Mitigating Discontinuance in Medical AI Systems: The Role of AI Explanations

Research Paper

Aycan Aslan[1], Maike Greve[1], and Lutz Kolbe[1]

[1] University of Goettingen, Chair of Information Management, Goettingen, Germany
{aycan.aslan,maike.greve,lkolbe}@uni-goettingen.de

Abstract. Despite significant advancements in medical artificial intelligence (AI) systems, these technologies are prone to mistake in their predictions. These mistakes can significantly affect medical experts' willingness to continue using these systems. To mitigate potential discontinuation, existing research indicates that providing additional information alongside predictions, can lessen negative outcomes like discontinuation. Given the potential impact on users' information processing, we hypothesize that AI explanations, detailing the system's decision-making process, can also influence the likelihood of discontinuing use after an AI mistake. Through an online experiment with medical experts (n=227), we demonstrate that such explanations can influence medical experts' information processing and, consequently, mitigate the adverse effects on the actual discontinuation of AI systems following a mistake.

Keywords: Artificial intelligence, decision-making, explainability, discontinuance, medicine.

1 Introduction

The deployment of artificial intelligence (AI) based systems offers great potential for medical applications. AI systems can enhance the decision-making of medical experts by supporting them in complex medical image analysis (Mehta and Pandit, 2018) or automatic extraction of relevant information from clinical notes with the use of natural language processing (Roski et al., 2014). Despite the potential of such systems, recent studies highlight that current AI is imperfect and hence may provide wrong AI advice (Ali et al., 2023). AI systems remain error-prone and the possible discontinuance of use when such systems make mistakes are a serious problem. Such mistakes are especially prevalent in highly complex medical applications. AI-based decision-support in medical settings is plagued with high levels of uncertain, incomplete, or imbalanced datasets (Holzinger et al., 2019). All these factors increase the probability of the AI making mistakes, even when the systems are constantly improved. Previous studies show that in the event of an AI mistake, users often question its inherent capabilities to support them and lose trust in the system (Adam et al., 2021; Weiler et al., 2022).

19th International Conference on Wirtschaftsinformatik,
September 2024, Würzburg, Germany

Prior literature highlights, that it is possible to counter such discontinuance with targeted messages. For example, studies have explored the option to provide warning messages to counteract usage discontinuance (Weiler et al., 2022). Such warning messages, also called inoculation messages, state the possibility of AI mistakes prior to an interaction of the user with the system and aim at decreasing the likelihood of discontinuance (McGuire, 1964). This kind of immunization through inoculation messages preemptively protects against stronger stimuli, i.e., discontinuance, by forewarning the users and realigning users' expectations to potential threats (McGuire, 1964).

However, such inoculation messages must be designed specifically (Weiler et al., 2022), which raises the questions whether already provided information with AI advice can function in a similar role to inoculation messages. AI explanations, which supplement predictions and are increasingly mandated by regulatory authorities (European Commission, 2021), emerge as a viable candidate for this role. AI explanations generally aim at making complex AI outcome generation more transparent by explaining to the end-user, in a humanly understandable way, how and why the AI system arrived at a certain prediction (Ribera and Lapedriza, 2019). Recent research points to the fact that providing users with an explanation for the given AI prediction not only reflects *more* information but reshapes their overall decision-making, for example influencing which information decision-makers include or disregard in their decisions (Bauer et al., 2023; Jussupow et al., 2021).

Nevertheless, while prior research indicates the influence of explanations on the decision-making of users, there are no experimental studies that investigate the role of explanations on the discontinuance behavior after a mistake. If explanations can effectively mitigate the discontinuance of AI systems similarly to established methods like inoculation messages, this would be beneficial. Explanations are increasingly becoming a regulatory requirement and thus must be implemented regardless, whereas other methods such as inoculation messages are voluntary and need to be purposefully crafted. Hence, we ask following research question:

RQ: What is the effect of AI explanations on the user's decision to discontinue the use of a medical AI system following a mistake, i.e. an incorrect medical diagnosis?

To answer this research question, we performed a 2x2 vignette study through an online scenario experiment with medical experts (n=227), grounded on a conducted workshop with two doctors. In this experiment, medical experts were asked to carry out three medical decision tasks supported by the advice of an AI-based system. To understand the perception of the treatments after a mistake, the AI system gave the right diagnosis for the first two tasks but provided a mistake, i.e. an incorrect diagnosis, for the third task. In this experiment, we use a dissimulative design to measure the actual discontinuance by the medical experts, rather than only their intent. We did so, as prior research has shown that measuring only the intent of participants, especially in complex medical settings, is not sufficient and distorts the true perception of the user (Limayem et al., 2007; Maillet et al., 2015; Straub et al., 1995; Turner et al., 2010). The main goal of the experiment is to understand whether the provision of an explanation influences discontinuance when the system makes a mistake. To testify the direct comparison to

AI explanations, our experiment also tests the effect of inoculation messages amongst medical experts. Our findings demonstrate that providing AI explanations increases the likelihood that medical experts will continue using AI support. Our findings contribute to the literature by showing the positive effect of explanations on users' system usage after a mistake and to practice by offering explanations as a potent tool to preemptively protect against mistake-triggered discontinuance of AI systems.

