Doctors' Dilemma – Understanding the Perspective of Medical Experts on AI Explanations

Aycan Aslan
University of Goettingen, aycan.aslan@uni-goettingen.de

Maike Greve
University of Goettingen, maike.greve@uni-goettingen.de

Marvin Braun
University of Goettingen, marvin.braun@uni-goettingen.de

Lutz M. Kolbe
University of Goettingen, lutz.kolbe@wiwi.uni-goettingen.de

Doctors' Dilemma – Understanding the Perspective of Medical Experts on AI Explanations

Completed Research Paper

Aycan Aslan
University of Goettingen
Humboldtallee 3, 37073 Goettingen
aycan.aslan@uni-goettingen.de

Maike Greve
University of Goettingen
Humboldtallee 3, 37073 Goettingen
maike.greve@uni-goettingen.de

Marvin Braun
University of Goettingen
Humboldtallee 3, 37073 Goettingen
marvin.braun@uni-goettingen.de

Lutz M. Kolbe
University of Goettingen
Humboldtallee 3, 37073 Goettingen
lutz.kolbe@wiwi.uni-goettingen.de

Abstract

As a solution for the pressing issue in medicine of "black-box" artificial intelligence (AI), models that are hard to understand, explainable AI (XAI) is gaining in popularity. XAI aims at making AI more understandable by explaining its working, e.g., through human understandable explanations. However, while prior research found that such explanations must be adapted for the given expert group being addressed, we find limited work on explanations and their effect on medical experts. To address this gap, we conducted an online experiment with such medical experts (e.g., doctors, nurses) (n=204), to investigate how explanations can be utilized to achieve a causal understanding and respective usage of AI. Our results demonstrate and contribute to literature by identifying transparency and usefulness as powerful mediators, which were not known before. Additionally, we contribute to practice by depicting how these can be used by managers to improve the adoption of AI systems in medicine.

Keywords: Explainable AI, Local explanations, Medical explainable AI, Causability

Introduction

Arguably, artificial intelligence (AI) is among the most influential technology trends of our time. As often discussed, AI offers a range of promising opportunities that is expected to impact many areas of human life, including medicine. Here, the potential of AI is evident, for example, in the form of medical image analysis (Mehta and Pandit 2018) or the use of natural language processing to extract information from clinical notes in an automated form (Roski et al. 2014). Accordingly, the global market for AI in medicine is steadily growing. Estimates show that the global medical AI market is slated to expand to around 67 billion US-dollars by 2027 (ReportLinker 2021).

Nevertheless, this described growth of potential and use cases of AI is accompanied by a growing complexity of the underlying AI algorithms. The increase in the performance of such AI-based systems usually come from more complex and sophisticated algorithms, which are capable of analyzing a larger body of data. In this context, especially high performant AI models, such as neural networks are deemed to be highly complex and therefore appear as "black-boxes". Such models are coined "black-boxes", as the processes that lead to their internal learning as well as resulting outcomes are, parts or completely, not comprehensible (Meske et al. 2020). While such "black-box" models are often more performant, they come at the cost of explainability. This lack of explainability poses a particular obstacle in the medical domain since applications of AI can have potentially severe consequences, such as treatment decisions or diagnostic choices (Meske et al. 2020). Consequently, the provision of explainability in such models, is one the biggest challenges for the diffusion of AI for the medical domain (Holzinger et al. 2020). Prior research stated, that medical experts must have the possibility of understand how and why a certain decision by an AI algorithm was made (Holzinger et al. 2017). Being able to support the medical experts in answering these questions with sufficient technical measures is crucial to build, trust and comprehension in the AI systems and therefore ensure long-term success of such (Meske et al. 2020; Woodcock et al. 2021).

As a possible solution to this pressing challenge of AI, explainable AI (XAI) is gaining in popularity (Meske et al. 2020). Explainable AI generally tries to make the AI system more transparent by explaining the inner working of it and hence unrevealing the "black-box" (Meske et al. 2020). While this goal can be achieved by different technical instruments, one way to implement XAI is by providing usable explanations for the respective outcome of the AI system (Holzinger et al. 2017, 2020). As such explanations are clearly addressed to the end-user of the system, prior research has showed that the user-perspective in designing value-adding explanations are highly important (Ribera and Lapedriza 2019). Studies demonstrate the importance of adapting explanations for the given target group and context, since different groups have varying needs and preferences (Gregor and Benbasat 1999; Ribera and Lapedriza 2019). For example, clinicians show to have a strong need to see supporting information for the given explanation and to understand the problem-solving strategy of the system (Schoonderwoerd et al. 2021). Consequently, involving the perspective of medical experts in designing AI explanations is crucial to incorporate respective needs and preferences that are unique to their domain. Nevertheless, to this date, there is a lack of medical expert involvement in the design of AI explanations (Barda et al. 2020). The definition of what constitutes a "good" explanation is often determined by model developers. These are more likely to focus on technical challenges, rather than having and contributing deep domain knowledge with respect to challenges (such as complexity of medical processes, level of detail needed, or understanding of medical workflows) present in the medical domain.

