# Human Factors in XAI: Enhancing User Reliance Through Emotional Alignment in Decision-Making

Der Fakultät für Wirtschaftswissenschaften der
Universität Paderborn
zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
- Doctor rerum politicarum -
vorgelegte Dissertation
von
Olesja Lammert, M.Sc.
geboren am 12.02.1994 in Arykbalyk

Erscheinungsjahr 2025

## Acknowledgments

The last four years have been a meaningful chapter in my life, filled with valuable experiences and thoughtful exchanges. I'm deeply thankful to everyone who supported me throughout this period. First, I'd like to warmly thank my doctoral advisor, Professor Dr. Kirsten Thommes, for her steady support and inspiring guidance throughout this journey. I truly appreciate the time Kirsten took for our discussions and I learned something valuable from each one. Her caring mentorship and great enthusiasm for research have consistently motivated and supported me throughout these years, for which I am deeply grateful. A special thanks also goes to the chair team for making this time such an enriching experience. I'm especially thankful to Jarek, Miro, and Conny, who were always there for me and whose support meant a lot throughout this process. I'd also like to thank Anastasia and Jörg, who played a valuable and supportive role in the early phase of my research. A big thank you to all my colleagues in the Transregional Collaborative Research Center 318 "Constructing Explainability." Working and growing alongside you was truly motivating. I'm also grateful to the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for their financial support, without which this work would not have been possible. I'd like to thank Professor Dr. Matthias Trier, Professor Dr. Martin Schneider, and Professor Dr. Lena Steinhoff for their time and support in reviewing and discussing this dissertation. Last but not least, I want to express my deepest gratitude to my family and friends for their support and encouragement throughout this time. A warm thank you to each and every one of you. I truly couldn't have done it without you.
Thank you all!

# Table of contents

## List of papers

**Paper 1**

Schütze, Christian*; Lammert, Olesja*; Richter, Birte; Thommes, Kirsten; Wrede, Britta (2023). Emotional debiasing explanations for decisions in HCI

**Paper 2**

Lebedeva, Anastasia*; Kornowicz, Jaroslaw*; Lammert, Olesja*; Papenkordt, Jörg* (2023). The role of response time for algorithm aversion in fast and slow thinking tasks

**Paper 3**

Booshehri, Meisam*; Buschmeier, Hendrik*; Cimiano, Philipp*; Kopp, Stefan*; Kornowicz, Jaroslaw*; Lammert, Olesja*; Matarese, Marco*; Mindlin, Dimitry*; Robrecht, Amelie Sophie*; Vollmer, Anna-Lisa*; Wagner, Petra*; Wrede, Britta* (2024). Towards a computational architecture for co-constructive explainable systems

**Paper 4**

Lammert, Olesja; Richter, Birte; Schütze, Christian; Thommes, Kirsten; Wrede, Britta (2024). Humans in XAI: increased reliance in decision-making under uncertainty by using explanation strategies

**Paper 5**

Thommes, Kirsten*; Lammert, Olesja*; Schütze, Christian; Richter, Birte; Wrede, Britta (2024). Human emotions in AI explanations

**Paper 6**

Lammert, Olesja (forthcoming). Can AI regulate your emotions? An empirical investigation of the influence of AI explanations and emotion regulation on human decision-making factors

---

*Note:*

* Authors contributed equally to the work.

## List of figures

### Synopsis



### Paper 1



### Paper 2



### Paper 3



### Paper 4

## List of tables

**Paper 5**



**Paper 6**

## List of abbreviations

| | |
|---|---|
| DSS | Decision-support system |
| EE | Explainee |
| ER | Explainer |
| HCI | Human-computer interaction |
| HHI | Human-human interaction |
| HRI | Human-roboter interaction |
| HTI | Human-technology interaction |
| M | Mean |
| Mdn | Median |
| SD | Standard deviation |

# Part I

# Synopsis

# 1 Introduction

Decision-making often results in outcomes that are perceived as unsatisfactory by the decision-maker. This dissatisfaction might stem from challenges in comprehensively assessing all relevant factors or fully considering the potential consequences in the decision-making process (Kahneman, 2017). In addition, emotional influences and cognitive biases can impair human judgment and increase the likelihood of outcomes that deviate from the individual's expectations (Kensinger, 2009; Lerner et al., 2007, 2015).

An established approach to simplifying and optimizing decision-making processes is the use of decision support systems (DSS) (Cao et al., 2023). Many of them are now AI-driven, such as in healthcare (Bussone et al., 2015; Wang et al., 2019), finance (Binns et al., 2018; Green and Chen, 2019), and cybersecurity (Capuano et al., 2022; Holder and Wang, 2021), where decisions often have significant consequences. To make it easier for users to understand why the DSS has provided a particular recommendation, eXplainable Artificial Intelligence (XAI) has become a major focus of research within this context (Alufaisan et al., 2021; Chazette and Schneider, 2020; Zhang et al., 2020). Its goal is to enable AI systems to explain their outputs in a way that is understandable to humans (Adadi and Berrada, 2018; Miller, 2019). XAI addresses one of the central challenges in contemporary AI systems: the lack of transparency in many algorithms (Miller, 2019). These systems often cannot clearly explain how they have arrived at their recommendations. XAI attempts to overcome this limitation by providing explanations that encourage users to make well-informed and independent decisions using the system's insights (Adadi and Berrada, 2018). In summary, XAI aims to make the decision-making processes of AI models more transparent, provide relevant background information and present results in a clear and understandable way (Adadi and Berrada, 2018; Miller, 2019).

Although XAI is considered a promising approach to helping users understand AI recommendations, its capabilities are still evolving and its practical implementation faces numerous challenges (Laato et al., 2022). Research findings indicate that not every explanation leads to greater acceptance (Schemmer et al., 2022), highlighting the need to understand how users actually process these explanations. To date, XAI has been developed primarily for a technical audience, which means that the explanations provided are often only comprehensible to data scientists and machine learning experts (Bauer et al., 2023; Neerincx et al., 2018; Ren et al., 2016; Riedl, 2019; Weitz et al., 2019). Another problem is that users do not always interpret the explanations as intended (Ehsan et al., 2024). Their conclusions sometimes deviate from logical or evidence-based considerations. It is therefore essential that explanations are tailored to the respective target groups to ensure comprehensibility and effectiveness (Bronner, 2006). In particular, as of today there has been limited emphasis on addressing the needs and level of understanding of end users (Laato et al., 2022; Riedl, 2019; Weitz et al., 2019).

In response to this and with the aim of strengthening acceptance and trust in AI-based DSS, research has increasingly recognized a user-centered design of XAI as essential (Cau et al., 2023; Wang et al., 2019). Human-centered XAI prioritizes the inclusion of human factors by considering human perception mechanisms. Moreover, the goal is to assess what information is understandable and useful to human users so that it can be effectively incorporated into system explanations. In essence, it involves determining the content, timing, and format of explanations that best assist human users (Schoonderwoerd et al., 2021).

As part of this human-centered shift in XAI, emotions are becoming increasingly important for understanding user involvement and decision-making, because they determine how individuals perceive, interpret and evaluate information (Lerner et al., 2015; Schwarz, 2000). In particular, when making decisions under risk or uncertainty, individuals do not act purely rationally by choosing the most optimal option based solely on their needs or preferences, but instead are also influenced by their emotions (Angie et al., 2011; Croskerry et al., 2013). Overlooking emotions in human-technology interaction (HTI) can have considerable repercussions for human users, with these effects falling into two main categories: 1. human cognitive processes and decision-making, and 2. users' overall attitudes toward AI-based DSS.