2 Background and Related Work

This section provides the background for our work, by explaining what AI explanations are and clarifying what the relevant literature streams for our work are.

Generally speaking, the goal of AI explanations is to make AI systems more transparent and understandable and to combat the emerging problem of black-box AI systems, a term that describes the phenomenon that such systems can be, in part or completely, not comprehensible for individuals (Meske et al., 2020). Here, explanations can address the inner workings of the AI system or a specific prediction, by showcasing what the system did and how it performed such action (Meske et al., 2020).

In terms of related work, our study relates to two intersecting research streams: 1) the existing literature on negative responses to algorithmic mistakes and 2) the effect of explanations on the cognitive processes of humans. Regarding the first research stream, prior studies have investigated the negative reactions users exhibit after interacting with algorithms that make mistakes. These studies reveal that users lose confidence in algorithmic agents more quickly than in human agents after observing similar mistakes (Dietvorst et al., 2015; Jones-Jang and Park, 2022). Additionally, research indicates that this loss of confidence stems from the users' assumption that algorithms are less capable of learning effectively compared to their human counterparts (Reich et al., 2023). On the second research stream, studies implicate that a provided explanation, additional to the prediction, is not only perceived as *more* information but also has the ability to reshape the user's sense-making more generally (Bauer et al., 2023; Jussupow et al., 2021). Explanations influence users' situational weighting of available information, leading them to include or disregard certain information (Bauer et al., 2023). In this context, explanations can, for instance, have a negative effect by fostering confirmation bias, enforcing users' tendency to selectively process information that strengthens their beliefs (Bauer et al., 2023). Similarly, studies show that explanations can influence users' cognitive processes by enticing them to rely on certain information sources (Jussupow et al., 2021). Here, explanations lead users to rely upon beliefs rather than actual data, when disconfirming with the explanation (Jussupow et al., 2021).

Upon reviewing these two research streams, the gap we aim to address in this study becomes evident. While there are studies examining the effects of algorithmic mistakes, they predominantly focus on the impact of predictions, overlooking the increasingly significant role of explanations. Our newly acquired understanding that explanations influence users' cognitive processes underscores the need for this study. Therefore, our research seeks to fill the gap by investigating how AI explanations affect users' decisions to continue using the system after it has made a mistake.

3 Theorizing the Effect of Explanations on Discontinuance

Now that we understand the background on explanations influencing users' decision-making, we want to theorize the effect of explanations on users' actual discontinuance behavior after an AI mistake. As prior research has proven the positive effect of messages such as inoculation messages (Adam et al., 2021; Compton, 2012; Weiler et al., 2022), these messages will serve as the baseline against which explanations can be compared. Hence, we will first build on the research on inoculation messages and theorize their mitigating effect on discontinuance as a baseline for our experiment. After that we theorize that explanations have a similar mitigating effect on discontinuance.

Generally, mistakes of IS, including AI systems, provoke negative reactions from users (Mahmood et al., 2022; Spreng et al., 1995). In response, according to the inoculation theory, users can be inoculated against attitude change, through inoculation messages (Compton, 2012). To achieve such resistance against attitude change, the user has to be warned about the potential threat to them (Adam et al., 2021; Compton, 2012; Weiler et al., 2022). In the case of medical AI systems, the potential threat to the users, i.e., the medical experts, is the AI system's possible underperformance. By warning the medical expert about possible mistakes of the AI system, the expectations of the experts are altered and consequently lowered. If a mistake occurs following this realigning of expectations, the negative attitude change, in the form of discontinuing the use of the AI system, will be alleviated, as the expectations have been lowered by the inoculation message. Essentially, the gap between the user's expectation and the worst outcome, an AI mistake, is reduced by the inoculation message. This effect has been shown for other domains and interactions with AI systems. For example, studies demonstrate that inoculation messages can alleviate user discontinuance after a mistake in the context of customer-service chatbots for online banking (Adam et al., 2021; Weiler et al., 2022). Hence, we hypothesize that:

H1: The provision of an inoculation message decreases the level of actual discontinuance after an AI mistake.