Consequently, our goal is to utilize a medical domain-specific perspective on how to design AI explanations to ensure perceived causal understanding of AI-based systems in medicine. For this, we build on prior work on AI explanations in other fields and identify transparency and usefulness as factors that are shown to have an influence on causal understanding. Additionally, while we highlight the importance of causal understanding of the AI system, we argue that understanding is not the end-point of the analysis but rather the basis for behavioral intention, such as the use of the AI system. Against this background, this paper investigates the following research question (RQ):

***RQ**: What are antecedents in the medical domain that influence how AI explanations lead to causal understanding and ultimately usage intention?*

To answer this research question, we will draw on information systems (IS) literature to derive different forms of AI explanations. These will be adapted with respective to current XAI literature, which highlights the importance of explaining single AI outcomes with the goal of generating a causal understanding (Holzinger et al. 2019). Consequently, the designed AI explanations are tested in a vignette-study through an online experiment with participants that work in the medical domain (e.g., doctors, nurses) (n=204), to investigate the influence of AI explanation perception of medical experts to deepen the understanding on their specific needs and preferences. To achieve this goal, we conducted a workshop with three doctors to design the respective vignette-study to ensure that the medical case and respective information are presented appropriately. In this context, we utilize the signaling theory as our theoretical lens to derive how explanations can serve as signals for the overall AI system. We demonstrate, that to achieve causal understanding for medical experts, AI explanations must address the need for transparency and usefulness. These are antecedents on causal understanding, which in turn influences the intention to use the AI system. We contribute to literature by identifying mediating effects to achieve causability in the medical domain and to practice by showing how explanations can be used strategically to increase the use of AI systems by medical experts.

Research Background

This section will provide the theoretical and conceptual background for our work. For this purpose, we first introduce the discourse on explainability of system in IS research. This serves as a foundation to adapt the classification of explanations provided by Gregor and Benbasat (1999) for the subsequent experiment. Additionally, we discuss research on AI explanations in general and specifically as used for medical experts to highlight how this study build on and contributes to previous research.

Explainability in IS Research

While the discussion on explainability is growing in the context of AI, the general study of explainability is not new to IS research. With the rise and growth of functionalities of technology in the 1980ies and 1990ies, IS research started to investigate the need for explanations in the interaction with such systems (Mao and Benbasat 2000; Ye and Johnson 1995). In this context, prior IS research elaborated many factors that influence how different target groups perceive certain explanations for information systems. Studies generally prove the positive effect of explanations on the overall understanding of the user with regard to the systems it interacts with (Dhaliwal and Benbasat 1996; Mao and Benbasat 2000). This effect on the understanding of the system have been found to further influence the decision making and the system perception by the user. It has been shown that explanations can alter the perception of the systems usefulness, ease of use, satisfaction, and ultimately trust positively (Dhaliwal and Benbasat 1996; Mao and Benbasat 2000; Ye and Johnson 1995). Hence, the body of work in IS research shows, how explanations can be utilized strategically, to achieve a more positive user perception of the system.

Further, IS research also offers frameworks for classifying different types of explanations. This includes the classification by Gregor and Benbasat (1999). In their work, they classify explanations on multiple dimensions to provide a profound understanding on the respective benefits of the explanation and how the respective explanation should be constructed (Gregor and Benbasat 1999). They classify explanation types based on three dimensions: Content, presentation format, and provision mechanism. While the presentation format describes whether the explanation is delivered as text-based or multimedia and the provision mechanism whether the explanations is provided for example automatically, we will focus on the classification of the content itself. Gregor and Benbasat (1999) classify the content in four types: 1) *Trace or line of reasoning*, 2) *Justification or support*, 3) *Control or strategic*, and 4) *Terminological*.

The first type of explanation is *trace or line of reasoning*. It summarizes explanations which explain why certain decisions were or were not made by explicitly referencing the data and/or rules used for the particular case. The second type is *justification or support* and describes explanations which justify part of the reasoning process of the system by linking the explanation provided to the deep knowledge from which it was derived. Third, *control or strategic* summarizes explanations that describe the control behavior and/or the problem-solving strategy of the system. Fourth and last, *terminological* summarizes explanations that supply some form of definitional or terminological information.

In summary, we note that there is a rich and sound body of work on explanations in IS research. As prior IS literature already hinted on, this classification can be utilized in the context of AI explanations (Schneider and Handali 2020). Consequently, we will adapt the classification on explanations by content type provided by Gregor and Benbasat (1999) to design our AI explanations for the later experiment.

Research on AI Explanations

In XAI literature, explanations are generally split into global and local explanations (Plumb et al. 2018). Global explanations describe information related to the overall logic of the AI model such as the fundamental structure of the model, underlying assumptions, and parameters (Zhang et al. 2021). They include all the representations that are learned by the model, hence helping users to understand the underlying structures of the system (Ribera and Lapedriza 2019) and therefore the overall behavior (Plumb et al. 2018). While described global explanations are more general for the overall AI model, local explanations are more specific in its statements. Local explanations explain the rationale for individual predictions (Zhang et al. 2021), hence promoting the understanding of small parts of the AI model (Elshawi et al. 2019). By focusing on specific outputs, local explanations allow users to understand better the reasons why the specific output of the AI model occurred (Ribera and Lapedriza 2019).

However, both global and local have advantages and disadvantages making them suitable for different situations. Prior research has established that different context, e.g., with regard to different goals or differences in prior knowledge, require different forms of explanations (Holzinger et al. 2017; Ribera and Lapedriza 2019). In this context, studies show that generally, domain experts (in contrast to lay users) should be provided with local explanations for two main reasons (Ribera and Lapedriza 2019). First, given the context of their work, domain experts, such as medical experts, are interested in clear information (Ribera and Lapedriza 2019). This is usually achieved more easily with local explanations, as it gives them actionable insights for their work. Second, as domain experts are specialists in the area of expertise where the decisions of the AI system take place, they are familiar with contextual information such as dependencies. This general understanding of the situation strengthens the demand for explanations which provide a high level of detail, such as local explanations that are tailored to specific outputs of the system (Ribera and Lapedriza 2019).