1. Regarding the first category, research indicates that emotions have an influence on **human cognitive processes and decision-making** and therefore can impair users' ability to engage with and understand AI-generated explanations (Bertrand et al., 2022; Guerdan et al., 2021). This may hinder rational consideration and thereby increase the risk of errors and encourage impulsive, biased or otherwise suboptimal decisions.

2. The second category emphasizes the **users' overall attitudes** toward AI-based DSS. Findings show that emotional reactions such as frustration or anger can reduce users' trust and satisfaction with AI systems (Irfan et al., 2018; Lerner and Tiedens, 2006). Negative experiences with DSS that are caused by emotional influences can have a lasting negative impact on the general acceptance and foster prejudice toward such technologies. Research already indicates that algorithmic decisions are often perceived as less fair and trustworthy than human decisions (Lee, 2018). Prejudice and a negative attitude toward AI mitigates its potential (Jussupow et al., 2020; Mahmud et al., 2022; Turel and Kalhan, 2023). This is particularly problematic in the long term, as it may ultimately lead users to reject AI recommendations, even

though the DSS provides clear and helpful explanations.

Relevant emotional influences can arise independently of the HTI, for instance as pre-existing emotional states before entering an interaction, which may influence the experience and behavior as a carry-over effect (Lerner and Keltner, 2000). However, emotions that emerge during the interaction itself also pose a relevant problem. In fact, research shows that algorithmic systems are capable of evoking negative emotions in users (Chaminade et al., 2010; Rosenthal-von der Pütten et al., 2013). More specifically, negative emotions such as dissatisfaction or anger may be triggered during an HTI, often due to the user's perception of inadequate system performance (Hornung and Smolnik, 2022).

Given the complexity and ongoing challenges in this area, a particular need for research that systematically integrates emotional and cognitive components into the development of XAI systems becomes apparent. Research suggests that incorporating emotional aspects into HTI outperforms traditional recommendation methods (Polignano et al., 2021) and enhances trust and user acceptance (Jeon, 2024). In light of this, particularly the influence of emotions on user behavior in HTI contexts remains insufficiently explored, highlighting a critical need for further investigation. To advance the development of emotion-aware systems that understand the role of emotions in decision-making, implement mechanisms to mitigate emotional influence, and generate contextually appropriate explanations to users' emotional states, the decision-making of emotionally affected individuals in the context of HTI should be examined as a first step. Therefore, this dissertation sets out to examine human cognitive processes and decision-making as well as the attitude toward AI-based DSS, with particular attention to the interplay between emotions and cognition. In line with this objective, the topic is tackled across six research articles.

This dissertation adopts two complementary approaches. The first draws on behavioral experiments to shed light on how individuals perceive and process information, and how they make decisions in an HTI (Paper 1 – 2 and Paper 4 – 6). The second approach, presented in Paper 3, introduces a conceptual framework that systematically embeds human-centered and emotional factors into the technical design. This framework extends traditional models to better account for user-related considerations.

By transferring well-founded theoretical concepts from psychology into the context of HTI and empirically analyzing their impact, this dissertation seeks to increase awareness among researchers and developers about the value of user-centered design while providing practical insights for incorporating emotion-related factors into the development of AI-based systems.

Figure 1 provides a structured overview of the six scientific articles included in this dissertation, illustrates their thematic development, and highlights the most important areas of investigation.

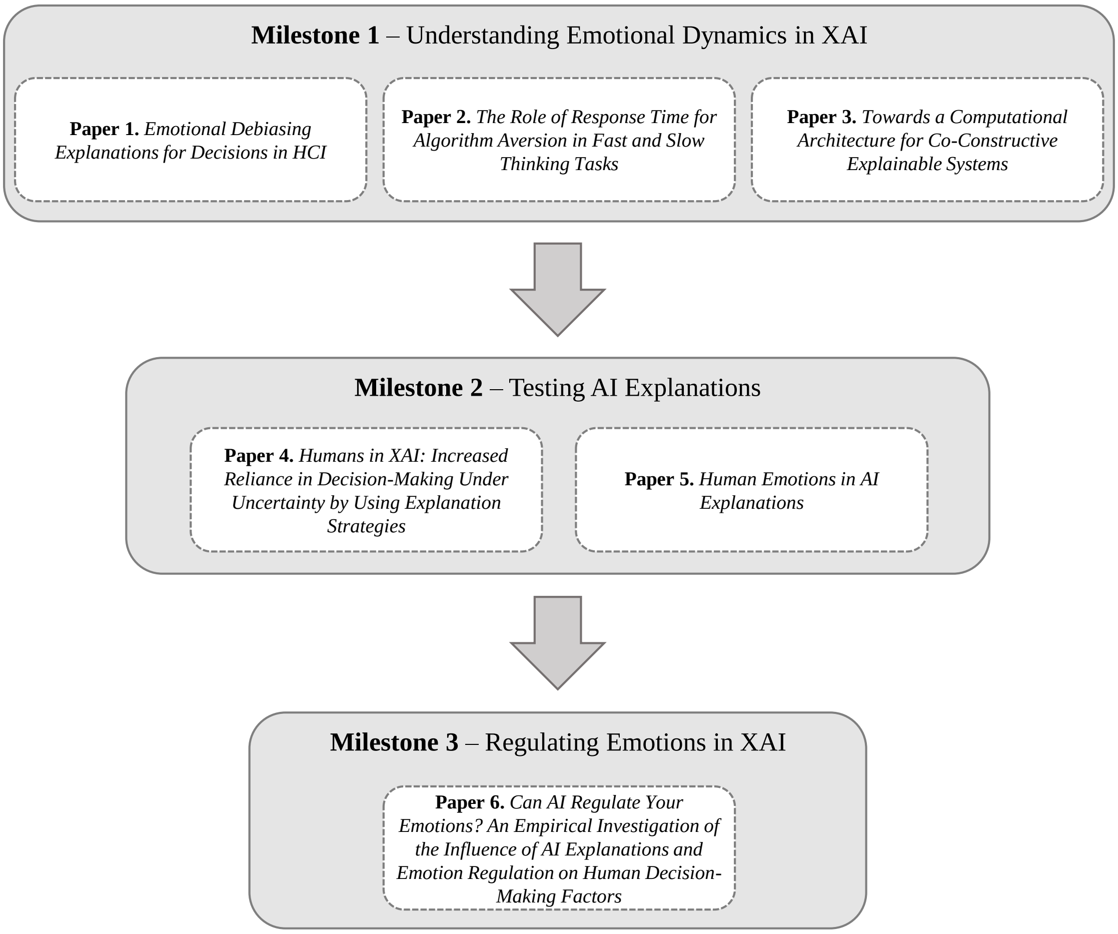


Figure 1: Research milestones and systematic overview of the papers

The main contents of each milestone are briefly summarized in the following, highlighting the structure from broader theoretical, empirical, and technical foundations to more specific analytical perspectives:

- **Milestone 1**: Paper 1 establishes a foundational understanding of the influence of emotional factors on human decision-making in HTI. Given the close relationship between emotions and cognition, Paper 2 turns to the cognitive processes involved in HTI. Paper 3 complements these behavioral insights with a technical synthesis that incorporates user-centered approaches. Considered together, Papers 1 to 3 contribute to a more comprehensive view of human decision-making behavior within HTI contexts after receiving recommendations, explanations or a combination of both. This forms a solid foundation for the subsequent studies involving cognitive and emotional components.

