Intelligent Human Systems Integration (IHSI 2025), Vol. 160, 2025, 194–205
https://doi.org/10.54941/ahfe1005825

Investigating the Impact of Confidence Scores in AI-Based Decision Support Systems on Decision Quality and Reliance in Work Contexts

Antonia Markus, Lea M. Daling, Esther Borowski, and Ingrid Isenhardt

Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University, Germany

ABSTRACT

This paper reviews the current literature on uncertainty communication in AI-based Decision Support Systems (DSS) and underscores the necessity of practice-oriented studies to better understand the effects of uncertainty communication on performance, task load, and attitudes toward AI. It identifies limitations in existing research regarding different forms of uncertainty communication and proposes a conceptual framework for an experimental study design. This framework aims to guide researchers interested in pursuing practice-oriented investigations in this field, ultimately contributing to the development of more effective AI-DSS.

Keywords: Confidence score, Uncertainty communication, AI-assisted decision-making

INTRODUCTION

Artificial Intelligence-Based Decision Support Systems (AI-DSS) are emerging as a transformative force within the framework of Industry 4.0, offering unprecedented capabilities to enhance operational efficiency, quality control, supply chain management, and risk mitigation (Burggräf et al., 2020). Despite the challenges associated with implementing these technologies, ongoing advancements in AI promise to further revolutionize industrial decision-making processes (Kasie et al., 2017).

However, effective interaction between users and AI systems is crucial for successful adoption and utilization. Research indicates that users often hesitate to engage with these systems due to difficulties in evaluating their quality and inconsistencies in the results produced. As the prevalence of AI systems in workplace environments continues to rise, it becomes increasingly important for users to be able to evaluate these systems and interpret their output.

One promising aspect of enhancing human integration in the decision process is the use of confidence scores. These scores play a vital role in helping users assess the reliability of AI outputs, thereby influencing their reliance on the system's recommendations and overall decision quality.

This concept paper aims to contribute to this area by analyzing the theoretical background related to uncertainty communication in AI-DSS. It will derive research questions that address identified gaps in the existing literature and present a discussion of a potential study design aimed at investigating these questions.

Theoretical Background

Artificial Intelligence-based decision support systems (AI-based DSS) are increasingly utilized across various decision-making contexts. DSS are computer-aided information systems that support the preparation of decisions at various management levels by condensing decision-relevant information and presenting it appropriately (Bhatt & Zaveri, 2002). The ability of AI systems to extract insights from large datasets has made them valuable collaborators to support humans in domains such as medicine, business, and design (Buch et al., 2018; Nagar & Malone, 2011; Patel et al., 2019; Zhang et al., 2021). Especially in work contexts, AI-based systems are used to support users in work-related decisions. Since the ongoing trend of automation offers the potential to take over repetitive tasks, employees are more and more confronted with complex decision-making scenarios. In medicine, for example, AI-based DSS can assist radiologists in the diagnostic process based on the analysis and interpretation of chest radiographs (Patel et al., 2019). However, lower-threshold areas of application can also make everyday work easier for employees. For example, AI-based DSS can be used successfully in factory planning, job scheduling, project management, or answering customer queries in customer support (Münker et al., 2023). Making effective and efficient decisions is often something that requires a lot of domain knowledge and many years of work experience and is therefore a huge challenge, especially for novices. Supporting employees by gathering huge amounts of data from many different sources and thus facilitating decision making is the aim of the implementation of AI-DSS systems. A successful AI-DSS is able to increase employee performance by enabling them to make decisions more effectively and efficiently, as well as reducing their perceived workload (Zhang et al., 2020). Moreover, a positive attitude towards AI-based DSS is both a prerequisite and a consequence of successful interaction between humans and AI (Lai & Tan, 2018).

These three target variables (performance, perceived task load and attitude towards AI) of successful AI-based DSS are explained in more detail below.