Building on the demonstrated impact of explanations on cognitive processes (see section 2), we can theorize the expected effect of providing AI explanations on users' discontinuance after an AI mistake. As mentioned, the additional provision of an explanation for the AI prediction influences which information is considered for the user's assessment of the situation (Bauer et al., 2023; Jussupow et al., 2021). We argue that a provided explanation, when the outcome is in line with their own beliefs of the situation, may lead users to be more forgiving, even if the system eventually makes mistakes. This is based on the pervasive human inclination to process information in a way that confirms their existing preconceptions. Here, the provision of an explanation is essential, as it lets the user compare the decision-making process of the system with their own beliefs and logic. This comparison would not be possible with a "traditional" black-box AI system, which only provides a prediction without any explanations. By opening the possibility of such comparison, the explanation-triggered higher level of forgiveness may be reflected in alleviated users' discontinuance decision after an AI

mistake. Operationalized for our medical context, we expect that the explanation will allow medical experts to compare the process of the medical AI with their logic based on their medical education and expertise. If the AI makes a mistake, we expect the medical expert to show lower tendencies of discontinuing the use of the system, as the explanation represents a unique tool for the medical expert to compare the system's decision-making with its own. Hence, we hypothesize that:

H2: The provision of AI explanations decreases the level of actual discontinuance after an AI mistake.

4 Research Design and Experimental Setting

To test the stated hypotheses, we used a 2x2 between-subjects experimental design in an online scenario experiment. Our goal was to test our hypotheses through multiple medical scenarios in which the participating experts will be asked to diagnose a disease based on a patient's presented symptoms using an AI system. In a joint workshop with two doctors, we designed an experimental scenario that took into consideration all relevant medical information to ensure that the scenario is logical, realistic, and can be understood by medical experts. The participants for the experiment were recruited from the online platform Prolific. Here, we used the criterion that only participants active in the medical sector can participate in our study. After removing 13 incomplete (e.g., missing built-in attention checks) responses, the final sample contained 227 survey responses. The average age of the participants was 38.9 years, of which 135 were women (59.9%), 90 were men (39.6%), and 2 indicated that they have a gender other than male or female (0.1%). In this group of participants, we had a strong representation of doctors. Out of the 227 participants, 132 were practicing doctors (58.1%). Further, 65 participants were nurses (28.6%). The rest can be grouped as being active in medical research (13.2%). The average reported years of medical expertise was 11.1 years. Additionally, we asked for the IT and AI knowledge of the participants. Out of all participants, 156 (68,7%) stated that they would rate their IT knowledge as high or very high. Only 48 participants (21.5%) stated that they would rate their AI knowledge as high or very high. Manipulation and attention checks were performed to ensure participants were able to relate to the shown medical scenario. The survey was conducted in February 2023.

4.1 Manipulations

All participants were presented with the same medical scenarios and were asked to solve the same task, i.e., deciding on which diagnosis likely persists for the given patient in cooperation with the AI system. The control group was asked to diagnose the patients with an AI system which only presented the prediction of the system, with no further information. The treatment group with the inoculation scenario was presented with an inculcation message which warned them about possible AI mistakes. The formulation chosen for the message was based on prior literature but adapted to our context

(Weiler et al., 2022). The explicit formulation chosen was: “The algorithm is still learning and produces errors. Consider that the wrong assessment of medical information by the AI system can have severe consequences for the patient.” For the explanation scenario, the participants were shown an explanation for the given AI advice. Here, we based the presented structure on prior literature (Ribera and Lapedriza, 2019) and made sure that it was in line with other similar experimental set-ups (Aslan et al., 2022). The fourth group received the inoculation message as well as the explanations. These four groups (1) control, 2) inoculation, 3) explanation, 4) inoculation and explanation) represent our 2x2 between-subjects design. Examples of the described treatment groups can be seen in Figure 1.