- **Milestone 2**: Papers 4 to 5 expand on this work by empirically testing four explanation strategies in experimental settings, with a focus on cognitive and emotional factors.

- **Milestone 3**: Paper 6 follows up on findings from Papers 4 and 5 and empirically investigates explanation strategies in emotionally charged contexts, particularly those involving anger. In addition, it introduces a supplementary approach that seeks to regulate emotional responses through an emotion regulation nudge.

The following Chapter 2, “Content of the research papers,” provides an overview of the most important findings and contributions of the individual papers, and explains their interrelationships. After each paper is presented, a corresponding table is provided containing information on the relevant conference presentations and the publication status.

# 2 Content of the research papers

Following the introductory chapter, which established the overarching structure of this dissertation and clarified the interrelationships among the individual papers, this chapter explores the three central milestones depicted in Figure 1 in greater detail. It begins by defining the objective and thematic focus of the first milestone, providing the necessary theoretical foundation for understanding the contributions of Papers 1 to 3. Throughout this chapter, each paper is presented with respect to its research aim, methodology, and key findings. Building on the insights gained, the transition to the second milestone is outlined, and the influence of the initial findings on the evolution of the subsequent research focus is made clear. Drawing from the first milestone's results, the presentation of Papers 4 and 5 contributes to a deeper understanding of existing scholarly work and elaborates on the empirical analysis that characterizes the second milestone. The progression to the third milestone is described thereafter, emphasizing the cumulative development of the research agenda. This is followed by an overview of Paper 6, which extends the analytical scope into new thematic directions. The last chapter concludes with a summary of the most important findings from all papers, evaluating the extent to which the initial research objectives have been achieved. In addition, open questions are identified and suggestions for further research and practice are made based on the results of the contributions presented.

(1) *Schütze, C., Lammert, O., Richter, B., Thommes, K., Wrede, B. (2023)*
*Emotional debiasing explanations for decisions in HCI*
To contribute to the research of AI systems that incorporate both the cognitive and emotional components of human users, **Paper 1** aims to develop a technical setup for explaining strategies through a DSS within an HTI. The DSS is intended to provide recommendations and explanations to human users, thereby assisting them in their decision-making. The reaction of emotional users after receiving such a recommendation and an explanation by a DSS is of central interest, particularly regarding whether they will subsequently rely on the recommendation. In addition, the paper seeks to assess how emotion induction and emotion recognition methods can be applied and evaluated in HTI.

In order to address this, two consecutive experiments were conducted under laboratory conditions. In the first study, participants were randomly assigned to either a "happiness" group or an "anxiety" group. A 5-minute autobiographical recall phase was used as an emotion induction procedure to elicit the respective emotional states with the aim of investigating how emotional states influence decision-making under risk and uncertainty. The emotional states before and after induction were compared using Wilcoxon rank-sum tests. The results reveal that the emotion induction successfully evoked the respective emotions. Furthermore, the findings provide evidence of biased decision-making in the given context.

Building on this finding, Study 2 replicated the experimental design of Study 1. In addition, it focused on how emotional states can be monitored and how interaction with a DSS influences both emotions and subsequent decision-making behavior. A Wizard-of-Oz approach was used to simulate the behavior of the DSS. Physiological data were collected via a smartwatch to measure heart rate, and video data were recorded with a webcam to capture the participants' facial expressions. An emotion recognition framework was then applied to evaluate its performance within the specified experimental setup. The results demonstrate again that the emotion induction procedure successfully elicited the respective emotions in Study 2 as well. Furthermore, the findings indicate that participants in the "anxiety" treatment group have a higher delta heart rate than participants in the "happiness" group. These differences could not be detected using the emotion recognition framework. In terms of behavior, the observed data show that only six participants changed their decision after receiving the DSS explanation.

Overall, the results show that emotion induction procedures are effective for experimentally manipulating emotions, which is a prerequisite for investigating emotional effects in HTI. In addition,the findings support the empirical evidence that users' actions are influenced not only by factual information but also by their emotional state, and that emotions in HTI can lead to biased decision-making. It also aligns with existing literature which suggests that AI explanations do not always offer added value compared to no explanation at all, in terms of effectiveness (Schemmer et al., 2022). To design DSS that are capable of helping individuals understand and make informed decisions, further human-centered research is needed to identify and evaluate new explanation strategies. This research will contribute to a deeper understanding of how to communicate complex information effectively through DSS. In particular, the results open up several directions for further research, including the questions: *What constitutes a good explanation from a human perspective? Which explanation strategies are appropriate for emotional users?*

Table 1: Scientific conference presentations and publication status – **Paper 1** "Emotional debiasing explanations for decisions in HCI"

| **Presentation at scientific conferences** | **Speaker** |
|---|---|
| September 21, 2022 – Data Society. Opportunity – Responsibility – Innovation, Paderborn | Olesja Lammert |
| September 26, 2022 – Faculty Workshop of Paderborn University, Melle | Olesja Lammert |
| July 28, 2023 – HCI International 2023 (HCII 2023), the 25th International Conference on Human-Computer Interaction and affiliated Conferences, Copenhagen, Denmark | Christian Schütze<br>Olesja Lammert |
| **Scientific journal or conference publication** | **Status** |
| Schütze, C., Lammert, O., Richter, B., Thommes, K., & Wrede, B. (2023). Emotional debiasing explanations for decisions in HCI. In International Conference on Human-Computer Interaction (pp. 318-336). Cham: Springer Nature Switzerland. | Published<br>(July 9, 2023) |

(2) *Lebedeva, A., Kornowicz, J., Lammert, O., Papenkordt, J. (2023)*
*The role of response time for algorithm aversion in fast and slow thinking tasks*
**Paper 2** explores how cognitive processing shapes the attitude toward AI systems. Applying the Dual Process Theory (Kahneman, 2017) as proposed by Bonnefon and Rahwan (2020) within the context of HTI, this paper investigates how variations in AI response time influence user reliance[1] across tasks with different cognitive demands and domains. According to Kahneman, human cognition operates through two systems: System 1, which is fast, instinctive, emotional, and System 2, which is slower, analytical, and logical (Kahneman, 2017). Within the scope of this paper, tasks that can be solved through intuitive, fast thinking are referred to as "fast-thinking tasks." Those that require analytical, slow thinking are designated as "slow-thinking tasks." Unlike interactions between humans, users tend to view longer response times from AI negatively when engaging in analytical tasks (Efendić et al., 2020). This perception likely stems from the assumption that such tasks are considered simple for machines and should therefore be delivered immediately. Longer response times can be interpreted as a sign of system failure or inefficiency in slow-thinking tasks in this context. In contrast, from a human user perspective,

[1]While the focus of this paper lies on algorithm aversion, the term algorithm acceptance, more specifically **user reliance**, is used to ensure consistency with the other papers within this dissertation. In this context, shifting the focus from algorithm aversion to user reliance offers a conceptually meaningful perspective, as the underlying operationalization remains consistent across the papers and merely reflects the opposing perspective (Jussupow et al., 2020).

longer AI response times in fast-thinking tasks may be perceived as necessary due to the task's complexity for non-human agents, which could foster greater user reliance. The fast-thinking of tasks are commonly viewed as human strengths and therefore perceived as more difficult for AI systems to handle quickly.

Paper 2 addresses the question of the effect AI response time has on user reliance for slow-thinking and fast-thinking tasks. To better understand this, a 2×2 between-subjects design was used in a controlled laboratory experiment. Participants were randomly assigned to one of the four groups. In addition, three task domains were employed. The results show that long response times are associated with higher user reliance in both fast- and slow-thinking tasks. Furthermore, significant differences were found in user reliance between task domains in fast-thinking tasks, whereas in slow-thinking tasks, the task domain does not significantly influence user reliance.