Performance

Enhancing performance often primarily relates to improvements in the accuracy and efficiency of processes facilitated by AI implementation. However, this effect does not necessarily involve collaborative decision-making between humans and AI, where the goal is to enhance the worker's performance through AI support. In such cases, the worker may complete tasks more quickly, increasing efficiency, or making fewer errors, thereby improving effectiveness and decision quality. In a recent meta-analysis, Vaccaro and co-authors investigated human and AI performance in comparison to a collaborative scenario of both. Their findings showed that when humans outperform AI, the combination tends to result in performance gains human-AI combinations often perform worse than the best-performing human or AI alone (Vaccaro et al., 2024). Specifically, when AI outperforms humans, combining both can lead to performance losses, whereas.

Despite these findings, in many work contexts, accountability for decisions must remain with humans. An AI system alone cannot be solely responsible for most decisions. Thus, identifying the factors that contribute to enhanced performance and thus success of joint human-AI decision-making remains a critical area of research.

Even if performance is one of the decisive criteria for evaluating the use of new systems, the importance of subjective criteria is repeatedly emphasized, particularly in the field of technology acceptance research (Venkatesh et al., 2003). The acceptance of users and thus also the likelihood of using a system depends to a large extent on user satisfaction during use. The following section looks at a common and well-researched criterion of user satisfaction that focuses on the perceived workload when using a system.

Perceived Task Load

The complexity of decisions where AI is employed can be highly demanding, making it essential that AI does not further increase the cognitive load on users. Task characteristics such as complexity directly influence cognitive load, which in turn affects worker performance (Hancock et al., 1995). Ideally, AI systems should alleviate this load for their users (Bläsing & Bornewasser, 2021).

Research indicates that task complexity and uncertainty significantly affect user reliance on AI systems. Specifically, complex and uncertain tasks tend to increase user reliance on AI, but this reliance is often inappropriate (Salimzadeh et al., 2024). One major factor contributing to this behavior is overtrust. High cognitive load during complex tasks can lead individuals to over-rely on AI, often resulting in an overestimation of the AI's capabilities (Goddard et al., 2014).

These findings highlight the importance of investigating factors that help reduce overtrust in AI systems. A critical factor in this context is task complexity. Previous research has often missed integrating task complexity as well as practice-oriented decisions, instead constructing experiments with simplified, binary decisions and artificial tasks that lack comparability to real-world decision-making scenarios (Lai et al., 2023).

However, while the use of such systems is beneficial, their effectiveness largely depends on user acceptance, attitude, and appropriately calibrated trust. Therefore, special attention should be given to these factors when analyzing or implementing AI systems.

Attitude Towards AI

The attitude of the users towards the systems and AI in general is a predisposition for the actual use of the systems. This has a long research history and is described by models such as the technology acceptance model (TAM) by Davis (*Technology Acceptance Model*, 1989). However, the attitude towards AI is mixed among the people (Lichtenthaler, 2020), especially those who are not that experienced with the use of those systems. According to the TAM a negative attitude towards a technology, for example AI, can result in less use of such systems. Attitude is crucial, but not so easy to control and influence. All three mentioned variables (performance, perceived task load, and attitude towards a system) are influenced by many different factors - these include relevant user variables (previous experience, age, gender) but also, for example, the way in which the AI system is designed. Uncertainty communication in particular plays a role here (Zhang et al., 2020). In the following, the crucial role of uncertainty communication in system design is described. Furthermore, we will analyse which influence it can potentially have on the three target variables performance, perceived task load and attitude towards a system.

Uncertainty Communication

The design of an AI system can have a significant influence on performance and user satisfaction. One design strategy to enhance collaboration with AI-based systems is uncertainty communication. This refers to the communication of uncertainty that is inherently present in machine learning models for example, as they rely on statistical and mathematical calculations.

Global and Local Uncertainty Communication

Uncertainty communication can be categorized into global and local uncertainty communication. Global uncertainty communication includes metrics like accuracy, which, in the context of decision support systems, indicates the overall proportion of correct system recommendations. Providing information about the system's accuracy can lead to beneficial outcomes, such as reduced decision-making time and more efficient strategies in managing system recommendations (Lukashova-Sanz et al., 2023). However, global uncertainty communication has a key limitation: it does not offer insights into the correctness of the AI's recommendation for a specific case. This is where local uncertainty communication comes into play. Local uncertainty communication provides probabilistic information on a case-by-case basis, often by displaying confidence scores. These scores help users gauge the likelihood that the AI's recommendation is correct for a particular instance. Research suggests that confidence scores can help calibrate users' trust in an AI model (Zhang et al., 2020). Communicating uncertainty can reduce overreliance by forcing them to think analytically in cases of high uncertainty (Prabhudesai et al., 2023).