Despite this shown need to adapt explanations for the given domain, we note that there is only limited prior work on explanations that are tailored for the medical domain. Here, the limited scope investigates explanations for lay users such as patients (e.g., Zhang et al., 2021) and for medical domain experts (e.g., Schoonderwoerd et al., 2021).

For the former group, prior studies investigated how explanations can help patients to overcome their limited medical knowledge and build trust in AI systems (Alam and Mueller 2021; Woodcock et al. 2021; Zhang et al. 2021). In this context, Woodcock et al. (2021) analyze the provision of AI explanations in the context of an AI driven symptom checker that is used by patients, with the goal of exploring how different forms of explanations by a symptom checker affect the trust of patients (Woodcock et al. 2021). They find, that the sentiment of patients is dependent on the type of disease and that with little prior knowledge, explanations get more important and lead to trust in the system (Woodcock et al. 2021). Similarly, Zhang et al. (2021) analyzed the value of explanations for the task of patients comprehending radiology reports, with the goal of understanding the impact of transparency and model performance in the context of explanations (Zhang et al. 2021). The authors find that revealing the model performance can promote trust and perceived usefulness of the system (Zhang et al. 2021). Further, Alam and Müller (2021) investigate explanations in the context of patient-facing explanatory diagnosis system. The goal of their study was to understand the impact of explanations in a diagnosis scenario (Alam and Mueller 2021). Here they find that explanations can help to improve overall satisfaction measures (Alam and Mueller 2021).

However, while these shown studies lay a basic understanding on how explanations in the medical domain are perceived, they do not provide useful actionable insights for the use by medical experts. The shown studies explore the use of explanations with patients, why their findings cannot be mapped to medical experts. As demonstrated by Woodcock et al. (2021) as well, prior knowledge in the domain plays an important role, in how explanations are perceived, a characteristic in which patients and medical experts differ drastically. Contrary to patients, medical domain experts process the information they get, for example an explanation, in the context of a huge body of domain knowledge, yielding a special need to explore explanations specifically for them. Nevertheless, we could find very limited work that focuses specifically on the perspective of medical experts in AI explanations. In this context, Schoonderwoerd et al. (2021) conducted a study on understanding needs and preferences of clinicians. They show that clinicians want to know about the information that the system used to make the respective diagnosis. Additionally, clinicians want to see supporting or contradicting information for the given explanations and wish to know how certain the system was in the given diagnosis (Schoonderwoerd et al. 2021). Further, the authors find that clinicians want to understand the problem-solving strategy of the system in the form that the explanations addresses potential differential diagnosis (Schoonderwoerd et al. 2021). However, while these findings are highly interesting and valuable for the design of explanations for medical experts, their explanatory powers are limited, as the study was a qualitative study only conducted with six clinicians (Schoonderwoerd et al. 2021). Therefore, the shown findings form a solid starting point, but require a more in-depth investigation.

In sum, while we note different methods to provide explainable AI, prior literature emphasizes the importance of local explanations for domain experts, such as experts in the medical domain. While other groups in the medical domain (such as patients) have been addressed in prior studies, the perception of medical experts on AI explanations have not been covered sufficiently yet. Therefore, we note the need for a deepened understanding on how local explanations, as these are shown to be sufficient for domain experts, are perceived by medical experts, which will be addressed by this study.

Linking AI Explanations and Causability

While the prior sub-section gives an understanding on what kind of explanation to design, the overall goal of providing AI explanations is addressed by the concept of *causability*. The notion of causability, emerged from the discussion what explainability of an AI system really means (Holzinger et al. 2019). In this discussion, Holzinger et al. (2019) pointed out, that explainability itself is a property of a system, which is not very helpful when trying to understand the user-perspective on explanations. Therefore Holzinger et al. (2019), proposed causability, which represents a property of person, rather than a system. In essence, causability describes the level of generated causal understanding of a human through a provided explanation (Holzinger et al. 2019). This causal understanding is defined by the dimensions of effectiveness, efficiency and satisfaction in a specific use context (Holzinger et al. 2019). Hence, a causal understanding based on these dimensions refers to a human understandable model that is to be achieved through explanations provided (Holzinger et al. 2019). Consequently, causability represents a measurement for the quality of explanation delivered (Holzinger et al. 2019). The better the explanation, the better will be the generated causal understanding of the user, i.e., causability.

In this context, prior research discussed different factors that might influence the causability for a given context. As gaining an understanding of the underlying relationships of the AI system is a crucial part of causability, *transparency* has been discussed as a possible factor to link AI explanation and causability and to ultimately influence the level of causability perceived. The rationale being that making the systems and the respective relationships transparent is the foundation for any advanced (causal) understanding for humans. Therefore, prior studies have highlighted the role transparency can play as a lens to link to AI explanations and causability. Shin (2021) shows, that transparency has a positive influence on the causal understanding of users in the context of communication, as users can utilize the increase in transparency to have a better understanding of the underlying causal relationships. Additional to transparency, the *usefulness* of the system has also been highlighted as a possible lens to link AI explanations and causability. Prior research has shown that explanations perceived as useful lead to higher levels of causal understanding, as the users of the system see a clear benefit by the given explanation for the respective task (Gefen and Straub 2000).