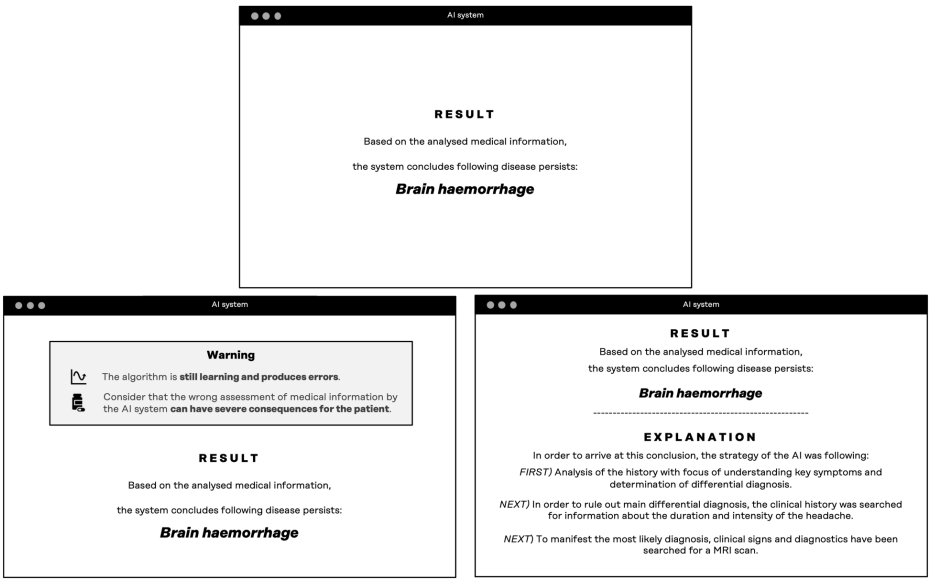

Figure 1. Control Scenario (Top), Inoculation Scenario (Bottom Left) and Explanation Scenario (Bottom Right)

4.2 Procedure

We set up the participants in a medical scenario in which they were supposed to solve three medical tasks in collaboration with the AI system. Even though the AI system used was only a mock-up version that presented the participants with the screenshots of the final decision, the functioning still closely aligns with those of a fully-implemented medical AI system (Panigutti et al., 2020). For each of the cases, a different hypothetical patient with respective symptoms was presented. After being shown the AI advice (either control, inoculation, or explanation), the expert was asked to indicate whether they agree or disagree with the diagnosis. The procedure consisted of four main steps (see Figure 2):

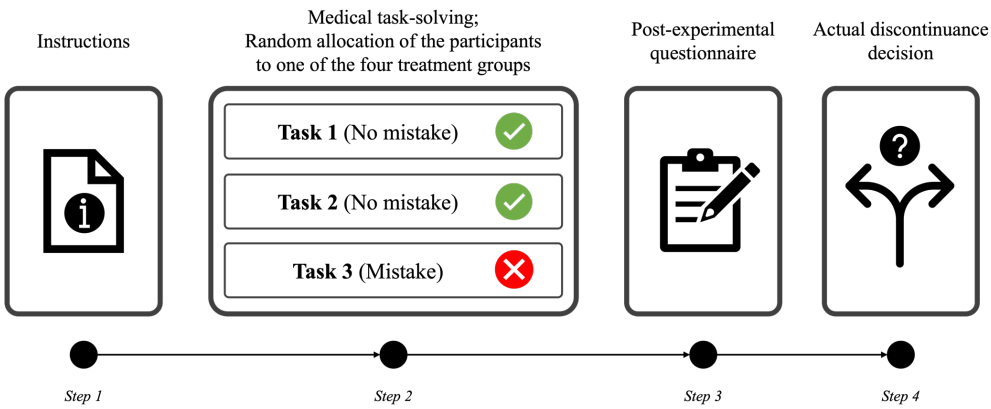

Figure 2. Experimental Procedure

(1) The experiment started with a short introduction to the participants that included information about the experimental procedure and instructions for the task. The participants were asked to put themselves into a situation where they adopt an AI system as a primary care doctor which analyzes the medical documents of the patients and supports them in their diagnosing process. The specific formulation was following: "Please imagine that you are a primary care doctor which looks into the medical cases of your patients to decide on what disease persists, based on present symptoms. In this context, you are able to use a new Artificial Intelligence (AI) based system, called AuMEDa, that supports you in deciding what disease persists for the given patient based on the analysis of the symptoms.". We told the participants that the entire study consisted of six medical tasks and after half of the tasks, they would be asked to complete a questionnaire. After the questionnaire, they would decide whether to continue with or without the AI system for the second half of the experiment. In actuality, the experiment ended after the first three tasks and the questionnaire, but we used this dissimulative design based on prior literature (Weiler et al., 2022) as we wanted to capture the medical experts' actual discontinuance decisions rather than only statements of intent.

(2) Next, the participants were randomly allocated to one of the four treatment groups. The participants were asked to diagnose the following three medical cases with the help of an AI system: 1) A 20-year-old female with *appendicitis*, 2) A 67-year-old male with a *myocardial infarction*, and 3) A 34-year-old female with a *severe migraine*. For each of the cases, the medical experts were presented with realistic symptoms based on the workshop we conducted with the doctors. Crucially, while the first two cases were assessed correctly by the AI system, it made a mistake in the third task. While the symptoms indicated a severe migraine for the 34-year-old female patient, the system gave the (wrong) diagnosis of *brain hemorrhage* as it incorrectly interpreted the MRI results.

(3) After the completion of the third task, participants answered the post-experimental questionnaire about their experience with the AI system and other questions, capturing our control variables. We also measured latent constructs to gain a deeper understanding of how the medical experts perceived the treatment.