The findings of this study contribute to the existing literature on algorithm appreciation and algorithm aversion by applying the Dual Process Theory in an HTI experiment. Taken together, it may seem reasonable to expect opposite response patterns in HTI compared to HHI, particularly in tasks involving fast- and slow-thinking. However, the findings indicate that user expectations are not simply reversed. Instead, they reflect complex, context-dependent perceptual mechanisms. These expectations often do not align with users' actual behavior. The results offer practical insights for optimizing AI response time to enhance user reliance. Additionally, by examining three distinct task domains, the study offers a deeper understanding of how user reliance emerges across varying cognitive demands. It emphasizes that both task domain and response time play a significant role in shaping user reliance in fast- and slow-thinking tasks. This has direct implications for the development of explanation strategies in HTI, raising the question of *how cognitive factors can be systematically integrated to ensure explanations that are both understandable and cognitively efficient, thereby supporting informed user decisions.*

Table 2: Scientific conference presentations and publication status – **Paper 2** "The role of response time for algorithm aversion in fast and slow thinking tasks"

| **Presentation at scientific conferences** | **Speaker** |
|---|---|
| July 28, 2023 – HCI International 2023 (HCII 2023), the 25th International Conference on Human-Computer Interaction and affiliated Conferences, Copenhagen, Denmark | Anastasia Lebedeva |
| **Scientific journal or conference publication** | **Status** |
| Lebedeva, A., Kornowicz, J., Lammert, O., & Papenkordt, J. (2023). The role of response time for algorithm aversion in fast and slow thinking tasks. In International Conference on Human-Computer Interaction (pp. 131-149). Cham: Springer Nature Switzerland. | Published (July 9, 2023) |

(3) *Booshehri, M., Buschmeier, H., Cimiano, P., Kopp, S., Kornowicz, J., Lammert, O., Matarese, M., Mindlin, D., Robrecht, A. S., Vollmer, A.-L., Wagner, P., Wrede, B. (2024)*
*Towards a computational architecture for co-constructive explainable systems*
**Paper 3** addresses the question of how a computational architecture can foster a paradigm shift in XAI, from viewing explanations as static outcomes to conceiving them as interactive processes that position the user as an active participant.

To support this shift, the paper proposes an architecture for XAI systems that centers on a co-constructive approach to explanation generation (Rohlfing et al., 2021). Within this architecture, the explanation recipient is actively involved in a bidirectional and iterative process of jointly constructing the explanation content, rather than being a passive recipient of preformulated explanations. Two components are central to this architecture: 1. Monitoring the user's understanding, for instance by capturing signals related to the user's cognitive or emotional state. 2. The use of scaffolding techniques to bridge knowledge gaps and dynamically adapt explanations to the user's needs and feedback, for example by rephrasing system statements or sending emotional signals to the user. Both processes are aligned with the MAPE-K model (IBM, 2006), an established framework in the field of autonomous computing, and are specifically adapted for interactive and co-constructive XAI.

While the architecture is presented primarily as a descriptive proposal, it nevertheless demonstrates how the integration of monitoring and scaffolding modules can support a continuous user-system adaption loop, which remains rare in current XAI research. It contributes to the advancement of XAI frameworks by placing an emphasis on user-centered approaches that support active participation in the creation and interpretation of explanations. As a result, the developed framework raises the question of *how user needs can be systematically incorporated into the design of explanations*. To address this question, experimental methods provide a clear direction for further investigation.

Table 3: Scientific conference presentations and publication status – **Paper 3** "Towards a computational architecture for co-constructive explainable systems"

| **Presentation at scientific conferences** | **Speaker** |
|---|---|
| April 20, 2024 – ExEn '24: 2024 Workshop on Explainability Engineering, Lisbon, Portugal | Dimitry Mindlin |
| **Scientific journal or conference publication** | **Status** |
| Booshehri, M., Buschmeier, H., Cimiano, P., Kopp, S., Kornowicz, J., Lammert, O., Matarese, M., Mindlin, D., Robrecht, A. S., Vollmer, A., Wagner, P. & Wrede, B. (2024). Towards a computational architecture for co-constructive explainable systems. In Proceedings of the 2024 Workshop on Explainability Engineering (pp. 20-25). | Published (July 25, 2024) |

To summarize, **Milestone 1** analyzed the relevance and interaction of cognitive and emotional influences in HTI. By linking theoretical, empirical, and technical insights, the findings demonstrate that cognitive factors and emotions play a crucial role in AI-assisted decision-making. Additionally, they show that user expectations toward AI differ from those directed at human counterparts. These findings emphasize the importance of more user-centered designs in XAI, particularly those that account for emotional and cognitive aspects of interaction. They also raise a further question about how explanations should be designed to effectively support users in AI-assisted decision-making. Building on this foundation, **Papers 4 and 5** experimentally examine diverse explanation strategies within XAI. These studies extend earlier insights and contribute to a deeper understanding of adaptive explanation systems.

(4) *Lammert, O., Richter, B., Schütze, C., Thommes, K., Wrede, B. (2024) Humans in XAI: increased reliance in decision-making under uncertainty by using explanation strategies*
While recent studies have highlighted the relevance of explanation styles in XAI (Cau et al., 2023; Gedikli et al., 2014; Shulner-Tal et al., 2025), there remains a lack of empirical evidence on how specific designs affect user behavior. **Paper 4** contributes to the research field of human-centered XAI by designing and

developing explanation strategies as well as examining which ones are perceived by users as appropriate and effective and how these perceptions impact user reliance and trust.

To achieve this, the study employed a randomized 5x1 between-subjects design in an online experiment conducted via Prolific. As part of the experiment, participants engaged in a decision-making task in which the DSS provided a recommendation and explanation, which varied depending on their treatment group assignment. Figure 2 presents the chronology of events during the decision-making process.

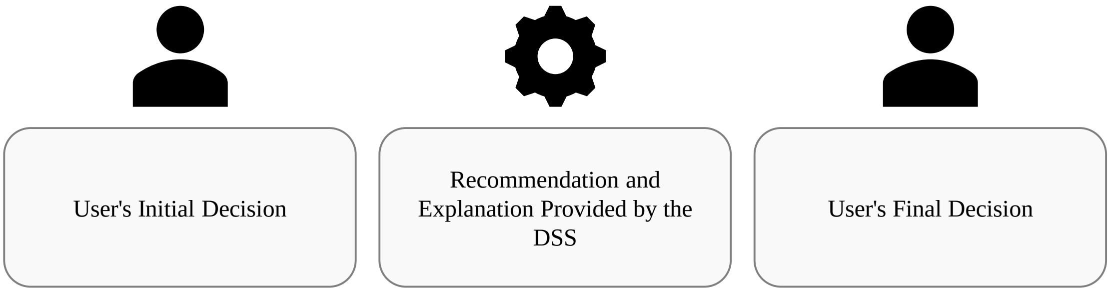


Figure 2: Sequence structure of the decision-making process

Besides the control group, in which no explanation of the DSS was provided, four explanation strategies were developed based on interdisciplinary insights from both human-human interaction (HHI) and HTI research (Amershi et al., 2019; Rosenthal-von der Pütten et al., 2013; Schoonderwoerd et al., 2021; Schütze et al., 2023). The two HTI-based strategies focus on accuracy by presenting the features that led to the recommendation. The number of displayed features was varied: either only the top three features were shown (*cognitive*[2] *explanation*), or all contributing features were displayed (*transparent explanation*). In contrast, strategies inspired by HHI were designed to resemble the way people naturally explain decisions to each other. Unlike the HTI-based strategies, they do not display any underlying features. One takes an authoritative tone (*authoritarian explanation*), while the other delivers an emotion-based explanation (*emotional explanation*), encouraging the user to reconsider their decision due to possible emotional factors that may have influenced their decision. For a more structured

[2]Initially referred to as "guided" in Papers 4 and 5, this explanation strategy is now more accurately termed "cognitive," highlighting its core aim of minimizing cognitive overload through selective feature presentation. The adoption of this terminology reflects both a theoretical refinement and increasing conceptual alignment within the field.

overview, a summary of all strategies and their areas of origin is presented in Table 4.