However, trust calibration alone may not be sufficient to improve AI-assisted decision-making. The effectiveness of decision-making may also depend on how well users can leverage their unique knowledge to compensate for potential AI errors (Zhang et al., 2020). Additionally, focusing solely on AI confidence scores is not enough; it is also important to consider the likelihood of human correctness when evaluating the recommendations (Ma et al., 2024). These mixed findings highlight that the conditions under which local uncertainty communication is most beneficial in AI-based decision support systems are still not fully understood and warrant further exploration.

Numerical and Visual Forms of Uncertainty Communication

Beside the effect of local uncertainty communication, also the representation varies between the studies. The approaches to decision support can be categorized into numerical and visual designs. The numerical approach involves presenting the probabilistic information of each decision option in the form of confidence scores e.g. percentages or decimals. These scores quantify the likelihood that a given option is correct. In contrast, the visual approach translates these numerical values into symbols, such as traffic lights, arrows, or smiley faces. For example, a green dot may indicate a confidence score above 0.70, a yellow dot could represent scores between 0.50 and 0.79, and a red dot may denote confidence levels below 0.40. Besides this direct translation violin plots or an indication with question marks have been used (Zhao et al., 2024). Both approaches have their respective strengths and weaknesses. Numerical communication offers greater precision, allowing users to understand the exact probability of each decision option. However, this precision can also be a drawback, as users might misinterpret or place undue emphasis on small percentage differences, particularly when the values between options are close. Visual indicators, on the other hand, are generally less detailed but can simplify decision-making. For instance, a simple three-color traffic light system conveys less information than a gradually filling progress bar, which provides a more nuanced representation of confidence levels. Despite this, visual symbols may reduce the cognitive load on users and facilitate quicker decision-making, albeit at the cost of reduced precision. A strength of symbols is that they are more accessible and easier to interpret for most people. Particularly, because research suggests that many people have problems interpreting probabilistic information (Gigerenzer et al., 2007). Even more surprising is that people, if asked, tend to prefer numerical information over symbols and claim to be able to interpret the percentages. However, interpreting percentages is very complex. Depending on the calculation they can mean different things. The total for each of the total is split between different options. For example, 50% in a binary decision scenario would state that both options are equally likely. But in a scenario with 10 options, if one option has 50% it means that the other 50% is potentially distributed to 9 options and the first option can be very likely. However, empirical evaluation of the effectiveness of those visualizations is rare. Regarding its influence on the reliance on decision support tools, it is known that the cognitive accessibility of the visualization technique, the perception of the model as well as the task difficulty matters.

Uncertainty communication might have a positive impact on the performance of AI-assisted decision, especially when overtrust is an issue. But it is not yet clear if the numerical representation or the visual representation is more beneficial.

Aim of the Present Work and Derivation of the Research Question

AI-DSS have not yet reached their full potential in maximizing user performance and enhancing satisfaction. Uncertainty communication, as an integral part of system design, could play a crucial role in positively influencing performance, task load, and attitudes toward AI. However, there remains a significant gap in understanding the effects of different forms of uncertainty communication. Current studies reveal several weaknesses, particularly concerning sample representativeness, the one-dimensional nature of decision-making frameworks, and insufficient consideration of the application context, such as the work environment (Lai et al., 2023). Most existing research relies on samples composed primarily of students who lack experience in real work environments, rendering them unsuitable for assessing work-related decisions. Additionally, many decision scenarios are artificially constructed to facilitate control within labor settings but are less reflective of actual workplace realities. Furthermore, numerous scenarios focus solely on binary decisions—such as "sick" versus "healthy"—which overlook more complex decision-making processes that involve multiple options. This simplistic decision structure may contribute to increased complexity related to task load and performance (Lai et al., 2023).