(4) Finally, participants could decide whether they would like to continue the second part of the experiment with the AI system or without it. This decision was used as the dependent variable.

After capturing the participants' discontinuance decision, we debriefed them and thanked them for their participation.

4.3 Operationalization

We measured our dependent variable (actual discontinuance), as a binary coded variable by asking the medical experts to choose between the two options: "I would like to continue using the AI system" or "I would like to discontinue using the AI system". Since our dependent variable is binary coded, we conducted a logistic (logit) regression, a probabilistic discriminative classifier, to understand the impact of our treatments on the dependent variable, an approach that is in line with related research (Murphy, 2012; Weiler et al., 2022).

Further, we made sure that the two right and one wrong (i.e., the AI mistake) diagnoses by the system were perceived as such. For this, we asked the participants to indicate whether they would agree or disagree with the provided AI prediction and specify what they believed was the correct diagnosis if they disagreed. For our first task, in which the AI system provided the correct diagnosis, on average over all treatments, 97.9% of all participants (223 participants) agreed with the prediction. For task two, in which the AI system also provided the correct diagnosis, on average over all treatments, 95.2% of all participants (216 participants) agreed with the prediction. For the third task, for which we deliberately planned an AI mistake, on average over all treatments, 86.0% of all participants (195 participants) disagreed with the given prediction, of which a majority indicated that they would rather conclude that a migraine persists (which was the correct assessment). This indicates that the last task was sufficiently perceived as an AI mistake of the system, as planned for in our manipulation.

The respective numbers, detailed for the given treatment group, can be seen in Table 1. To sum up, the data shows that the predictions in the first two tasks were sufficiently perceived as the correct diagnoses whereas the prediction in the third was perceived as an AI mistake.

Table 1. Perception of the Correct Diagnoses and AI Mistake

	Task 1 (Correct diagnosis)		**Task 2 (Correct diagnosis)**		**Task 3 (Incorrect diagnosis)**	
	Agree	**Disagree**	**Agree**	**Disagree**	**Agree**	**Disagree**
Control	52	1	49	4	6	47
Inoculation	53	3	51	5	7	49
Explanation	58	1	58	1	9	50
Inoculation & Explanation	59	0	58	1	10	49

In addition, we measured latent constructs such as perceived enjoyment (Koufaris, 2002), perceived transparency (Schnackenberg et al., 2021), and perceived usefulness (McKinney et al., 2002) all on a 7-point-Likert scale ranging from strongly disagree to strongly agree. We chose these constructs to understand the overall interaction between the expert and AI system in more detail, as prior research has shown that they can play a relevant factor in the interaction (Araujo et al., 2020; Aslan et al., 2022; Bao et al., 2021). We establish internal consistency and reliability by examining Cronbach's alpha and composite reliability. Both of these are fulfilled if the threshold of 0.7 is met (Ko et al., 2005; Nunnally and Bernstein, 1994), which is the case for all our scales. Further, the results indicate that all discriminant validity requirements were met, as each scale's average variance extracted exceeded the multiple squared correlations (Fornell and Larcker, 1981).

5 Data Analysis and Results

First, we present some descriptive analysis of the data by highlighting the rate of discontinuance in percentage (see Table 2). For our control scenario (no inoculation and no explanation), after the AI made a mistake, 15 out of 53 medical experts decided to discontinue the use of the AI system, which is a discontinuance rate of roughly 28.3%. For our inoculation scenario, 7 out of 56 medical experts decided to discontinue the use of the AI system, amounting to a discontinuance rate of only 12.5%. For the explanation scenario, 7 out of 59 medical experts decided to discontinue the system, which is a discontinuance rate of 11.9%. Lastly, for the group that received the inoculation message and the explanation, 6 out of 59 experts decided to discontinue their system use, amounting to a discontinuance rate of 10.2%.

Table 2**.** Discontinuance Rate of the Treatment Groups

	n	**Decision to continue (with the AI system)**	**Decision to discontinue (with the AI system)**	**Discontinuance rate**
Control	53	38	15	**28,3%**
Inoculation	56	49	7	**12,5%**
Explanation	59	52	7	**11,9%**
Inoculation & Explanation	59	53	6	**10,2%**

The discontinuance rates show that, generally, most people would like to continue using the AI system despite the mistakes it makes. Nevertheless, we also see that the discontinuance rates are lower for the inoculation and explanation treatment group, showing indications toward the hypothesized positive effect of inoculation messages and explanations on the user's information processing and subsequent decision to discontinue usage.

