THE DIFFERENTIAL EFFECT OF EXPLAINABLE ARTIFICIAL INTELLIGENCE ON TECHNOSTRESS: TOWARD AN EXPLANATION OF USAGE INTENTION, SATISFACTION, AND PERFORMANCE OUTCOMES

Research-in-Progress

1 Introduction

New digital technologies are fueled by health data and machine learning algorithms. They will have a substantial impact on the transformation process of new diagnostics, innovative treatments, or predictive healthcare. For instance, artificial intelligence (AI) integration into the clinical workflow will lead to a new form of hybrid work for physicians in complex decision-making (Jussupow et al., 2021). In the development of such algorithms, increasing attention is being paid to ensuring that AI-based recommendations not only work effectively but also appear understandable and comprehensible to their users. The ability to explain system decisions is critical for the success of the system (Fernández-Loría et al., 2022). However, balancing both, the performance of a system and comprehension by its user, remains one of the biggest challenges in information systems (IS) and computer science (Rai, 2020). Building on well-documented technical challenges of explainable artificial intelligence (XAI), more information and additional explanations are only useful if the users can process them effectively. Research on algorithmic control from Cram et al. (2022) points to adverse effects when users are confronted with algorithms and cannot effectively process information from guidance. As a result, it may be desirable that AI systems provide understandable guidance but only if it facilitates effective interaction with the system (i.e., potential negative effects must be considered). Based on this observation, we see an area of tension in data-driven decision-making with XAI in healthcare. In this area, XAI is intended to support users but could simultaneously serve as an additional trigger for stress. Mastering this challenge and understanding the phenomenon is instrumental for enabling better decisions and improving innovative solutions (e.g., value-based care for patients). However, the way

XAI is appraised by its users will determine whether its technological potential can be realized, and stress can be channeled into positive outcomes.

Recent research in IS distinguishes two types of stressors that can arise in technology interactions: challenge technostressors as the "bright side" and hindrance technostressors as the "dark side" of IS use stress (Benlian, 2020; Califf et al., 2020; Maier et al., 2021; Tarafdar et al., 2019). The confrontation with XAI to make the right decision has the potential to facilitate or mitigate both types of technostress (Califf et al., 2020). We argue that XAI can occupy a role to support both, bearing challenge technostressors and carrying hindrance technostressors. Therefore, the right XAI design to raise understanding of the AI while interacting in daily life could shape a new opportunity for hybrid intelligence and an emergent configuration in future well-being (Dellermann et al., 2019).

Previous research on XAI has focused on the positive effects of explainability (Hamm et al., 2021; Meske et al., 2022; Rai, 2020). However, to the best of our knowledge, this is the first research that takes a broader perspective on XAI considering its adverse effects and whether XAI should be included in the decision-making process to support and/or undermine stress and finally improve outcomes. Considering the potential of XAI for decision-making and the associated risk, there is an urge to understand the implications of explainable designs to realize their benefits. We, therefore, aim at answering the question:

"How does explainability influence AI-based decision-making, technostress, and downstream consequences?"

To answer this question, we ground our research on perceptions of XAI and the holistic stress process model (Califf et al., 2020). The holistic stress process model postulates that the appraisal of a specific technology environment condition decides whether the user appraises the technology as a hindrance or a challenge (Tarafdar et al., 2019). Based on this theoretical background, we observe end users of an AI-based nutrition app in an online lab experiment. We aim to shed light on how end users' (1) *performance*, (2) *usage intention*, and (3) *satisfaction* manifest depending on the design of decision-making support through XAI in the context of an AI-based nutrition app. Our intended contributions will be the exploration of adverse effects of XAI, and a nuanced understanding of its influence on technostress and outcomes. In this sense, we aim at contributing to research on XAI and technostress in healthcare.

2 Theoretical Foundation

Following Berente et al. (2021, p. 12), we define AI "as the frontier of computational advancements that references human intelligence in addressing ever more complex decision-making problems." This frontier embeds two dimensions, performance, and scope. Performance describes the "ever-improving execution of tasks to which AI is applied while scope describes the "ever-expanding range of contexts to which AI is applied" (Berente et al., 2021, p. 12).