There is need for further research in real-world settings, emphasizing the application of qualitative research designs and the inclusion of representative samples to better understand the implications and effectiveness of AI-assisted decision-making (Mahmud et al., 2022). The presented studies in this field do not sufficiently consider the specific characteristics of work contexts and focus on artificially constructed decision-making situations that are not transferable to professional settings.

The aim of this study is to address the methodological weaknesses identified in previous research and to analyze the effects of different forms of uncertainty communication on three target variables: performance, perceived load, and attitudes toward AI. This paper presents a proposed study design to investigate the following research questions:

Do individuals differ in their performance, perceived load, and attitudes toward AI based on whether and what type of uncertainty communication the AI system provides?

In the proposed study, we anticipate differences in performance among participants who receive no uncertainty communication compared to those who receive visual uncertainty communication (using arrows) as well as those who receive numerical uncertainty communication. Additionally, we expect variations in perceived load between participants with no uncertainty communication and those with visual or numerical uncertainty communication. Furthermore, we anticipate differences in attitudes toward AI among participants based on the type of uncertainty communication received—specifically comparing those with no uncertainty communication to those receiving visual or numerical formats.

The following paragraphs outline a planned study concept aimed at investigating these research questions.

Methodical Approach

The theoretical challenges associated with researching AI-DSS and the derived hypotheses necessitate a comprehensive and holistic study design. This design must consider the working context, the unique characteristics of the target population, and the multidimensional nature of decision-making situations. Below, we present the components of our methodological approach and the planned study design that address these challenges.

Sample

The recruitment goal for this sample is to ensure it accurately represents the working population. Therefore, participants must be employed and aged between 18 and 65 years. There are no restrictions regarding industry sectors, as AI-DSS can serve as valuable tools for supporting novices in various work activities. Additionally, participants should possess basic computer skills. During recruitment, AI was not mentioned to avoid deterring individuals with negative attitudes toward AI. The targeted sample size is 150 participants.

Instruments

This study aims to investigate how uncertainty communication influences objective performance, perceived task load, and user attitudes toward AI. The operationalization of these variables is detailed below.

Objective Performance Measures: One indicator of efficiency will be the time taken by participants to respond to each decision scenario. Effectiveness will be assessed through decision quality, measured by the number of correct decisions made.

Perceived task load: The NASA Task Load Index (NASA TLX) questionnaire (Hart & Staveland, 1988) will be used to measure participants' cognitive load. This measurement allows for comparisons between perceived load during decision-making without AI support versus with AI support.

Attitude towards AI: Attitudes toward AI will be measured both before and after interaction with AI support to determine whether this interaction impacts acceptance levels. The assessment will utilize the Attitudes Towards Artificial Intelligence Scale (ATTARI-12) (Stein et al., 2024).

Finally, we will employ the MAILS AI literacy questionnaire (Carolus et al., 2023) to evaluate participants' competencies in dealing with AI—an important control variable given its potential influence on interactions with AI-DSS outputs. Additional demographic information such as age, gender, and profession will also be collected as they may affect performance, perceived load, and attitudes toward AI.

Task

A task was selected that exemplifies a typical application of artificial intelligence (AI) in the workplace, relevant to both experts and novices. This use case was developed in collaboration with a company and focuses on addressing customer inquiries within technical customer support.

In designing the task, particular attention was given to ensuring that participants with varying levels of prior experience could quickly acclimate to the requirements. Additionally, the task is inherently complex that the integration of an AI Decision Support System (DSS) is considered suitable. Specifically, the partcipants tasks was the following: the participants receive emails describing different problems the customer has with a product. Their task is to select the fitting solution text from a list of eight possible texts.

The design of this task is a critical component of the study, as it ensures realism within the work context while allowing for completion under controlled and measurable conditions.

Study Design

This study aims to investigate the effects of uncertainty communication on performance, task load, and attitudes toward AI. These variables are measured both before and after the use of AI support, allowing for within-subject comparisons. Additionally, different types of uncertainty communication are examined. A between-subject design is employed to assess three distinct groups, resulting in a 2x3 factorial design.