To test the hypotheses, we conducted a logit regression analysis on our binary-coded dependent variable *actual discontinuance* (see Table 3). The results demonstrate that

providing an inoculation message negatively impacted the dependent variable (actual discontinuance) (β = -1.0758, p = .0332), i.e., proving lower levels of discontinuance. Hence, **hypothesis 1 is supported**. Additionally, the results demonstrate that the AI explanation also negatively impacted the dependent variable (β = -1.0164, p = .0447). Therefore, we find **support for hypothesis 2** as well.

As mentioned above, we measured perceived levels of enjoyment, transparency, and usefulness to gain a better understanding of how the overall interaction with the AI system was perceived by the medical experts. Here, we conducted ANOVAs to investigate the effect of the treatment groups on the construct's enjoyment, transparency, and usefulness. We find, that for a p-level < .05, medical experts receiving the inoculation messages reported significantly higher levels of transparency (β = .134, p = .024) and enjoyment (β = .151, p = .011). For the same p-level, the medical experts receiving an explanation reported significantly higher levels of transparency (β = .407, p = < .001) and usefulness (β = .199, p = < .001), but not enjoyment.

Further, we tested our dependent variable for our controlled variables and found that all control variables (age, gender, medical expertise, IT knowledge, and AI knowledge) do not have a significant effect on our dependent variable.

Table 3**.** Results of Logistic Regression on the Dependent Variable Actual Discontinuance in Relation to Control Group

	β	**Std. error**	**z value**	**p-value**
Intercept	-0.9295	0.3049	-3.048	**.0023 ****
Inoculation	-1.0758	0.5050	-2.130	**.0332 ***
Explanation	-1.0164	0.5062	-2.008	**.0447 ***
*Note: n = 227, R^2= 0.202 (Cox & Snell), *p < .05; **p <.01*				

In summary, the conducted logistic regression show support for both our hypotheses, indicating that inoculation messages and AI explanations can lower the tendency of medical experts to discontinue their use of an AI system after it makes a mistake. The conducted analysis of the latent constructs reveals, among other things, that despite the positive effect of AI explanation on discontinuance, it is not perceived as enjoyable as the inoculation treatment.

6 Discussion

Since the existing literature does not address the effect of AI explanations on users' behavior in terms of discontinuance after an AI mistake, our study addresses this gap by showing that explanations can act as a powerful tool to significantly reduce a user's tendency to stop using an AI system. Our experiment shows that AI explanations, like inoculation messages, can influence medical experts processing process by allowing them to compare the AI system's process of arriving at a diagnosis with their own. Giving them this option (which would not be possible with only an AI prediction) allows users to be more forgiving when an AI mistake occurs.

Our analysis of latent constructs reveals that the inoculation group experiences higher levels of enjoyment compared to the explanation group. This could be attributed to the greater cognitive effort needed to fully understand the provided explanations. This suggests that the implementation of explanations in AI systems must be approached with caution, considering both potential positive and negative impacts on users. In the following, we discuss the implications of these findings for the literature and practice and point out future research opportunities.

6.1 Contributions to Literature and Practice

Our study contributes to both to literature and practice in several ways. In terms of literature, our study contributes to the growing literature on the influence of AI explanations on users' cognitive information processing. As shown in the research background, prior literature established that explanations can have a unique impact on users by influencing which and how information are perceived in a decision-making process (Bauer et al., 2023; Ebermann et al., 2022; Jussupow et al., 2021). We contribute to this literature by indicating that such influence on the cognitive processes of users can be used to trigger a desirable behavioral outcome. Prior research shows that for many domains a state of human-AI collaboration is preferable (van den Broek et al., 2021), implying that we would not want the user to stop the use of the AI system after a mistake. In this context, our experiment indicates that providing explanations alters the cognitive process of medical experts and leads them to be more forgiving towards the AI system. This leads to the desired outcome that they are less likely to discontinue the AI system use following a mistake. Thus, we strengthen the literature on the influence of explanations on cognitive processes by showing how such influence lowers the discontinuance rates of AI systems.

In terms of practice, our study shows medical managers that explanations can be used as a strategic tool against the discontinuance of the use of AI. As the overall diffusion of AI promises immense potential for the medical and healthcare sector, there is growing pressure for medical managers to ensure that their organizations invest in and adopt AI systems. Among other factors, this will ensure that their organization does not miss a major technological development and prepare their organization for the future. This trend is proven by the ever-growing sums that are being invested in the global medical AI market (ReportLinker, 2021). However, while AI for medical applications gets more and more important, medical managers struggle with the question of how to positively affect the adoption of such systems by the end-users, i.e., medical experts. Ultimately, the success of deployed medical AI systems depends on whether the users use and continue using them. Here, our study shows that explanations, as a system functionality, can be used strategically against the discontinuance of AI systems after an AI mistake. As discussed, AI mistakes, due to the very probabilistic nature of AI systems, cannot be completely ruled out. Hence, our study shows that explanations can be used as a countermeasure against a problem present today, which will also stay relevant in the future. This potential of explanations promises to be a great advantage for medical managers and organizations that heavily invest in AI systems and consequently hope to reap the benefits of such financial investment.