In progressively complex situations, AI promises to take a game-changer role to satisfy the demand for quicker and validated decisions by making large amounts of data accessible, usable and utilizable. The roles of humans when interacting with AI are not clearly defined and AI can take on a superior role in interaction with humans. Precisely, AI can outperform humans (Shen et al., 2019) or outperform human crowds (Fu et al., 2021). Seeing AI as a frontier expands our horizon of understanding by showing that AI is perceived not only as a phenomenon but rather as a moving target of evolving phenomena (Berente et al., 2021). AI is an example of a new generation of agentic IS artifacts that require revisiting the human agency primacy assumption (Baird & Maruping, 2021). Those technologies are no longer passive tools but can assume certain delegations of tasks (Baird & Maruping, 2021).

Being confronted with such technologies as AI could lead to insecurity because end-users receive it as a black-box and are not able to comprehend the decision suggestions. One solution for the black-box problem is to make the AI *explainable*. We define *explainability* as "generating decisions in which one of the criteria taken into account during the computation is how well a human could understand the decisions in the given context, which is often called interpretability or explainability." (Miller, 2019, p. 3). However, recent IS research reveals that AI interactions come along with other challenges. Those challenges discussed in the literature include aversion (Berger et al., 2021), cognitive challenges (Jussupow et al., 2021), confirmation challenges (Jussupow et al., 2022), lack of "true" ground truth (Lebovitz et al., 2021), opacity (Lebovitz et al., 2022), or unique human knowledge (Fügener et al., 2021). Surprisingly little is known about decision-making with XAI and the resulting influence on technostress and outcomes. However, there are major challenges, especially in the interaction between users and XAI (Barredo Arrieta et al., 2020; Berente et al., 2021). For this reason, we aim to understand challenge and hindrance technostressors with the tension of explainability and its diverse influences on downstream consequences.

The main concept of technostress is defined as “a modern disease of adaptation caused by an inability to cope with new computer technologies in a healthy manner” (Ragu-Nathan et al., 2008, p. 418). This concept gains particular importance when interacting with new innovative technologies, such as AI, as these technologies not only assume a passive role but in the future will actively delegate tasks or suggest solutions to end-users for decision-making (Baird & Maruping, 2021). In the past, most research so far has focused on negative technostressors (Tarafdar et al., 2019). Negative technostressors are associated with constraining work-related tasks and are appraised by users as destructive (Tarafdar et al., 2019). This results in a majority of negative effects on work outcomes from the literature in the past (Tarafdar et al., 2019). Techno-distress stands for end-users experiencing “bad” stress and IS as a threat (Tarafdar et al., 2019). In turn, techno-eustress stands for end-users interacting with technology and receiving the appraisal as challenging. First studies have provided conceptual (Tarafdar et al., 2019) and empirical insights into their impact on IS use in general (Benlian, 2020; Maier et al., 2021) and on medical professionals (Califf et al., 2020). From previous research, we find that the design of technology decides whether the end-users appraise the conditions as positive or negative stress.

There are many different types of technostressors and their instantiations depend on the research context. In the context of XAI, feature overload (Zhang et al., 2016) should be particularly prominent as XAI provides the end-user with more comprehensible features for decision-making, delegation, or prevention as an IS. Feature overload means that end-users of a technology perceive the features of the technology and its interaction as threatening. Following Califf et al., (2020), we identify involvement as our main challenge technostressor. Involvement means that end-users of a technology perceive the technology and its interaction as an opportunity to improve their work (Califf et al., 2020). We see involvement as our main positive technostressor because the higher users are trying to understand and/

or comprehend the IS, they involve more in the interaction with the technology. Building upon the nature of XAI and a contextualized understanding of technostress, we postulate eight hypotheses. We first expect that the additional information provided by XAI compared to AI decision support without explanation will foster positive and negative technostress because of the imbalance effects of cognitive support in hybrid decision-making with AI. In particular, we expect the explanation to facilitate involvement (H1a; challenge technostressor) as well as feature overload (H1b; hindrance technostressor. In line with prior research, we expect challenge technostressors to influence three outcome variables such that they increase performance (H2a), satisfaction (H2b), and intention to use (H2c). Conversely, we expect hindrance technostressors to reduce performance (H3a), satisfaction (H3b), and intention to use (H3c).