In conclusion, we note that AI explanations should be designed and provided with the goal of generating a respective causal understanding, i.e., causability for the user that is exposed to the explanation. In the context of causability, we identify transparency and usefulness as possible factors that might influence causability and therefore might serve as helpful constructs to deepen our understanding on the perception of AI explanations in the medical domain.

Understanding the Perception of Medical AI Explanations

As the research background demonstrated, we can conclude that the most value-adding explanation for medical experts are local explanations. Additionally, in designing such local explanations, the goal should clearly be increasing the causability of such experts. However, currently it is unclear how such causability can be achieved through local explanations in the medical domain. Hence, to fill this gap, in the following we will adapt the signaling theory as our lens and propose our research model which aims at shedding light on achieving causability in the medical domain.

Conceptualizing Explanations as Signals

To deepen our understanding on how explanations can influence the perception of the overall AI system, we will utilize the signaling theory as our theoretical lens. The signaling theory helps to understand how two parties, e.g., buyer and seller, address limited information with each other with regard to a product and service (Wells et al. 2011). The main tool in this interaction are signals or cues, that one party can use to convey information that is otherwise not or hardly observable (Wells et al. 2011). In this context, signals are extrinsic to the products are service itself, but rather product-related attributes that do not alter the fundamental nature of the product (Wells et al. 2011). Using the purchasing interaction between buyer and seller for a personal computer as an example, the price or the brand of the computer would be an extrinsic cue, i.e., signal, which do not change the intrinsic attributes of the computer. Nevertheless, these signals can be used to convey the buyer to buy the respective computer. The focal asymmetry of information between the parties is characterized by the signaling theory by two phenomenon that must be given: 1) *Pre-purchase information scarcity*, and 2) *Post-purchase information clarity* (Wells et al. 2011). Pre-purchase information scarcity describes that the consumer cannot access or interpret the product quality attribute prior to the signal (Wells et al. 2011). Post-purchase information clarity is achieved when the information clarity is higher than before the signal has been send (Wells et al. 2011). Finally, the goal of signaling is depicted through the *signal outcome*. The signal outcome aims to signal quality regarding the product or service to the less-informed party, to ultimately influence the desired outcome such as attitude to use the product or service, positively (Wells et al. 2011). As demonstrated by prior literature, the signaling theory is particularly suitable if the less-informed party has low product familiarity, why he finds it harder to judge intrinsic product attributes and therefore more easily understand extrinsic cues in the form of signals (Wells et al. 2011). There are a variety of strong empirical studies that support the usefulness of the signaling theory (Wells et al. 2011). For example, Wang et al. (2004) investigates privacy seal of approvals as signals to promote trust in online retailers. Further, Durcikova and Gray (2008) use the knowledge validation process as a signal to influence the overall knowledge quality.

These proven explanatory powers of the signaling theory can be used in our context to explore how explanations can function and enhance the use of the overall AI system. It can offer us a theoretical lens on not to view the AI system and explanations as separate, but rather in conjunction, where explanations can function as a cue for the overall AI system. Hence, mapping the parties to our context, we will view the interaction between AI developers and medical experts, since the AI developers design the AI system and explanation, and medical experts ultimately use the system and will be addressed by explanations. As stated, we will utilize the explanations as signals for the AI system. Explanations are suited for the role of signals as the medical expert have a difficult time to evaluate the intrinsic quality of the AI system due to the black-box nature of such. Hence, explanations can be used as signals to improve the understanding of the overall AI system and to debilitate the information asymmetry. Here, the asymmetry is characterized by pre-purchase information scarcity, as the medical experts lack the information to assess whether the given recommendation of the AI system was accurate. However, after receiving the signal, i.e., the explanation, post-purchase clarity is achieved since the explanation supports the medical experts in assessing whether the given recommendation was accurate. Ultimately, the signal outcome consists in the change in attitude towards the use of the AI system. The signal, in form of the explanation, should help to convince the medical experts of the AI system and ultimately persuade them to use it.

Deriving a Research Model

As one of the core goals of XAI in general and AI explanations in specific is to provide more clarity, the concept of transparency plays an essential role in the perception of explanations. Therefore, we argue that providing explanations for the recommendation of the AI system will influence the perceived transparency by the medical experts. The rationale being, that explanations help to "open" the "black-box" for them, whereby they can start to get a feel for how the system works. Explanations, for example in form of references to source data, also give the medical experts a transparent opportunity to peak into the decision-making of the AI system. The importance of transparency has also been noted by prior studies. For example, in their interview-based study Ochmann et al. (2021) find that a lack of transparency is one of the main obstacles for experts to adopt AI systems. Accordingly, multiple studies support the notion that explanations improve the transparency perceived (Bussone et al. 2015; Horsky et al. 2012; Meske et al. 2020), due to the stated reasons. For example, in the context of comprehending radiology reports, Zhang et al. (2021) shows that provided explanations lead to users feeling that the AI system is more transparent.

This overall goal of XAI matches with the goal of the signaling theory to achieve post-purchase clarity, through explanations. Hence, transparency, in the form that the recommendation is traceable for the user, is crucial to achieve higher levels of clarity after the signal, i.e., the explanation. In line with this statement, prior studies investigated the effect of transparency in the context of the signaling theory. For example, Durcikova et al. (2008) demonstrated how transparency can be utilized to achieve such post-purchase clarity after a signal, in the context of knowledge repositories. Thus, building on the central goals of AI explanations and the need for post-purchase clarity, we hypothesize that:

> **Hypothesis 1**: The presentation of an AI explanation will have a positive influence on the perceived transparency with respect to how the AI system came to that given medical recommendation.