The course of the study, as illustrated in Figure 1, begins with participants reading information about the product relevant to the emails. This information includes a product description, potential malfunctions, and corresponding response texts. To familiarize participants with the email system, a practice phase consisting of two emails is integrated.

Following this, we assess how well participants understand the functionality of the described product by including a control block at the beginning. This block consists of four emails that participants must respond to without AI support. After completing the control block, the investigator provides verbal feedback on how many emails were answered correctly. This feedback serves as a benchmark for participants to evaluate their own performance and competence regarding the task.

The subsequent AI-assisted interaction phase comprises two blocks of five emails each, with verbal feedback provided based on performance after each block.

Before entering the AI-assisted interaction phase, participants complete questionnaires related to the targeted variables. First, they fill out the NASA Task Load Index (TLX) to assess task load during their completion of tasks without AI support. Next, they complete questionnaires evaluating their attitudes toward AI and an AI literacy questionnaire. After participating in the AI-assisted interaction phase, participants repeat the questionnaires concerning task load and their attitudes toward AI.

During the AI-assisted interaction phase, in all three groups the AI ranks possible solutions according to their likelihood of addressing the problems described in each email. In two experimental groups, additional uncertainty communication is presented. In one condition, likelihoods are displayed numerically as decimal percentages (e.g., 0.63 for 63%). The second experimental condition conveys this information using arrows for symbolic uncertainty communication. Three different arrow variations represent likelihood: an upward-pointing green arrow indicates a likelihood over 0.80; a sideward-pointing arrow signifies a range between 0.79 and 0.40; and a downward-pointing red arrow denotes a likelihood below 0.39. Participants are randomly assigned to these groups.

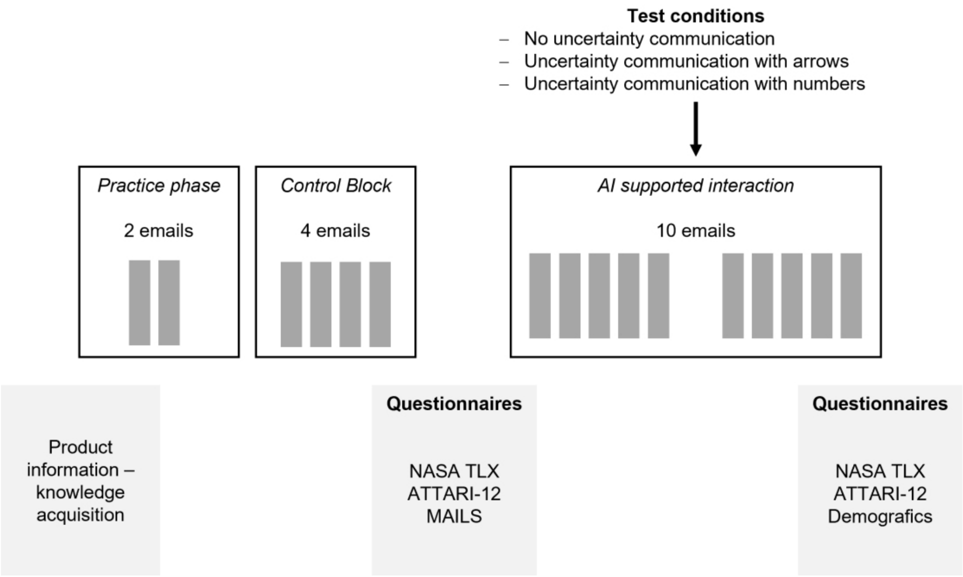

Figure 1: The course of the study displaying the different email blocks and measuring times.

DISCUSSION

The proposed study adopts a practice-oriented approach and is expected to utilize a sample composed primarily of working professionals rather than an academic population. Unlike many studies that rely on student samples or theoretical simulations, our use of a nonacademic, real-world sample increases the external validity of our findings and offers insights that may be more directly applicable to practical settings. By focusing on individuals engaged in active professional environments, we gain a better understanding of how decision-making, model interaction, and uncertainty communication might function in everyday work contexts.