3 Methodology

To test our hypotheses, we conducted an online lab experiment with users and an AI-based smart nutrition app called *eatbetter*. The main purpose of this app is to support its users in exchanging ingredients for sugar reduction while maintaining the flavor of the dish. We decided to investigate a nutrition context because a healthy diet and nutritional intolerances are still a big challenge for the majority of the western population. An online experiment fitted the purpose of our study because it allowed us to measure the potential effects precisely and with high internal validity (Karahanna et al., 2018). In the task, the participants were asked to replace one ingredient with the most sugar in a dish and replace it with a lower-sugar alternative while maintaining the original flavor. An AI helped the user to make decisions by recognizing the main ingredients of the dish via image recognition and then recommending which ingredient is the best for replacement with recommending an alternative. We manipulated the explainability of AI decision-making support. In two conditions, the AI support consists of (1) a recommendation of the AI either with explanations of flavor similarity and sugar reduction (white box) or (2) a recommendation with no information (black box). We measure involvement, feature overload, satisfaction, and intention to use on seven-point Likert scales (1=”strongly disagree” and 7 = “strongly agree”) established in previous studies and aligned to an AI-based recommendation system. Performance was measured by correct answers based on real-world flavor molecules and sugar ingredients. Further, we checked the manipulation of explainability by measuring comprehension and understanding (Shin, 2021). At the end of the questionnaire, we measured age, gender, experience with AI, experience with cooking, and information technology self-efficacy as controls. We used a panel of 101 diverse, heterogeneous participants from the United Kingdom and collected the data after a pretest on Prolific. For the data analysis, we adopted seemingly unrelated regressions to test our hypothesis. We present the results of our data analysis in Figure 1.

4 Results

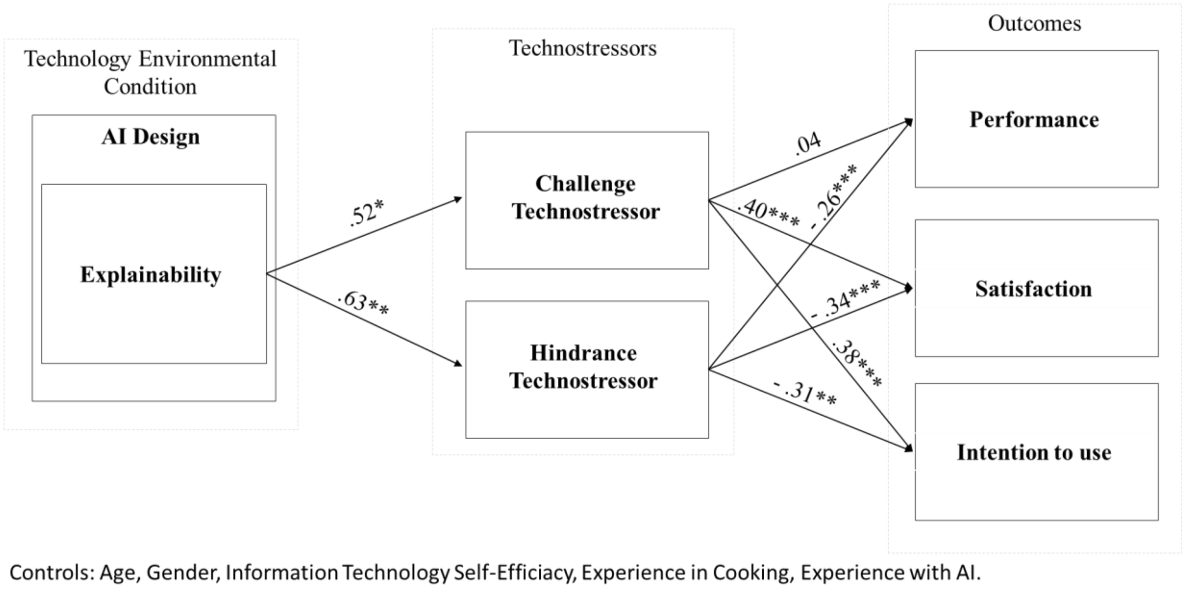

*Figure 1. Research Model. Note. n=101. * $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$.*

5 Expected Contribution

Upon completion, our findings will contribute to IS research in three ways: First, we introduce a theoretical model that links the tension of XAI to technostress, and outcomes. Thereby, we contribute to research on technostress by linking the design of AI-based recommendation apps to positive and negative technostress perceptions. Second, we expand our understanding of decision patterns in the interaction between end-users and XAI in replacement tasks in an experimental setting with explainability of AI (Rai, 2020). Third, we complement prior research focusing on the positive effects of XAI with our investigation of the adverse effects of XAI. We would like to develop our research-in-progress paper and are highly interested in valuable feedback for further development in a different health IT context for a journal submission.