As stated, in line with the signaling theory, we aim at evaluating the explanation as part of the overall AI system. As such, we note that the sole purpose of the overall AI system is to support the expert in their medical work. Consequently, we argue that providing an explanation for the recommendation given will influence the perceived usefulness of the AI system. As the explanation will provide more information on why the respective recommendation was shown, the medical expert will perceive the recommendation to be more informative and therefore useful for the medical case at hand. This argument is further strengthen, when considering that AI-based systems are often used for decision-support (Meske et al. 2020). In the context of decision-support systems, explanations will help the medical experts in providing additional information, which are than applicable for their medical work context. In line with this argument, prior literature investigated the role that explanations have on the perceived usefulness of AI systems. For example, Zhang et al. (2021) showed the influence of revealing information in form of explanations to the users. They find, that explanations, in the form disclosing the model performance, positively influences the perceived usefulness of the system as users value the additional information (Zhang et al. 2021). Building on these findings, we hypothesize that:

> **Hypothesis 2**: The presentation of an AI explanation will have a positive influence on the perceived usefulness of the medical recommendation given by the AI system.

As stated in the research background a main component of causability is that medical experts are in the position to trace causal relationships derived by the AI system (Holzinger et al. 2019, 2020; Shin 2021). Hence, we argue that higher levels of transparency will influence the perceived level of causability. As transparency, achieved through an explanation, gives the medical expert the possibility to recognize which causal relationships with respect to the present medical information exist and are in play. For example, an explanation that displays the problem-solving strategy of the AI system, helps the medical experts to understand the sequence of decision made by the system, hence helping the expert in understanding present causal relationships among different medical documents and medical information in such documents. This effect of transparency on causability has been analyzed by prior studies as well. For example, Shin (2021) demonstrate, that in the context of media recommendations, present transparency of the recommendation has a positive influence on causability of the user. Therefore, based on prior literature investigating the relationship between the perceived transparency of a given recommendation and causability, we hypothesize that:

> **Hypothesis 3**: Higher levels of transparency regarding the medical recommendation will lead to an increase in causability.

As stated in the research background, causability is a multi-dimensional notion (Holzinger et al. 2019, 2020). One of these dimensions relevant for the causal understanding is effectiveness (Holzinger et al. 2019, 2020). In this context, Holzinger et al. (2019) highlight that the dimensions must be seen through the lens of the specific use context, as such context will influence the perceived causal understanding. In this regard, we argue that the perceived usefulness will have an influence on causability, as its specifically addresses the dimension effectiveness regarding the medical work of the expert. If medical experts perceive the recommendation of the system to be useful it will help them to generate a causal understanding, as they see a benefit in the recommendation for their medical work. Therefore, higher levels of usefulness have an effect on the effectiveness dimension of causability. This relationship between usefulness and effectiveness in the context of technology use is supported by prior studies. For instance, Gefen and Straub (2000) demonstrate the influence of perceived usefulness on efficiency in the context of E-commerce adoption. Thus, reflecting on the usefulness of explanations for the medical work of the experts and building on prior literature on the effect of perceived usefulness on perceived effectiveness, with regard to the influence of usefulness on causability, we hypothesize that:

> **Hypothesis 4**: Higher levels of usefulness regarding the medical recommendation will lead to an increase in causability.

Looking into the current literature on AI explanations, we note that usually causability is the end-point of the analysis. Therefore, most of the studies addressing causability analyze the antecedents of such (Shin

2021). While without a doubt causability is a crucial factor in understanding the effect of explanations, which is why it is also included in our model, we argue that the real goal should be to influence the users, with the help of explanations, to actually *use* the provided AI system.

This conviction is supported by the signaling theory, which calls for a change in attitude as the ultimate goal for providing signals. Consequently, prior literature utilizing the signaling theory have analyzed signals that ultimately affect the intention to use a product or service. For example, in the context of health intermediaries, Song and Zahedi (2007) show how signals such as third party seals and health infomediary reputation, ultimately have on the intention to use the health infomediary. In line with our argument that AI explanations should help to increase the use of AI systems and the signaling theory that calls for signal outcomes in form of an attitude change, we hypothesize that:

> **Hypothesis 5**: Higher levels of causability will lead to an increase in the intention to use the AI system.

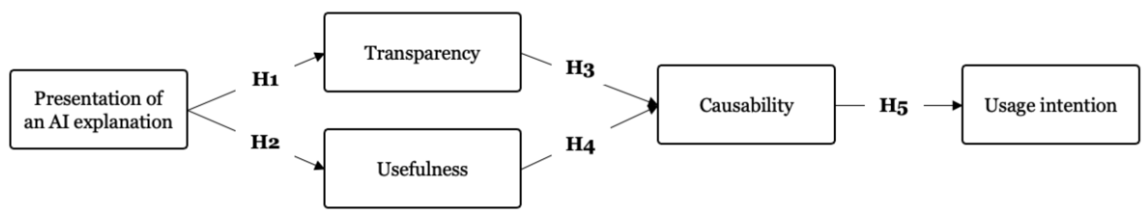

Figure 1. Proposed Research Model

The full research model is depicted in Figure 1.

