatricians, and radiologists) from a local hospital selected 12 real pediatric forearm fractures from their database as stimuli. The experts simulated all system outputs, including recommendations, feedback, explanations, and X-ray markings. In 3 of the 12 cases, incorrect recommendations were introduced by simulating miscalculated dislocations, leading to inappropriate treatment suggestions.

## 4.2. Scales and Measures

Medical decision performance was evaluated by (1) decision accuracy and (2) decision confidence (0–10 scale: 0 = not confident, 10 = very confident). Automation-related UX was assessed through perceived information processing awareness, trustworthiness, and diagnosticity, with trustworthiness measured using cognition-based trust questions from Madsen and Gregor [23]. Perceived diagnosticity was assessed with a self-developed 4-item questionnaire (The system's outputs helps me to check my assumptions; The system's output supports me to justify my decision; The system's output enables me to rule out wrong decisions; Looking at the system's output does not help me to make a decision). SIPA was measured with the SIPA Scale [24]. To account for user diversity, we measured affinity for technology interaction (ATI) [25] and self-rated preference for automation in pediatric fracture diagnosis and treatment using the PATS scale [26]. All questionnaires employed a 6-point Likert scale (1 = completely disagree, 6 = completely agree). We used two-sided $t$-tests for independent samples to compare the difference values (L4 - L3) for all dependent variables (decision accuracy, decision accuracy in wrong cases, confidence, SIPA, trust, and diagnosticity) between experts and novices. As a consequence of our underpowered study, we focus on interpreting effect sizes, we report as Cohen's $d$ (small: $d = 0.2$, medium: $d = 0.5$, and large: $d = 0.8$) [27]. Prior to all performed $t$-tests we tested for normality distribution (Shapiro-Wilk Test) and homogeneity of variance (Levene Test), which were given for all variables.




(a) Prototypical L3 system, showing the diagnosis evaluation step.




(b) Prototypical L3 system, illustrating the treatment recommendation step.

**Figure 2:** Interfaces of System A and System B, illustrating differences in their automation levels through distinct feedback.

### 4.3. Recruitment and Procedure

Ethics approval was granted by the local university's Ethics Committee [anonymous for submission]. Participants ($N$ = 21) included experts ($n$ = 11; physicians and pediatricians) and novices ($n$ = 10; medical students without pediatric specialization). Each received €15 compensation. Participants completed a sociodemographic questionnaire, received instructions, and watched a DSS introduction video. They then worked on four patient cases using L3 and L4, reporting decisions and confidence. After each system, they rated information processing awareness, trustworthiness, and diagnosticity. System order was randomized, with every third case in each block featuring incorrect recommendations.

## 5. Results

A total of 22 participants ($N$ = 22) participated, with one removed due to logging errors. The final sample ($N$ = 21) included 11 experts and 10 novices. The mean ATI score for the sample was $M$ = 3.32 ($SD$ = 0.94), close to the general population value of 3.61 [25]. Experts had a similar score ($M$ = 3.36, $SD$ = 0.94), while novices had a slightly lower score ($M$ = 3.28, $SD$ = 0.97). The mean PATS score for the sample was $M$ = 4.21 ($SD$ = 0.58), with experts preferring slightly lower automation ($M$ = 4.05, $SD$ = 0.49) compared to novices ($M$ = 4.40, $SD$ = 0.64). We did not find significant differences between experts' and novices' differences values (L4 - L3) for all variables, which led to the rejection of all Hypotheses (**H1 A-C, H2 A-C**). We did observe medium effect sizes for decision accuracy in wrong cases, SIPA and perceived diagnosticity. The results are depicted in Table 1.

**Table 1**
Results of measured variables for experts ($n$=11) and novices ($n$=10)

| | Experts | | Novices | | $t(19)$ | $p$ | $d$ |
|---|---|---|---|---|---|---|---|
| | $M$ | $SD$ | $M$ | $SD$ | | | |
| H1: Task Performance | | | | | | | |
| H1.A: Decision Accuracy | 0.39 | 0.68 | 0.25 | 0.89 | 0.40 | .700 | 0.17 |
| H1.B: Decision Accuracy (Wrong Cases) | -0.27 | 1.10 | 0.70 | 2.21 | -1.29 | .211 | 0.57 |
| H1.C: Decision Confidence | 0.52 | 1.00 | 0.42 | 1.37 | 0.19 | .853 | 0.08 |
| H2: User Experience | | | | | | | |
| H2.A: SIPA | 0.27 | 0.51 | 0.05 | 0.44 | 1.07 | .300 | 0.47 |
| H2.B: Perceived Trustworthiness | 0.07 | 0.52 | -0.11 | 0.55 | 0.74 | .468 | 0.32 |
| H2.C: Perceived Diagnosticity | 0.07 | 0.73 | -0.32 | 0.73 | 1.58 | .130 | 0.69 |

**Note:** The table presents the differences in mean values (S4 - S3) of for each variable between experts and novices. Positive values indicate a preference for the L3 System, negatives for the L4 System