However, there are also some limitations to this approach. Although the participants will be working professionals, they are not experts in the specific domain relevant to the study. Their lack of expertise may limit the generalizability of the findings to highly specialized fields, where domain knowledge and experience may play a larger role in interpreting and acting on AI-driven insights.

It is also questionable if the proposed design provides a high enough level of difficulty that reflects the complexity of tasks in real working scenarios. On the one hand, the complexity of the task had to be low enough to acquire sufficient expertise in a short amount of time. On the other hand, it must be complex enough so that integrating the support of the AI system into participants' decision-making process is sensible.

Future research could benefit from expanding the participant pool to include domain experts, which would allow for comparisons between novice and expert decision-makers in their interactions with AI models.

Despite these limitations, the findings contribute to the growing understanding of how uncertainty communication and decision-support tools are perceived and utilized by everyday users. This practice-oriented focus may provide valuable guidance for developing AI systems that are accessible and beneficial to a broader, non-specialist audience.

CONCLUSION

This paper summarizes the current state of the literature on uncertainty communication in AI-DSS and highlights the importance of practice-oriented studies to enhance our understanding of how uncertainty communication affects performance, task load, and attitudes toward AI. It identifies shortcomings in existing research concerning various forms of uncertainty communication and proposes a conceptual framework for an experimental study design. This framework may serve as a guideline for other researchers interested in conducting practice-oriented research.

ACKNOWLEDGMENT

AIXPERIMENTATIONLAB stands for Augmented Intelligence Experimentation Laboratory — the research project runs until the end of 2023 and is funded by the Federal Ministry of Labour and Social Affairs (BMAS) as part of the funding program “Future-oriented companies and administrations in the digital transformation (room for learning and experimental AI)” — EXP.01.00016.20.

REFERENCES

Bhatt, G. D., & Zaveri, J. (2002). The enabling role of decision support systems in organizational learning. *Decision Support Systems*, *32*(3), 297–309. https://doi.org/10.1016/s0167-9236(01)00120-8

Bläsing, D., & Bornewasser, M. (2021). Influence of Increasing Task Complexity and Use of Informational Assistance Systems on Mental Workload. *Brain Sciences*, *11*(1). https://doi.org/10.3390/brainsci11010102

Buch, V. H., Ahmed, I., & Maruthappu, M. (2018). Artificial intelligence in medicine: Current trends and future possibilities. *British Journal of General Practice*, *68*(668), 143–144. https://doi.org/10.3399/bjgp18X695213

Burggräf, P., Wagner, J., Koke, B., & Bamberg, M. (2020). Performance assessment methodology for AI-supported decision-making in production management. *Procedia CIRP*, *93*, 891–896. https://doi.org/10.1016/j.procir.2020.03.047

Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E., & Wienrich, C. (2023). MAILS - Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans*, *1*(2), 100014. https://doi.org/10.1016/j.chbah.2023.100014

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest : A Journal of the American Psychological Society*, *8*(2), 53–96. https://doi.org/10.1111/j.1539-6053.2008.00033.x

Goddard, K., Roudsari, A., & Wyatt, J. C. (2014). Automation bias: Empirical results assessing influencing factors. *International Journal of Medical Informatics*, *83*(5), 368–375. https://doi.org/10.1016/j.ijmedinf.2014.01.001

Hancock, P. A., Williams, G., Manning, C. M., & Miyake, S. (1995). Influence of task demand characteristics on workload and performance. *The International Journal of Aviation Psychology*, *5*(1), 63–86. https://doi.org/10.1207/s15327108ijap0501_5

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (pp. 139–183). Elsevier. https://doi.org/10.1016/s0166-4115(08)62386-9

Kasie, F. M., Bright, G., & Walker, A. (2017). Decision support systems in manufacturing: A survey and future trends. *Journal of Modelling in Management*, *12*(3), 432–454. https://doi.org/10.1108/JM2-02-2016-0015