6.2 Limitations and Opportunities for Future Research

Despite the illustrated contributions, we note two important areas in which future research could strengthen and extend the results of our study. First, we must note the importance of our decision on what type of mistake to build into the experiment. AI systems are complex systems based on many internal parameters so there are various types of mistakes that such systems can make. Here, we can expect that different forms of mistakes can trigger different types of reactions from the users. Second, we recognize that for better medical generalizability, the test of more diverse medical cases is needed. We considered this aspect in the joint workshop with the doctors and proactively designed cases that are not only limited to one medical specialty (appendicitis, myocardial infarction, and migraine are problems associated with different parts of the body), why we believe that our findings are not only bounded to one medical specialty. However, we recognize that different medical cases that also vary in their complexity and scope would help strengthen our results and make the presented findings more generalizable.

7 Conclusion

In conclusion, our work provides valuable insights into how AI explanations lower the likelihood of discontinuance of use after an AI mistake in the medical field. By conducting a between-subject online experiment (n=227) with medical experts we show that explanations, which allow the medical experts to compare their own decision with that of the system, make them more forgiving by significantly alleviating users' discontinuance decision. While other forms of messages, such as inoculation messages, must be designed on purpose, there is a strong regulatory push to implement explanations in AI systems for sensitive applications, such as the medical field. As organizations continue to adopt AI-based systems, they will have to implement more explainability measures due to regulatory pressure. In this context, our study highlights an advantage of explanations, previously unknown. Our study sheds light on the important effect of explanations in the event of an AI mistake, which will continue to be a challenge for the adoption of such a system as they will likely never reach a state of "infallibility". Hence, we contribute to the literature by showcasing a further example of how explanations influence the cognitive processes of users, by demonstrating the effect on discontinuance behavior. Additionally, we contribute to practice by offering a strategic tool for organizations as a countermeasure against the discontinuance of AI systems, in which medical organizations invest heavily.

8 Acknowledgements

This research was supported by the AuMEDa project, which is a joint research project between the University Medical Center Goettingen (UMG) and the University of Goettingen, Chair of Information Management.

References

Adam, M., Wessel, M., Benlian, A., 2021. AI-based chatbots in customer service and their effects on user compliance. Electron Markets 31, 427–445. https://doi.org/10.1007/s12525-020-00414-7

Ali, O., Abdelbaki, W., Shrestha, A., Elbasi, E., Alryalat, M.A.A., Dwivedi, Y.K., 2023. A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. Journal of Innovation & Knowledge 8, 100333. https://doi.org/10.1016/j.jik.2023.100333

Araujo, T., Helberger, N., Kruikemeier, S., De Vreese, C.H., 2020. In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI & Soc 35, 611–623. https://doi.org/10.1007/s00146-019-00931-w

Aslan, A., Greve, M., Braun, M., Kolbe, L.M., 2022. Doctors ' Dilemma – Understanding the Perspective of Medical Experts on AI Explanations. International Conference on Information Systems 2022 0–17.

Bao, Y., Cheng, X., De Vreede, T., De Vreede, G.-J., 2021. Investigating the relationship between AI and trust in human-AI collaboration. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2021.074

Bauer, K., von Zahn, M., Hinz, O., 2023. Expl(AI)ned: The Impact of Explainable Artificial Intelligence on Users' Information Processing. Preprint. https://doi.org/10.1287/isre.2023.1199

Compton, J., 2012. Inoculation Theory. The SAGE Handbook of Persuasion: Developments in Theory and Practice 220–236. https://doi.org/10.4135/9781452218410.n14

Dietvorst, B.J., Simmons, J.P., Massey, C., 2015. Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. Journal of Experimental Psychology: General, 144, 114–126.

Ebermann, C., Selisky, M., Weibelzahl, S., 2022. Explainable AI: The Effect of Contradictory Decisions and Explanations on Users' Acceptance of AI Systems. International Journal of Human–Computer Interaction 1–20. https://doi.org/10.1080/10447318.2022.2126812

European Commission, 2021. Artificial Intelligence Act COM(2021) 206 final. Brussels 0106, 1–108.