Table 4: Overview of the explanation strategies

| **Explanation Strategy** | **Description** | **Field of Origin** |
|---|---|---|
| Transparent | Displays all contributing features that led to the recommendation | HCI |
| Cognitive | Displays top three features that led to the recommendation | HCI |
| Emotional | Displays no feature details, frames recommendation with reference to emotional influences on user decision-making | HHI |
| Authoritarian | Displays no feature details, presents recommendation in a prescriptive tone | HHI |
| Control | Baseline — no explanation provided | – |

The results of the Mann-Whitney U test show that the *cognitive* explanation leads to significantly higher user reliance than the *transparent* explanation. In addition, the probit regression analysis further reveals that the likelihood of participants relying on the DSS is significantly higher under the *cognitive* explanation condition than under the *control* condition. An additional key finding is that participants in the *control* group who followed the DSS recommendation report significantly higher levels of subjective trust than participants in the *cognitive*, *emotional*, and *authoritarian* treatment groups.

Building on prior research, this study designs and develops explanation strategies by drawing on insights from HHI as well as HTI. This study contributes to the ongoing research in XAI by empirically demonstrating that the selective presentation of information through a *cognitive* explanation, rather than full *transparency*, can positively influence user behavior. This approach helps facilitate cognitive processes and reduce cognitive overload, which is particularly relevant in real-world scenarios involving high information density.

In addition, the results indicate that subjectively perceived trust in the AI system does not necessarily result in higher user reliance, pointing to a notable discrepancy between perception and behavior that should be carefully considered in future research. The results of this study underscore the importance of continued research on this topic, particularly into *how these explanation strategies are perceived under emotional conditions during decision-making and to what extent they can regulate or reduce emotional responses in order to mitigate biased decision-making.*

Table 5: Scientific conference presentations and publication status – **Paper 4** “Humans in XAI: increased reliance in decision-making under uncertainty by using explanation strategies”

| **Presentation at scientific conferences** | **Speaker** |
|---|---|
| November 06, 2023 – 2nd TRR 318 Constructing Explainability Conference Measuring Understanding, Paderborn | Olesja Lammert |
| November 09, 2023 – Faculty Workshop of Paderborn University, Paderborn | Olesja Lammert |
| **Scientific journal or conference publication** | **Status** |
| Lammert, O., Richter, B., Schütze, C., Thommes, K., & Wrede, B. (2024). Humans in XAI: increased reliance in decision-making under uncertainty by using explanation strategies. Frontiers in Behavioral Economics, 3, 1377075. | Published (March 8, 2024) |

(5) *Thommes, K., Lammert, O., Schütze, C., Richter, B., Wrede, B. (2024) Human emotions in AI explanations*

While Paper 4 has demonstrated that user reliance varies by explanation type, **Paper 5** extends this investigation by incorporating emotional dimensions. Specifically, it examines whether the explanation strategies outlined in Paper 4 are perceived differentially by users depending on their emotional state, with particular focus on emotional valence and arousal.

Thus, a similar experimental structure to that described in Paper 4 was used, involving five treatment groups, a decision-making task, and a DSS providing both a recommendation and an explanation. In Paper 5, however, participants were divided into low- and high-arousal groups using a median split in order to assess users’ responses to explanation strategies. In addition, a corresponding analysis was applied to emotional valence, whereby participants were divided into groups exhibiting either positive or negative valence.

The results of independent-samples t-tests indicated that emotionally aroused participants showed greater user reliance when receiving a *cognitive* or an *emotional* explanation compared to no explanation. In contrast, participants with lower arousal levels preferred no explanation over a *cognitive* one. Controlling for relevant variables, subsequent multiple linear regression analyses supported these findings and identified significant interaction effects on user reliance between arousal and the *cognitive* explanation strategy, as well as between arousal and the emotional explanation strategy. The results did not indicate any consistent effects with regard to varying levels of emotional valence.

These findings point to the importance of recognizing that users engage with and respond to explanations differently when they are emotionally aroused. Specifically, they underscore the need to adapt explanation strategies to better align with users' emotional states and their associated information processing needs. These findings give rise to a subsequent research question: *how effective are explanation strategies in emotionally charged contexts that involve strong negative emotions?* This question is particularly relevant because such emotional states frequently occur in HTI (Lee, 2018; Phillips and Madhavan, 2013), and can have potentially long-term negative consequences, such as reduced user reliance or diminished trust. A further point of consideration is to investigate *which additional approaches may be employed in XAI to mitigate the impact of negative emotional states.*

Table 6: Scientific conference presentations and publication status – **Paper 5** "Human emotions in AI explanations"

| **Presentation at scientific conferences** | **Speaker** |
|---|---|
| July 17, 2024 – xAI-2024: The 2nd World Conference on eXplainable Artificial Intelligence, Valletta, Malta | Olesja Lammert |
| **Scientific journal or conference publication** | **Status** |
| Thommes, K., Lammert, O., Schütze, C., Richter, B., & Wrede, B. (2024). Human emotions in AI explanations. In World Conference on Explainable Artificial Intelligence (pp. 270-293). Cham: Springer Nature Switzerland. | Published (July 10, 2024) |

Taken together, **Papers 4 and 5** were essential to achieving **Milestone 2**, which advances the understanding of how such strategies are received and processed by humans. These strategies were designed, developed, and evaluated in emotionally neutral experimental settings. The process was driven by the need to create human-centered explanations, drawing on research from human-machine and human-human interaction. The resulting approach emphasizes the importance of aligning explanations with natural communication patterns and users' cognitive mechanisms. In the empirical studies conducted, participants show greater user reliance on the system when presented with the *cognitive* explanation (Paper 4), and with either the *cognitive* or the *emotional* explanation (Paper 5), particularly when users are emotionally aroused.

Building on these results and addressing remaining open questions, the next step in this dissertation aims to investigate explanation strategies in emotionally charged contexts characterized by a specific high-arousal negative emotion, and to examine whether such strategies can help mitigate users' emotional responses. In addition, examining user reliance and self-reported trust is expected to yield a more comprehensive understanding of how explanation strategies function under emotional impact.

(6) *Lammert, O. (forthcoming)*
*Can AI regulate your emotions? An empirical investigation of the influence of AI explanations and emotion regulation on human decision-making factors*
Research shows that anger frequently arises in HTI and may impair essential outcomes, including trust, understanding, and the effectiveness of decision-making. **Paper 6** aims to foster a more comprehensive understanding of the decision-making processes of emotional users. Specifically, it examines how different factors affect angry users' trust and reliance, as well as the potential of these treatments to mitigate anger.