## 6. Discussion and Conclusion

This study examined the influence of user expertise on task performance and automation-related UX in XAI systems with differing automation patterns. None of the hypotheses (**H1.A, H1.B, H1.C, H2.A, H2.B, H2.C**) were supported. However, concerning our limited sample size and potentially under-powered study design, we found a medium effect size for decision accuracy of wrong cases when comparing experts and novices. Novices' decision accuracy of wrong cases seems to suffer more extensively from a lower level automation than experts' decision accuracy. Additionally, we found a medium effect size for SIPA and diagnosticity. Novices' perceived diagnosticity appears to increase to a greater extent when reducing automation levels compared to experts. Conversely, a decrease in automation levels shows a greater decrease on SIPA in experts than in novices. Based on our findings, we assume that expertise may enable experts to utilize low levels of automation effectively, whereas novices may lack the necessary knowledge to do so. This is evident in our observation that only novices were unable to rely on the low automation system accurately, while at the same time exhibiting an increased perception of trustworthiness. This is in line with the results of [4], which indicate a moderating effect of expertise on users' ability to accurately rely on DSS. Additionally, in their study, experts appeared to use the DSS for X-ray screening of passenger baggage to confirm their own hypotheses, whereas novices used it as a guide to base their decisions on. In our study, experts and novices might have applied different hypothesis strategies comparable to the ones described in [4]. The way L3 systems integrates users into decision-making and present information may be particularly suitable for users in the process of knowledge development, as such systems can help clarify relationships in complex situations more effectively [28]. This is reflected in our observed increase in perceived diagnosticity among novices when using systems with lower levels of automation. The present study exhibits certain limitations that need to be considered when interpreting our findings: First, the sample consisted solely of German practitioners or students. Second, we did not utilize an underlying AI-Model for both DSS. Third, our system was deigned based on the assumption that a definitive ground truth or correct decision in medical diagnosis can be defined, which is not always the case for complex medical decisions.

The present research provides insights into the potential impact of user expertise on performance and UX in the context of automation patterns (i.e. reducing automation levels in AI-based DSS). Findings indicate that novices are affected to a greater extent by reduced automation levels compared to experts. When designing DSS with lower automation levels, designers should carefully account for user expertise and tailor the design accordingly.

## Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT, in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

## References


[1] G. Bansal, T. Wu, J. Zhou, R. Fok, B. Nushi, E. Kamar, M. T. Ribeiro, D. Weld, Does the whole exceed its parts? the effect of ai explanations on complementary team performance, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 1–16. URL: https://doi.org/10.1145/3411764.3445717. doi:10.1145/3411764.3445717.

[2] T. Miller, 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, FAccT '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 333–342. URL: https://doi.org/10.1145/3593013.3594001. doi:10.1145/3593013.3594001.

[3] M. R. Endsley, Expertise and Situation Awareness, Cambridge Handbooks in Psychology, Cambridge University Press, 2006, p. 633–652.

[4] A. Chavaillaz, A. Schwaninger, S. Michel, J. Sauer, Expertise, automation and trust in x-ray screening of cabin baggage, Frontiers in Psychology 10 (2019). URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.00256. doi:10.3389/fpsyg.2019.00256.

[5] M. Szymanski, M. Millecamp, K. Verbert, Visual, textual or hybrid: the effect of user expertise on different explanations, in: Proceedings of the 26th International Conference on Intelligent User Interfaces, IUI '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 109–119. URL: https://doi.org/10.1145/3397481.3450662. doi:10.1145/3397481.3450662.

[6] R. Parasuraman, T. Sheridan, C. 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. URL: http://ieeexplore.ieee.org/document/844354/. doi:10.1109/3468.844354.

[7] L. Onnasch, C. D. Wickens, H. Li, D. Manzey, Human performance consequences of stages and levels of automation: An integrated meta-analysis, Human Factors 56 (2014) 476–488. URL: https://doi.org/10.1177/0018720813501549. doi:10.1177/0018720813501549. arXiv:https://doi.org/10.1177/0018720813501549, pMID: 24930170.

[8] M. Langer, C. König, M. Papathanasiou, Highly-automated job interviews: Acceptance under the influence of stakes, International Journal of Selection and Assessment 27 (2019). doi:10.1111/ijsa.12246.

[9] A. Klingbeil, C. Grützner, P. Schreck, Trust and reliance on ai — an experimental study on the extent and costs of overreliance on ai, Computers in Human Behavior 160 (2024) 108352. URL: https://www.sciencedirect.com/science/article/pii/S0747563224002206. doi:https://doi.org/10.1016/j.chb.2024.108352.

[10] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, G.-Z. Yang, Xai—explainable artificial intelligence, Science Robotics 4 (2019) eaay7120. URL: https://www.science.org/doi/abs/10.1126/scirobotics.aay7120. doi:10.1126/scirobotics.aay7120. arXiv:https://www.science.org/doi/pdf/10.1126/scirobotics.aay7120.

[11] V. Alexander, C. Blinder, P. J. Zak, Why trust an algorithm? performance, cognition, and neurophysiology, Computers in Human Behavior 89 (2018) 279–288. URL: https://www.sciencedirect.com/science/article/pii/S0747563218303480. doi:https://doi.org/10.1016/j.chb.2018.07.026.