References

Baird, A., & Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS Quarterly*, *45*(1), 315–341. https://doi.org/10.25300/MISQ/2021/15882

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Benlian, A. (2020). A Daily Field Investigation of Technology-Driven Spillovers from Work to Home. *MIS Quarterly*, *44*(3), 1259–1300. https://doi.org/10.25300/MISQ/2020/14911/

Berente, N., Gu, B., Recker, J., & Santanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, *45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). Watch Me Improve—Algorithm Aversion and Demonstrating the Ability to Learn. *Business & Information Systems Engineering*, *63*(1), 55–68. https://doi.org/10.1007/s12599-020-00678-5

Califf, C. B., Sarker, S., & Sarker, S. (2020). The Bright and Dark Sides of Technostress: A Mixed-Methods Study Involving Healthcare IT. *MIS Quarterly*, *44*(2), 809–856. https://doi.org/10.25300/MISQ/2020/14818

Cram, W. A., Wiener, M., Tarafdar, M., & Benlian, A. (2022). Examining the Impact of Algorithmic Control on Uber Drivers' Technostress. *Journal of Management Information Systems*, *39*(2), 426–453. https://doi.org/10.1080/07421222.2022.2063556

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid Intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Fernández-Loría, C., Provost, F., & Han, X. (2022). Explaining Data-Driven Decisions made by AI Systems: The Counterfactual Approach. *Management Information Systems Quarterly*, *46*(3), 1635–1660.

Fu, R., Huang, Y., & Singh, P. V. (2021). Crowds, Lending, Machine, and Bias. *Information Systems Research*, *32*(1), 72–92. https://doi.org/10.1287/isre.2020.0990

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will Humans-in-The-Loop Become Borgs? Merits and Pitfalls of Working with AI. *MIS Quarterly*, *45*(3), 1527–1556. https://doi.org/10.25300/MISQ/2021/16553

Hamm, P., Wittmann, H. F., & Klesel, M. (2021). Explain it to Me and I will Use it: A proposal on the Impact of Explainable AI on Use Behavior. In *ICIS 2021 Proceedings*. https://aisel.aisnet.org/icis2021/ai_business/ai_business/9

Jussupow, E., Spohrer, K., & Heinzl, A. (2022). Radiologists' Usage of Diagnostic AI Systems: The Role of Diagnostic Self-Efficacy for Sensemaking from Confirmation and Disconfirmation. *Business & Information Systems Engineering*, forthcoming. https://doi.org/10.1007/s12599-022-00750-2

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence. *Information Systems Research*, *32*(3), 713–735. https://doi.org/10.1287/isre.2020.0980

Karahanna, E., Benbasat, I., Bapna, R., & Rai, A. (2018). Editor's Comments: Opportunities and Challenges for Different Types of Online Experiments. *MIS Quarterly*, *42*(2), iii–xi.

Lebovitz, S., Levina, N., & Lifshitz-Assa, H. (2021). Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MIS Quarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis. *Organization Science*, *33*(1), 126–148. https://doi.org/10.1287/orsc.2021.1549

Maier, C., Laumer, S., Tarafdar, M., Mattke, J., Reis, L., & Weitzel, T. (2021). Challenge and Hindrance IS Use Stressors and Appraisals: Explaining Contrarian Associations in Post-Acceptance IS Use Behavior. *Journal of the Association for Information Systems*, *22*(6), 1590–1624. https://doi.org/10.17705/1jais.00709

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management*, *39*(1), 53–63. https://doi.org/10.1080/10580530.2020.1849465

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Ragu-Nathan, T. S., Tarafdar, M., Ragu-Nathan, B. S., & Tu, Q. (2008). The Consequences of Technostress for End Users in Organizations: Conceptual Development and Empirical Validation. *Information Systems Research*, *19*(4), 417–433. https://doi.org/10.1287/isre.1070.0165

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Shen, J., Zhang, C. J. P., Jiang, B., Chen, J., Song, J., Liu, Z., He, Z., Wong, S. Y., Fang, P.-H., & Ming, W.-K. (2019). Artificial Intelligence Versus Clinicians in Disease

Diagnosis: Systematic Review. *JMIR Medical Informatics*, *7*(3), e10010. https://doi.org/10.2196/10010

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Tarafdar, M., Cooper, C. L., & Stich, J. (2019). The technostress trifecta - techno eustress, techno distress and design: Theoretical directions and an agenda for research. *Information Systems Journal*, *29*(1), 6–42. https://doi.org/10.1111/isj.12169

Zhang, S., Zhao, L., Lu, Y., & Yang, J. (2016). Do you get tired of socializing? An empirical explanation of discontinuous usage behaviour in social network services. *Information & Management*, *53*(7), 904–914. https://doi.org/10.1016/j.im.2016.03.006