A randomized online experiment, employing a 2×3 between-subjects design, was conducted via Prolific to investigate the influence of an emotion regulation nudge (nudge, no nudge) and an explanation (cognitive, emotional, control). Participants assigned to the nudge condition received instructions to maintain neutrality and were explicitly encouraged to approach the following experiment objectively and analytically. The nudge is intended to reduce strong negative emotional reactions in order to support users in processing explanations and improve their acceptance. In contrast, participants in the no nudge condition proceeded without any instruction. Subsequently, all participants unterwent an anger induction procedure. Afterward, they participated in the decision-making task, during which they received a recommendation and an explanation from the DSS.

The observed patterns align with the findings in Paper 5, which suggest that high-arousal emotional states enhance user reliance when *cognitive* or *emotional* explanations are provided. In line with this, angry users in the present Paper 6 exhibited greater user reliance when receiving either *cognitive* or *emotional* explanations. Furthermore, the results indicate that introducing a nudge to foster emotional awareness is associated with increased user reliance. When the nudge is combined with an explanation strategy, user reliance appears to be even higher compared to providing an explanation alone. These findings extend the relevance of explanation styles (Cau et al., 2023; Gedikli et al., 2014; Shulner-Tal et al., 2025) to highly emotionally charged contexts, such as anger, emphasizing their importance when emotional arousal is present. Furthermore, they indicate that an emotional regulation nudge can be effective in AI-assisted decision-making.

Table 7: Scientific conference presentations and publication status – **Paper 6** “Can AI regulate your emotions? An empirical investigation of the influence of AI explanations and emotion regulation on human decision-making factors”

| **Presentation at scientific conferences** | **Speaker** |
|---|---|
| March 21, 2025 – Workshop: Examining the Reflexive Constitution of Context in-(inter)action, Paderborn | Olesja Lammert |
| To be presented in July, 2025 – xAI-2025: The 3rd World Conference on eXplainable Artificial Intelligence, Istanbul, Turkey | Olesja Lammert |
| **Scientific journal or conference publication** | **Status** |
| Lammert, O. (forthcoming). Can AI regulate your emotions? An empirical investigation of the influence of AI explanations and emotion regulation on human decision-making factors. In World Conference on Explainable Artificial Intelligence. Cham: Springer Nature Switzerland. | Accepted |

The subsequent Chapter 3 “Summary and future direction” revisits the findings of the papers from a broader perspective and discusses future directions for both research and practice. Subsequently, **Part II** provides a systematic presentation of the papers underlying this dissertation, following the structure introduced in this chapter.

# 3 Summary and future direction

XAI is already deeply embedded into various aspects of daily life and will continue to shape decision-making processes, particularly in areas where decisions carry substantial ethical, social, or practical implications (Gadekallu et al., 2024; Haque et al., 2023). This dissertation advances the human-centered research perspective that integrates psychological user characteristics, emphasizing emotion and cognition as essential components for the development of adaptive and context-sensitive explanations in HTI. Drawing on the findings of the six research papers, the findings support the view that AI-generated explanations should not only be technically comprehensible but must also align with users' cognitive and emotional processing styles (Riedl, 2019; Weitz et al., 2019), which may foster more accurate evaluations and more confident use of AI recommendations. Accordingly, the following section presents the essential findings to clarify the overall contribution of this dissertation and demonstrates how it extends current research in human-centered XAI.

The insights gained in **Milestone 1** reaffirm the established importance of emotional dynamics in human decision-making (Gonzalez et al., 2007; Pfister and Böhm, 2008) and broaden their relevance to the field of XAI. Given the minimal attention emotions have received in prior XAI research (Weitz et al., 2019), the findings underscore the essential role of emotions in shaping engagement and reliance on AI recommendations, ultimately affecting decision-making processes. Notably, emotional processes have a substantial impact on how users evaluate and accept AI-generated explanations (Gonzalez et al., 2007). More specifically, emotions affect thought and guide attention. Taken together, these findings show that the consideration of emotional dynamics is fundamental to HTI within XAI, as no decision is entirely unaffected by emotional influences, regardless of how rational the context may appear (Pfister and Böhm, 2008). Therefore, advancing research in XAI requires a shift from asking whether emotions influence cognitive processes in AI-assisted decision-making to exploring how these influences unfold. As demonstrated in this dissertation, established conceptual frameworks from psychology offer a robust foundation for experimental research in the field of XAI. They enable a thorough investigation of the extent to which emotional responses modulate cognitive processing in AI-assisted decision-making.

In addition, the research findings further suggest that the acceptance and effectiveness of AI systems depends largely on the extent to which these systems are adapted to human expectations of thinking and communication. Importantly, humans appear to apply different standards and expectations when interacting with AI compared to human agents. The trustworthiness and acceptance of the AI recommendations increase when the system exhibits signs of reflective processing, which can be likened to analytical or deliberative thinking in humans. Rather than favoring fast or intuitive responses, users tend to value signs of cognitive depth and intentionality, regardless of whether the task at hand would traditionally require rapid or slow thinking (Lebedeva et al., 2023).

This observation ties into another important finding: emotional communication in HTI is not solely unidirectional from the user to the system, for example, through gestures, facial expressions, vocal tone, or physiological signals. Instead, it constitutes a bidirectional exchange. AI-systems are also capable of conveying emotional signals, for example through response time, the personalization of explanations, or emotionally nuanced language. These signals influence how users cognitively process AI-generated recommendations. As a result, it becomes clear that the mere provision of an explanation is insufficient to foster trust or encourage user reliance. Comprehensibility and transparency may be necessary prerequisites, however they alone do not engender trust. Rather, it is the perceived cognitive depth (Lebedeva et al., 2023) and emotional coherence of a system (Schütze et al., 2023) that really shape user trust. The architecture proposed in this work also addresses the need for bidirectional interaction by integrating the mechanisms of monitoring and scaffolding (Booshehri et al., 2024; Rohlfing et al., 2021). It offers a foundation for future research to detecting, interpreting, and adaptively responding to users' emotional states. The architecture is suitable for implementation across various application domains. Future experimental studies on diverse explanation strategies can provide deeper insights into how emotional factors influence user behavior and interaction with systems.

The main conclusions drawn from **Milestone 2** suggest that explanations provided by a DSS significantly influence users' reliance on and trust in the AI system (Lammert et al., 2024; Thommes et al., 2024). The findings indicate that individuals prefer concise and contextually relevant explanations rather than complete representations that may result in cognitive overload. This supports an information-psychological approach to explanation design, where only the core features of a model or decision are emphasized in order to reduce cognitive effort (You et al., 2022). Such explanations are particularly effective when users in a heightened state of arousal have limited capacity to process complex information. In addition, explanations that make users aware of the emotional influences on their decisions can direct attention and influence judgment. This underlines the importance for system developers to consider cognitive and emotional factors in the design of explanations, ensuring they are meaningfully embedded in the development of XAI systems to enhance understanding, trust, and user reliance.

Beyond explanation design itself, these findings shed light on the broader interplay between user interaction and trust formation in AI systems, contributing to ongoing debates in the field. Previous research has shown that users can rely on a system without necessarily trusting it, and that high levels of trust do not always lead to increased user reliance (Lai et al., 2023; Scharowski et al., 2022). Consistent with this complexity, the findings of this dissertation reveal a similar divergence: participants reported higher levels of perceived trust in cases where the system did not provide an explanation, yet they were more likely to rely on the system when an explanation was offered (Lammert et al., 2024). This indicates that subjective trust and user reliance are different constructs that are characterized by different aspects of system interaction. These inconsistencies suggest promising directions for future research into underreliance, overreliance, and the broader constructs of overtrust and undertrust in AI systems. Effective AI explanations and recommendations should support and improve human decision-making and not encourage uncritical acceptance. For example, in high-stakes domains such as clinical care, overreliance can lead to errors or misdiagnoses, especially when time pressure causes users to rely on it, even if they do not fully trust the system. Consequently, future research should incorporate both subjective and objective measures of trust to develop a more comprehensive understanding of trust dynamics in AI-assisted decision-making.