Lai, V., Chen, C., Smith-Renner, A., Liao, Q. V., & Tan, C. (2023). Towards a Science of Human-AI Decision Making: An Overview of Design Space in Empirical Human-Subject Studies. In *2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1369–1385). ACM. https://doi.org/10.1145/3593013.3594087

Lichtenthaler, U. (2020). Extremes of acceptance: Employee attitudes toward artificial intelligence. *Journal of Business Strategy*, *41*(5), 39–45. https://doi.org/10.1108/JBS-12-2018-0204

Lukashova-Sanz, O., Dechant, M., & Wahl, S. (2023). The Influence of Disclosing the AI Potential Error to the User on the Efficiency of User–AI Collaboration. *Applied Sciences*, *13*(6), 3572. https://doi.org/10.3390/app13063572

Ma, S., Wang, X., Lei, Y., Shi, C., Yin, M., & Ma, X. (2024). "Are You Really Sure?" Understanding the Effects of Human Self-Confidence Calibration in AI-Assisted Decision Making. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–20). ACM. https://doi.org/10.1145/3613904.3642671

Mahmud, H., Islam, A. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, *175*, 121390. https://doi.org/10.1016/j.techfore.2021.121390

Münker, S., Padrón, M., Markus, A., Kemmerling, M., Abdelrazeq, A., & Schmitt, R. H. A Human-Centered Decision Support System in Customer Support. In (pp. 3–10). https://doi.org/10.1007/978-3-031-71304-0_1

Nagar, Y., & Malone, T. W. (2011). Making Business Predictions by Combining Human and Machine Intelligence in Prediction Markets. https://dspace.mit.edu/handle/1721.1/87989

Patel, B. N., Rosenberg, L., Willcox, G., Baltaxe, D., Lyons, M., Irvin, J., Rajpurkar, P., Amrhein, T., Gupta, R., Halabi, S., Langlotz, C., Lo, E., Mammarappallil, J., Mariano, A. J., Riley, G., Seekins, J., Shen, L., Zucker, E., & Lungren, M. (2019). Human-machine partnership with artificial intelligence for chest radiograph diagnosis. *Npj Digital Medicine*, *2*(1), 111. https://doi.org/10.1038/s41746-019-0189-7

Prabhudesai, S., Yang, L., Asthana, S., Huan, X., Liao, Q. V., & Banovic, N. (2023). Understanding Uncertainty: How Lay Decision-makers Perceive and Interpret Uncertainty in Human-AI Decision Making. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 379–396). ACM. https://doi.org/10.1145/3581641.3584033

Salimzadeh, S., He, G., & Gadiraju, U. (2024). Dealing with Uncertainty: Understanding the Impact of Prognostic Versus Diagnostic Tasks on Trust and Reliance in Human-AI Decision Making. In F. F. Mueller, P. Kyburz, J. R. Williamson, C. Sas, M. L. Wilson, P. T. Dugas, & I. Shklovski (Eds.), *ACM Digital Library, Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–17). Association for Computing Machinery. https://doi.org/10.1145/3613904.3641905

Stein, J.-P., Messingschlager, T., Gnambs, T., Hutmacher, F., & Appel, M. (2024). Attitudes towards AI: Measurement and associations with personality. *Scientific Reports*, *14*(1), 2909. https://doi.org/10.1038/s41598-024-53335-2

Technology acceptance model. (1989). https://link.springer.com/content/pdf/10.1007/978-3-030-45274-2.pdf

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour.* Advance online publication. https://doi.org/10.1038/s41562-024-02024-1

Venkatesh, Morris, & Davis (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540

Zhang, G., Raina, A., Cagan, J., & McComb, C. (2021). A cautionary tale about the impact of AI on human design teams. *Design Studies*, *72*, 100990. https://doi.org/10.1016/j.destud.2021.100990

Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020). 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.* ACM. https://doi.org/10.1145/3351095.3372852

Zhao, J., Wang, Y., Mancenido, M. V., Chiou, E. K., & Maciejewski, R. (2024). Evaluating the Impact of Uncertainty Visualization on Model Reliance. *IEEE Transactions on Visualization and Computer Graphics*, *30*(7), 4093–4107. https://doi.org/10.1109/tvcg.2023.3251950