Fornell, C., Larcker, D.F., 1981. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research 18, 39. https://doi.org/10.2307/3151312

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., Müller, H., 2019. Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 9, 1–13. https://doi.org/10.1002/widm.1312

Jones-Jang, S.M., Park, Y.J., 2022. How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. Journal of Computer-Mediated Communication 28, zmac029. https://doi.org/10.1093/jcmc/zmac029

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, 713–735. https://doi.org/10.1287/ISRE.2020.0980

Ko, Kirsch, King, 2005. Antecedents of Knowledge Transfer from Consultants to Clients in Enterprise System Implementations. MIS Quarterly 29, 59. https://doi.org/10.2307/25148668

Koufaris, M., 2002. Applying the Technology Acceptance Model and Flow Theory to Online Consumer Behavior. Information Systems Research 13, 205–223. https://doi.org/10.1287/isre.13.2.205.83

Limayem, Hirt, Cheung, 2007. How Habit Limits the Predictive Power of Intention: The Case of Information Systems Continuance. MIS Quarterly 31, 705. https://doi.org/10.2307/25148817

Mahmood, A., Fung, J.W., Won, I., Huang, C.M., 2022. Owning Mistakes Sincerely: Strategies for Mitigating AI Errors. Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3491102.3517565

Maillet, É., Mathieu, L., Sicotte, C., 2015. Modeling factors explaining the acceptance, actual use and satisfaction of nurses using an Electronic Patient Record in acute care settings: An extension of the UTAUT. International Journal of Medical Informatics 84, 36–47. https://doi.org/10.1016/j.ijmedinf.2014.09.004

McGuire, W.J., 1964. Some Contemporary Approaches, in: Advances in Experimental Social Psychology. Elsevier, pp. 191–229. https://doi.org/10.1016/S0065-2601(08)60052-0

McKinney, V., Yoon, K., Zahedi, F., 2002. The measurement of Web-customer satisfaction: An expectation and disconfirmation approach. Information Systems Research 13, 296–315. https://doi.org/10.1287/isre.13.3.296.76

Mehta, N., Pandit, A., 2018. Concurrence of big data analytics and healthcare: A systematic review. International Journal of Medical Informatics 114, 57–65. https://doi.org/10.1016/j.ijmedinf.2018.03.013

Meske, C., Bunde, E., Schneider, J., Gersch, M., 2020. Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. Information Systems Management 1–11. https://doi.org/10.1080/10580530.2020.1849465

Murphy, K.P., 2012. Machine learning: a probabilistic perspective. The MIT Press, Adaptive computation and machine learning series.

Nunnally, J.C., Bernstein, I.H., 1994. The Assessment of Reliability, in: Psychometric Theory. McGraw-Hill, New York, NY, pp. 248–292.

Panigutti, C., Perotti, A., Pedreschi, D., 2020. Doctor XAI: an ontology-based approach to black-box sequential data classification explanations. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency 629–639. https://doi.org/10.1145/3351095.3372855

Reich, T., Kaju, A., Maglio, S.J., 2023. How to overcome algorithm aversion: Learning from mistakes. J Consum Psychol 33, 285–302. https://doi.org/10.1002/jcpy.1313

ReportLinker, 2021. Artificial Intelligence in Healthcare Market by Offering, Technology, Application, End User and Geography - Global Forecast to 2027.

Ribera, M., Lapedriza, A., 2019. Can we do better explanations? A proposal of user-centered explainable AI. CEUR Workshop Proceedings 2327.

Roski, J., Bo-Linn, G.W., Andrews, T.A., 2014. Creating value in health care through big data: Opportunities and policy implications. Health Affairs 33, 1115–1122. https://doi.org/10.1377/hlthaff.2014.0147

Schnackenberg, A.K., Tomlinson, E., Coen, C., 2021. The dimensional structure of transparency: A construct validation of transparency as disclosure, clarity, and accuracy in organizations, Human Relations. https://doi.org/10.1177/0018726720933317

Spreng, R.A., Harrell, G.D., Mackoy, R.D., 1995. Service recovery: Impact on satisfaction and intentions. Journal of Services Marketing 9, 15–23. https://doi.org/10.1108/08876049510079853

Straub, D., Limayem, M., Karahanna-Evaristo, E., 1995. Measuring System Usage: Implications for IS Theory Testing. Management Science 41, 1328–1342. https://doi.org/10.1287/mnsc.41.8.1328

Turner, M., Kitchenham, B., Brereton, P., Charters, S., Budgen, D., 2010. Does the technology acceptance model predict actual use? A systematic literature review. Information and Software Technology 52, 463–479. https://doi.org/10.1016/j.infsof.2009.11.005

van den Broek, E., Sergeeva, A., Huysman, M., 2021. When the machine meets the expert: An ethnography of developing ai for hiring. MIS Quarterly 45, 1557–1580. https://doi.org/10.25300/MISQ/2021/16559

Weiler, S., Matt, C., Hess, T., 2022. Immunizing with information – Inoculation messages against conversational agents' response failures. Electronic Markets 32, 239–258. https://doi.org/10.1007/s12525-021-00509-9