Research Design and Experimental Setting

To validate the derived hypotheses, we conducted a vignette study with three different explanations and one control group in an online scenario experiment. As we aimed at building a sound medical scenario, we consulted three doctors: An anesthesiologist, an ophthalmologist, and a psychologist. In a joint workshop the scenario design was derived. This was done to ensure that the medical information provided in the medical scenario are logical and can be understood by medical experts. The participants for the online experiment were recruited from Prolific. We used the criterium that only participants active in the medical sector can participate in our study. After removing 19 responses due to being incomplete or missing build in attention checks, the final sample contained 204 survey responses. The average age of the participants was 39 years, of which 163 were women (79.9%), 38 were men (18.6%), 2 preferred not to say their gender (1.0%), and 1 indicated that they have another gender then male or female (0.5%). In this group of participants, we had a strong representation of doctors. Out of the 204 participants, 115 are practicing doctors (56.4%). Further, 35 participants are nurses (17.2%), while six participants are active in medical research (2.9%). Additionally, 48 participants are grouped into other medical activities, such as pharmacy or unspecified activities in the hospital (23.5%). The average reported years of medical expertise are 18.5 years. Additionally, we asked for the IT and AI knowledge of the participants. Out of all participants, 123 (60.3%) stated that they would rate their IT knowledge as high or very high. However, 124 participants (60.8%) stated that they would rate their AI knowledge as medium or high. We surveyed and obtained constructs (see Table 2), control variables, and demographic information. Manipulation and respectively attention checks were performed to ensure participants were able to relate to the shown medical scenario. The survey was conducted in March 2022.

Introductory medical scenario
Please imagine that you are a primary care doctor. As of lately, you recognized that more and more patients consult you for a secondary opinion. However, in this context, the patients send over a growing body of medical documents (in a digital format) with regard to e.g., the history of the present illness, their past medical history or laboratory tests. To cope with this growing body of medical information, which you have to consider when consulting the patients, you decide to introduce an artificial intelligence (AI)-based system, called AuMEDa, that will support you by analyzing a vast amount of documents and medical information for each respective patient. Ultimately, the AI system gives you a recommendation for the diagnosis of the disease, based on the conducted analysis of medical information.
Table 1. Medical scenario setting

The survey started for each participant with the introduction (see Table 1). The participants were asked to put themselves into a situation where they adopt an AI system as a primary care doctor. The AI system is supposed to help them to analyze the growing body of medical documents they must analyze to consult patients. After feedback from the three consulted doctors, we opted to specify that the case at hand is a secondary opinion, meaning that the patient already consulted a doctor and got a diagnosis but wants a secondary opinion by another doctor.

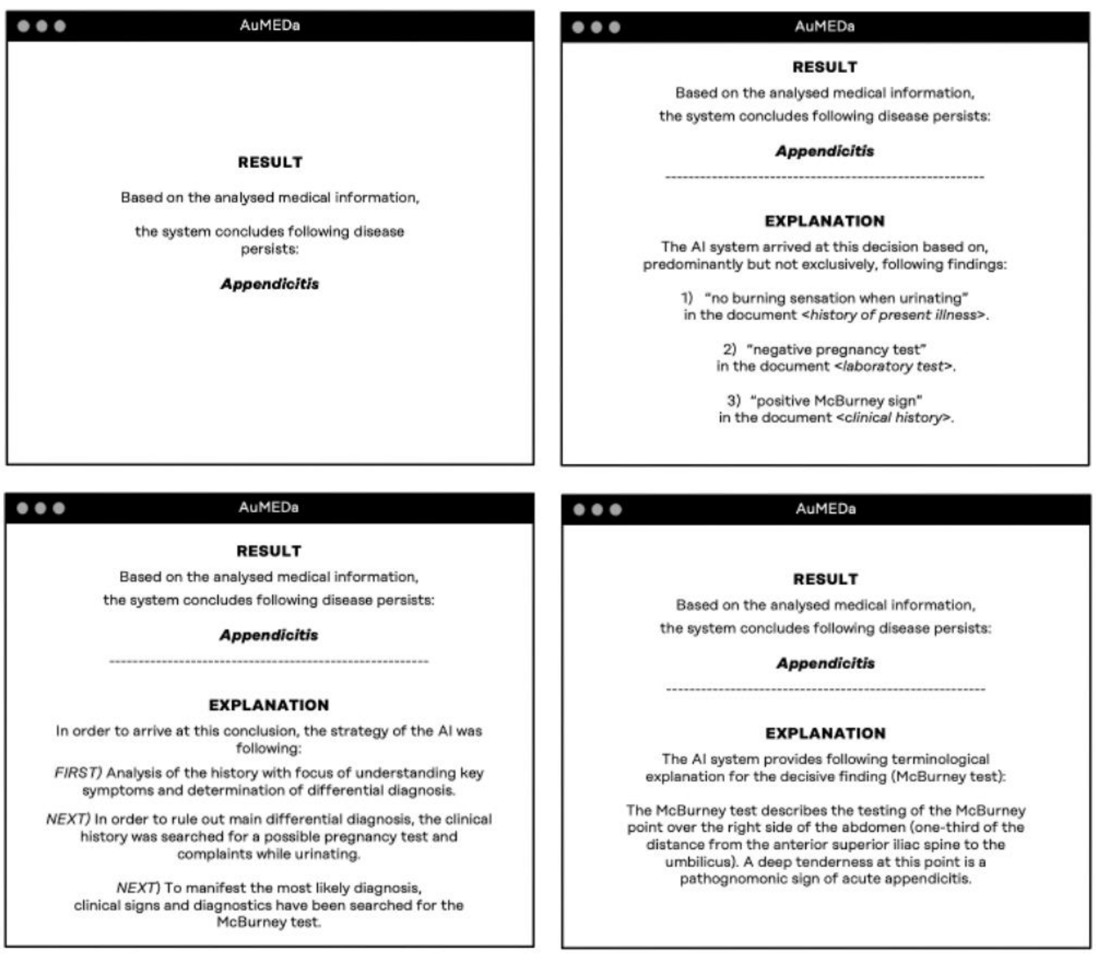

Figure 2. Control scenario (top left), trace explanation (top right), strategic explanation (bottom left), and terminological explanation (bottom right)

The consulted doctors told us, that especially in the context of secondary opinions, patients consult them with a large body of medical documents from prior doctor or hospital visits. Additionally, patients often ask for a secondary opinion via email and do not visit the respective doctor from whom they want a secondary opinion in person. Hence, the only information the doctors can rely on are in the medical documents send to them.