**Milestone 3** strengthens the view that explanations serve a crucial function in emotionally charged contexts: individuals in a state of high emotional arousal are more inclined to accept AI recommendations when explanations are offered (Thommes et al., 2024). Furthermore, the findings support previous research suggesting that emotion regulation nudges can help mitigate emotional responses in humans (Mauss et al., 2007), while also making a novel contribution by showing that such a nudge, when applied specifically within the context of XAI to evoke emotional awareness, even as a one-time intervention, can significantly increase the willingness to follow AI-assisted recommendations. Integrating these findings into practical applications can contribute to the development of human-centered XAI systems. In sensitive decision-making areas where wrong decisions can have serious consequences, such as in disaster control, air traffic control or clinical care, these insights can be applied strategically to support better outcomes. In such domains, specialist staff are trained in dealing with emotionally stressful situations. However, emotional reactions often persist or resurface later. These reactions can impair the quality of decisions. The integration of emotion regulation strategies in DSS therefore represents a promising method for minimizing the impact of emotional interference. The application of such nudges in high-stakes DSS contexts can help professionals to facilitate cognitive reframing of emotional situations, leading to more objective and informed decisions. This creates a need for further research, as different user groups of a DSS, such as end users, laypersons, developers and technical experts, may require different strategies and forms of explanations (Burnett, 2020; Ribera and Lapedriza, 2019; Schoonderwoerd et al., 2021). Future research should explicitly investigate these differences in AI-asissted decision-making, with clearly defined target groups.

As XAI continues to evolve, the integration of emotional dimensions into HTI will likely remain a key area of research and practical relevance. With the growing prominence of large language models like ChatGPT, text-based emotion recognition systems may play an increasingly important role. These systems can identify the emotional state by analyzing user input and generate personalized explanations accordingly. Such functions could enable HTI to better mimic the patterns and responsiveness found in HHI. These systems could also meaningfully influence HTI in environments where emotions are communicated via speech or video, and they may contribute to advancing multimodal explanation strategies in HTI.

Overall, emotion-aware XAI systems should be seen as powerful tools that enhance human decision-making by helping individuals make more informed and confident decisions. The development of emotion-aware systems requires a sophisticated understanding of how different emotional states and arousal levels affect users' needs and how previously studied combinations of explanations can be used to support decision-making. To improve the robustness and practical applicability of the current results of this dissertation, future research should investigate these effects under more realistic conditions, e.g. through field experiments. Such an approach would strengthen both external validity and practical relevance. Additionally, long-term studies may provide valuable insights into the effectiveness of emotion regulation strategies. Progress in this field requires more interdisciplinary collaboration that integrates insights from psychology, technology, and empirical research, with particular consideration of ethical, social and regulatory factors. This is especially important given the simultaneous challenge of designing explanations that inform users without being intrusive or controlling, and addressing broader concerns about diminished human control.

## References

Adadi, A. and Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access*, 6:52138–52160.

Alufaisan, Y., Marusich, L. R., Bakdash, J. Z., Zhou, Y., and Kantarcioglu, M. (2021). Does explainable artificial intelligence improve human decision-making? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6618–6626.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., et al. (2019). Guidelines for human-ai interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–13.

Angie, A. D., Connelly, S., Waples, E. P., and Kligyte, V. (2011). The influence of discrete emotions on judgement and decision-making: A meta-analytic review. *Cognition & Emotion*, 25(8):1393–1422.

Bauer, K., von Zahn, M., and Hinz, O. (2023). Expl (ai) ned: The impact of explainable artificial intelligence on users' information processing. *Information systems research*, 34(4):1582–1602.

Bertrand, A., Belloum, R., Eagan, J. R., and Maxwell, W. (2022). How cognitive biases affect xai-assisted decision-making: A systematic review. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, pages 78–91.

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., and Shadbolt, N. (2018). 'it's reducing a human being to a percentage' perceptions of justice in algorithmic decisions. In *Proceedings of the 2018 Chi conference on human factors in computing systems*, pages 1–14.

Bonnefon, J.-F. and Rahwan, I. (2020). Machine thinking, fast and slow. *Trends in Cognitive Sciences*, 24(12):1019–1027.

Booshehri, M., Buschmeier, H., Cimiano, P., Kopp, S., Kornowicz, J., Lammert, O., Matarese, M., Mindlin, D., Robrecht, A. S., Vollmer, A.-L., et al. (2024). Towards a computational architecture for co-constructive explainable systems. In *Proceedings of the 2024 Workshop on Explainability Engineering*, pages 20–25.

Bronner, S. J. (2006). Folk logic: Interpretation and explanation in folkloristics. *Western Folklore*, 65(4):401–433.

Burnett, M. (2020). Explaining ai: fairly? well? In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pages 1–2.

Bussone, A., Stumpf, S., and O'Sullivan, D. (2015). The role of explanations on trust and reliance in clinical decision support systems. In *2015 international conference on healthcare informatics*, pages 160–169. IEEE.

Cao, S., Gomez, C., and Huang, C.-M. (2023). How time pressure in different phases of decision-making influences human-ai collaboration. *Proceedings of the ACM on Human-computer Interaction*, 7(CSCW2):1–26.

Capuano, N., Fenza, G., Loia, V., and Stanzione, C. (2022). Explainable artificial intelligence in cybersecurity: A survey. *Ieee Access*, 10:93575–93600.

Cau, F. M., Hauptmann, H., Spano, L. D., and Tintarev, N. (2023). Effects of ai and logic-style explanations on users' decisions under different levels of uncertainty. *ACM Transactions on Interactive Intelligent Systems*, 13(4):1–42.

Chaminade, T., Zecca, M., Blakemore, S.-J., Takanishi, A., Frith, C. D., Micera, S., Dario, P., Rizzolatti, G., Gallese, V., and Umiltà, M. A. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS one*, 5(7):e11577.

Chazette, L. and Schneider, K. (2020). Explainability as a non-functional requirement: challenges and recommendations. *Requirements Engineering*, 25(4):493–514.

Croskerry, P., Singhal, G., and Mamede, S. (2013). Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ quality & safety*, 22(Suppl 2):ii58–ii64.

Efendić, E., Van de Calseyde, P. P., and Evans, A. M. (2020). Slow response times undermine trust in algorithmic (but not human) predictions. *Organizational Behavior and Human Decision Processes*, 157:103–114.

Ehsan, U., Passi, S., Liao, Q. V., Chan, L., Lee, I.-H., Muller, M., and Riedl, M. O. (2024). The who in xai: how ai background shapes perceptions of ai explanations. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–32.

Gadekallu, T. R., Maddikunta, P. K. R., Boopathy, P., Deepa, N., Chengoden, R., Victor, N., Wang, W., Wang, W., Zhu, Y., and Dev, K. (2024). Xai for industry 5.0-concepts, opportunities, challenges and future directions. *IEEE Open Journal of the Communications Society*.

Gedikli, F., Jannach, D., and Ge, M. (2014). How should i explain? a comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies*, 72(4):367–382.

Gonzalez, G., De La Rosa, J. L., Montaner, M., and Delfin, S. (2007). Embedding emotional context in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop*, pages 845–852. IEEE.

Green, B. and Chen, Y. (2019). The principles and limits of algorithm-in-the-loop decision making. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–24.