[12] S. Zhang, J. Yu, X. Xu, C. Yin, Y. Lu, B. Yao, M. Tory, L. M. Padilla, J. Caterino, P. Zhang, D. Wang, Rethinking Human-AI Collaboration in Complex Medical Decision Making: A Case Study in Sepsis Diagnosis, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing

Systems, CHI '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 1–18. URL: https://dl.acm.org/doi/10.1145/3613904.3642343. doi:10.1145/3613904.3642343.
[13] C. Attig, D. Wessel, T. Franke, Assessing personality differences in human-technology interaction: An overview of key self-report scales to predict successful interaction, in: C. Stephanidis (Ed.), HCI International 2017 – Posters' Extended Abstracts, Springer International Publishing, Cham, 2017, pp. 19–29.
[14] R. Nimmo, M. Constantinides, K. Zhou, D. Quercia, S. Stumpf, User characteristics in explainable ai: The rabbit hole of personalization?, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 1–13. URL: https://doi.org/10.1145/3613904.3642352. doi:10.1145/3613904.3642352.
[15] K. Goddard, A. Roudsari, J. C. Wyatt, Automation bias: a systematic review of frequency, effect mediators, and mitigators, Journal of the American Medical Informatics Association 19 (2011) 121–127. URL: https://doi.org/10.1136/amiajnl-2011-000089. doi:10.1136/amiajnl-2011-000089.
[16] C. Kupfer, R. Prassl, J. Fleiß, C. Malin, S. Thalmann, B. Kubicek, Check the box! how to deal with automation bias in ai-based personnel selection, Frontiers in Psychology 14 (2023). URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1118723. doi:10.3389/fpsyg.2023.1118723.
[17] B. Y. Lim, Improving understanding and trust with intelligibility in context-aware applications, Ph.D. thesis, Carnegie Mellon University, USA, 2012. AAI3524680.
[18] S. Guo, F. Du, S. Malik, E. Koh, S. Kim, Z. Liu, D. Kim, H. Zha, N. Cao, Visualizing uncertainty and alternatives in event sequence predictions, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 1–12. URL: https://doi.org/10.1145/3290605.3300803. doi:10.1145/3290605.3300803.
[19] T. Schrills, T. Franke, How do users experience traceability of ai systems? examining subjective information processing awareness in automated insulin delivery (aid) systems, ACM Trans. Interact. Intell. Syst. 13 (2023). URL: https://doi.org/10.1145/3588594. doi:10.1145/3588594.
[20] H. Vasconcelos, M. Jörke, M. Grunde-McLaughlin, T. Gerstenberg, M. S. Bernstein, R. Krishna, Explanations can reduce overreliance on ai systems during decision-making, Proc. ACM Hum.-Comput. Interact. 7 (2023). URL: https://doi.org/10.1145/3579605. doi:10.1145/3579605.
[21] D. Shin, A. Koerber, J. S. Lim, Impact of misinformation from generative ai on user information processing: How people understand misinformation from generative ai, New Media & Society 0 (2024) 14614448241234040. URL: https://doi.org/10.1177/14614448241234040. doi:10.1177/14614448241234040. arXiv:https://doi.org/10.1177/14614448241234040.
[22] I. Marzi (Ed.), Kindertraumatologie, Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. URL: http://link.springer.com/10.1007/978-3-642-44997-0. doi:10.1007/978-3-642-44997-0.
[23] M. Madsen, S. Gregor, Measuring human-computer trust, in: 11th australasian conference on information systems, volume 53, Citeseer, 2000, pp. 6–8.
[24] T. Schrills, M. Zoubir, M. Bickel, S. Kargl, T. Franke, Are Users in the Loop? Development of the Subjective Information Processing Awareness Scale to Assess XAI, in: Proceedings of the ACM CHI Workshop on Operationalizing Human-Centered Perspectives in Explainable AI ., 2021, pp. 1–10. URL: https://hcxai.jimdosite.com/hcxai-21-contents/.
[25] T. Franke, C. Attig, D. Wessel, A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale, International Journal of Human–Computer Interaction 35 (2019) 456–467. URL: https://www.tandfonline.com/doi/full/10.1080/10447318.2018.1456150. doi:10.1080/10447318.2018.1456150.
[26] M. Zoubir, Preference for Automation Types Scale (PATS), 2024. URL: https://rgdoi.net/10.13140/RG.2.2.22149.97769. doi:10.13140/RG.2.2.22149.97769.
[27] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2 ed., Routledge, 2013. URL: https://www.taylorfrancis.com/books/9781134742707. doi:10.4324/9780203771587.
[28] H. Khosravi, S. B. Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, S. Knight, R. Martinez-Maldonado, S. Sadiq, D. Gašević, Explainable artificial intelligence in education, Computers and Education: Artificial Intelligence 3 (2022) 100074. URL: https://www.sciencedirect.com/science/article/pii/

S2666920X22000297. doi:https://doi.org/10.1016/j.caeai.2022.100074.