Following the introductory scenario, all participants were introduced to specific medical case for a secondary opinion. This consisted of 25-year-old female who complains about abdominal pain (Specific text shown to the medical experts: “Given this described context, you get the request for a secondary opinion for a 25-year-old female who complains about abdominal pain for two days.”). Once again, this specific medical case was design in close collaboration with the three consulted doctors. Their feedback indicated that this is the level of information that is representative for a secondary opinion, without looking into the medical documents send. Additionally, we were told that a complaint such as abdominal pain happens fairly frequently and leaves some room for differential diagnosis, which gives a raison d'être for the AI system. Before shown the outcome of the AI system, the participants see following final information: “After using your AI system, which has analyzed the vast past medical information for the female including the doctor notes from her doctor’s visit last week, you get following recommendation of the AI system”.

After this information, participants are randomly assigned to one of the explanation scenarios (see Figure 2). The first one being the control scenario, where the participants are shown that the systems diagnosis “appendicitis”, without further information. With regard to the used explanations, we draw on the classification proposed by Gregor and Benbasat (1999). However, as we have an initial understanding on the needs of medical experts provided by Schoonderwoerd et al. (2021), we will include only three types of explanation in our analysis: 1) Trace, 2) Strategic, and 3) Terminological. We made this selection as we believe that these types best reflect the type of information needed by medical experts (Schoonderwoerd et al. 2021). Hence, for the trace explanation scenario we display the reasons the diagnosis of appendicitis by referencing to the respective features and source medical documents. Next, the strategic explanation addresses the problem-solving strategy of the system by showing how it proceeded. Finally, the terminological explanation provides a terminological explanation for the decisive medical finding. To check, whether the participants noticed the difference between the control scenario and the respective explanation scenario, we conducted manipulation checks. The results differ significantly between the participants that received the scenario with the trace explanation and control ($t(92)=4.173$, $p<.001$), strategic explanations and control ($t(89)=8.993$, $p<.001$), and terminological explanation and control ($t(89)=7.071$, $p<.001$). The respective medical information provided in each explanation scenario, has once again been designed with the feedback of the three doctors, to ensure that relevant medical checks and information with respect to abdominal pain is covered.

Data Analysis and Results

Due to the novel investigation with regard to the perception of AI explanations by medical experts, we deployed the partial least squares structural equation modeling (PLS-SEM) approach (Hair et al. 2011), to test the presented research model (see Figure 1). PLS-SEM is helpful for our context as we can analyze the effects between latent variables (Goodhue et al. 2012). Additionally, it enables us to analyze complex models, despite smaller sample sizes (Fombelle et al. 2016). This is particularly helpful in our study, where only medical experts were allowed to participate, which significantly reduces the pool of potential participants. The software used to conduct the partial least squares structural modeling approach, was Smart-PLS 3.0 (Hair et al. 2021).

Measurement Model Results

For the measurements of our constructs, all items were adapted based on relevant literature. Nevertheless, we did adjust them to our medical context. To follow presented best practices presented in prior literature, we assessed the validity and reliability of the model (Fornell and Larcker 1981). Hence, ensuring indicator reliability, internal consistency reliability, convergence validity, and discriminant validity. To address the indicator reliability of our model, we checked that the indicators of our used constructs explain more variance than the measurement error (Hair et al. 2019). This is given for all used items, except item three and seven for the construct *causability* (see Table 2). Consequently, these two items were dropped. Moreover, we establish internal consistency reliability by examining Cronbach's alpha and composite reliability. Both of these criteria are fulfilled if the threshold of 0.7 is met (Hair et al. 2021; Nunnally and Bernstein 1994). Table 4 indicates that this criterion is met for all used constructs. Further, we evaluated the convergent validity by examining the average variance extracted (AVE). With regard to the used indicators, the constructs should explain at least 50% of the variance (Henseler et al. 2009), which is met as well.