Guerdan, L., Raymond, A., and Gunes, H. (2021). Toward affective xai: facial affect analysis for understanding explainable human-ai interactions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3796–3805.

Haque, A. B., Islam, A. N., and Mikalef, P. (2023). Explainable artificial intelligence (xai) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technological Forecasting and Social Change*, 186:122120.

Holder, E. and Wang, N. (2021). Explainable artificial intelligence (xai) interactively working with humans as a junior cyber analyst. *Human-Intelligent Systems Integration*, 3(2):139–153.

Hornung, O. and Smolnik, S. (2022). Ai invading the workplace: negative emotions towards the organizational use of personal virtual assistants. *Electronic Markets*, 32(1):123–138.

IBM (2006). An architectural blueprint for autonomic computing. Technical report, IBM.

Irfan, B., Kennedy, J., Lemaignan, S., Papadopoulos, F., Senft, E., and Belpaeme, T. (2018). Social psychology and human-robot interaction: An uneasy marriage. In *Companion of the 2018 ACM/IEEE international conference on human-robot interaction*, pages 13–20.

Jeon, M. (2024). The effects of emotions on trust in human-computer interaction: A survey and prospect. *International Journal of Human–Computer Interaction*, 40(22):6864–6882.

Jussupow, E., Benbasat, I., and Heinzl, A. (2020). Why are we averse towards algorithms? a comprehensive literature review on algorithm aversion. *Proceedings of the 28th European Conference on Information Systems (ECIS)*, pages 1–16.

Kahneman, D. (2017). Thinking, fast and slow.

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emot Rev*, 2(1):99–113.

Laato, S., Tiainen, M., Najmul Islam, A., and Mäntymäki, M. (2022). How to explain ai systems to end users: a systematic literature review and research agenda. *Internet Research*, 32(7):1–31.

Lai, V., Chen, C., Smith-Renner, A., Liao, Q. V., and Tan, C. (2023). Towards a science of human-ai decision making: An overview of design space in empirical human-subject studies. In *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency*, pages 1369–1385.

Lammert, O., Richter, B., Schütze, C., Thommes, K., and Wrede, B. (2024). Humans in xai: increased reliance in decision-making under uncertainty by using explanation strategies. *Frontiers in Behavioral Economics*, 3:1377075.

Lebedeva, A., Kornowicz, J., Lammert, O., and Papenkordt, J. (2023). The role of response time for algorithm aversion in fast and slow thinking tasks. In *International Conference on Human-Computer Interaction*, pages 131–149. Springer.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big data & society*, 5(1):2053951718756684.

Lerner, J. S., Han, S., and Keltner, D. (2007). Feelings and consumer decision making: Extending the appraisal-tendency framework. *Journal of consumer psychology*, 17(3):181–187.

Lerner, J. S. and Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition & emotion*, 14(4):473–493.

Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. S. (2015). Emotion and decision making. *Annual review of psychology*, 66:799–823.

Lerner, J. S. and Tiedens, L. Z. (2006). Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition. *Journal of behavioral decision making*, 19(2):115–137.

Mahmud, H., Islam, A. N., Ahmed, S. I., and Smolander, K. (2022). What influences algorithmic decision-making? a systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, 175:121390.

Mauss, I. B., Cook, C. L., Cheng, J. Y., and Gross, J. J. (2007). Individual differences in cognitive reappraisal: Experiential and physiological responses to an anger provocation. *International Journal of Psychophysiology*, 66(2):116–124.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences.

Neerincx, M. A., van der Waa, J., Kaptein, F., and van Diggelen, J. (2018). Using perceptual and cognitive explanations for enhanced human-agent team performance. In *Engineering Psychology and Cognitive Ergonomics: 15th International Conference, EPCE 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings 15*, pages 204–214. Springer.

Pfister, H.-R. and Böhm, G. (2008). The multiplicity of emotions: A framework of emotional functions in decision making. *Judgment and decision making*, 3(1):5–17.

Phillips, R. and Madhavan, P. (2013). The role of affective valence and task uncertainty in human-automation interaction. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 354–358. SAGE Publications Sage CA: Los Angeles, CA.

Polignano, M., Narducci, F., de Gemmis, M., and Semeraro, G. (2021). Towards emotion-aware recommender systems: an affective coherence model based on emotion-driven behaviors. *Expert Systems with Applications*, 170:114382.

Ren, D., Amershi, S., Lee, B., Suh, J., and Williams, J. D. (2016). Squares: Supporting interactive performance analysis for multiclass classifiers. *IEEE transactions on visualization and computer graphics*, 23(1):61–70.

Ribera, M. and Lapedriza, A. (2019). Can we do better explanations? a proposal of user-centered explainable ai. CEUR Workshop Proceedings.

Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human behavior and emerging technologies*, 1(1):33–36.

Rohlfing, K., Cimiano, P., Scharlau, I., Matzner, T., Buhl, H., Buschmeier, H., Grimminger, A., Hammer, B., Häb-Umbach, R., Horwath, I., Hüllermeier, E., Kern, F., Kopp, S., Thommes, K., Ngonga Ngomo, A.-C., Schulte, C., Wachsmuth, H., Wagner, P., and Wrede, B. (2021). Explanation as a social practice: Toward a conceptual framework for the social design of AI systems. *IEEE Transactions on Cognitive and Developmental Systems*, 13:717–728.

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., and Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics*, 5:17–34.

Scharowski, N., Perrig, S. A., von Felten, N., and Brühlmann, F. (2022). Trust and reliance in xai–distinguishing between attitudinal and behavioral measures. *arXiv preprint arXiv:2203.12318*.

Schemmer, M., Hemmer, P., Nitsche, M., Kühl, N., and Vössing, M. (2022). A meta-analysis of the utility of explainable artificial intelligence in human-ai decision-making. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, pages 617–626.

Schoonderwoerd, T. A., Jorritsma, W., Neerincx, M. A., and Van Den Bosch, K. (2021). Human-centered xai: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human-Computer Studies*, 154:102684.

Schütze, C., Lammert, O., Richter, B., Thommes, K., and Wrede, B. (2023). Emotional debiasing explanations for decisions in hci. In *International Conference on Human-Computer Interaction*, pages 318–336. Springer.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition & emotion*, 14(4):433–440.

Shulner-Tal, A., Kuflik, T., Kliger, D., and Mancini, A. (2025). Who made that decision and why? users' perceptions of human versus ai decision-making and the power of explainable-ai. *International Journal of Human–Computer Interaction*, 41(7):4230–4247.

Thommes, K., Lammert, O., Schütze, C., Richter, B., and Wrede, B. (2024). Human emotions in ai explanations. In *World Conference on Explainable Artificial Intelligence*, pages 270–293. Springer.

Turel, O. and Kalhan, S. (2023). Prejudiced against the machine? implicit associations and the transience of algorithm aversion. *Mis Quarterly*, 47(4).

Wang, D., Yang, Q., Abdul, A., and Lim, B. Y. (2019). Designing theory-driven user-centric explainable ai. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–15.

Weitz, K., Schiller, D., Schlagowski, R., Huber, T., and André, E. (2019). " do you trust me?" increasing user-trust by integrating virtual agents in explainable ai interaction design. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pages 7–9.

You, S., Yang, C. L., and Li, X. (2022). Algorithmic versus human advice: Does presenting prediction performance matter for algorithm appreciation? *Journal of Management Information Systems*, 39(2):336–365.

Zhang, Y., Liao, Q. V., and Bellamy, R. K. (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*, pages 295–305.