Constructs and Items	**Loadings**
Transparency (α = .887, CR = .918, AVE = .691) (Schnackenberg et al. 2021)	
The recommendation provided by the AI system is well traceable.	.867
The recommendation provided by the AI system is clear.	.742
The approach for the recommendation of the AI system was transparent to me.	.854
The relevant information with regard to the recommendation of the AI system have been provided.	.882
I have all the information I need with regard to the recommendation provided by the AI system.	.804
Usefulness (α = .902, CR = .939, AVE = .837) (McKinney et al. 2002)	
The recommendation of the AI system was informative.	.886
The recommendation of the AI system was valuable.	.920
The recommendation of the AI system was useful.	.937
Causability (α = .848, CR = .888, AVE = .570) (Holzinger et al. 2020)	
I found that the recommendation included all relevant known causal factors with sufficient precision and granularity.	.708
I understood the recommendation within the context of my work.	.822
I did not need support to understand the recommendation.	*.492*
I found the recommendation helped me to understand causality.	.730
I was able to use the recommendation with my knowledge base.	.816
I think that most people would learn to understand the recommendation very quickly.	.748
I did not need more references in the recommendation: e.g., medical guidelines, regulations.	*.599*
I received the recommendation in a timely and efficient manner.	.630
Usage intention (α = .944, CR = .964, AVE = .901) (Bhattacherjee and Sanford 2006)	
I intend to use the AI system on my job in the next time.	.967
I intend to use the AI system on my job in the near future.	.972
I intend to use the AI system for more of my job responsibilities.	.906

α – Cronbach's alpha, CR – Composite Reliability, AVE – Average Variance Extracted. Items were measured on a 7-point Likert scale (1 = Strongly disagree, 7 = Strongly agree). Dropped items are indicated in italics.

Table 2. Construct Measurement, Reliability, and Convergent Validity

	Transparency	**Usefulness**	**Causability**	**Usage intention**
Transparency	**0.831**	*0.771*	*0.801*	*0.495*
Usefulness	0.690	**0.915**	*0.847*	*0.637*
Causability	0.703	0.748	**0.755**	*0.698*
Usage intention	0.455	0.590	0.634	**0.949**

FL-criterion in bold, HTMT in italics

Table 3. Discriminant Validity

Furthermore, we verified our model by applying the Fornell and Larcker (1981) criterion (FL) and the heterotrait-monotrait ratio (HTMT) (Hair et al. 2021). According to Hair et al. (2021), the FL-criterion is met if the squared AVE is greater than the correlations with all other constructs of the model. This is met for our measurement model accordingly (see Table 3). The HTMT ratio is met if values are below the threshold of 0.9 (Henseler et al. 2015). The discriminant validity of our model holds. We also tested for the existence of a common method bias, by following the method suggested by Kock (2015), as we employed a threshold of 5.0 for the variance inflation factor (VIF), which is met for all the VIF values.

Structural Model Results

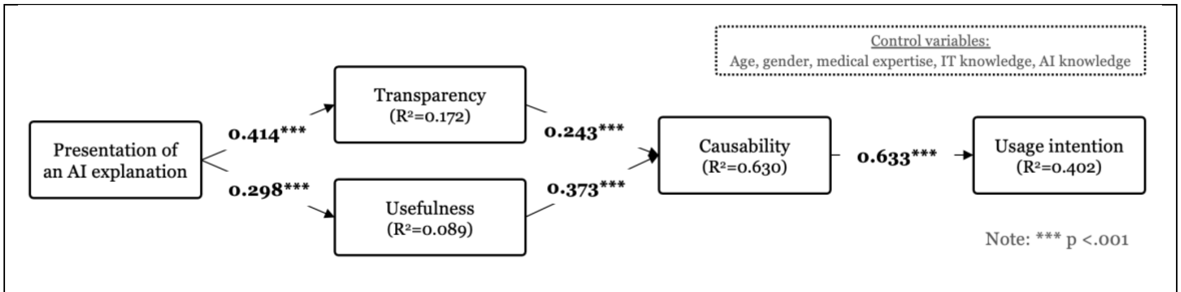

Figure 3. Proposed Research Model

With the goal of testing our measurement model and evaluate the significance of our hypothesized paths, we ran the bootstrap method with 10,000 samples (Hair et al. 2021). The results lend support to all of our five hypotheses (see Figure 3). Looking into the effects of providing an AI explanation, we note that explanations have a positive effect on transparency (β=.414, p<.001) and usefulness (β=.298, p<.001). Therefore, we can conclude that **H1 and H2 are supported**. Analyzing the effects on causability, we note that both transparency (β=.243, p<.001) and usefulness (β=.373, p<.001) have a positive significant effect on causability. Hence, **H3, and H4 are supported.** Finally, we find positive effect of causability on usage intention (β=.633, p<.001), with which we can state that **H5 is supported** as well.

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

Post-Hoc Analysis

To further investigate whether there are differences in the perception of the different explanations (trace explanation vs. strategic explanation vs. terminological explanation) we conducted an ANOVA. However, we couldn't find significant differences of the three explanation types on our analyzed constructs transparency and usefulness. Moreover, we tested the effect among the explanations on *relevance* (adopted from McKinney et al. 2002), which depicts how relevant and related the recommendation was with regard to the medical work and *competence* (adopted from Gulati et al. 2019), which depicts on how competent the AI system is perceived. However, in line with our analysis of transparency and usefulness, we couldn't find significant differences on these constructs regarding our explanation types.

This post-hoc analysis strengthens the notion that medical experts appreciate the provision of an explanation, however, are not selective in the sense that one explanation is clearly preferred over another (among the explanation types tested in this study).

Discussion

As prior literature neglected the perspective of medical experts on AI explanations, our study addresses this gap by showing how AI explanations can lead to a higher level of causability. Our experiments highlight the importance of transparency and usefulness in the context of providing AI explanations, as they have been shown to be strong factors to achieve causability. Additionally, we go one step further and demonstrate the influence of causability on the intention to actually use the AI system. This shows, how explanations can be used strategically to increase the use of AI systems. Nevertheless, as indicated by our post-hoc analysis, we did not find significant differences among the explanation types provided. In the following, we discuss the implications of these findings for literature and for practice and point out future research opportunities.

Contributions to Literature

Our study contributes to literature in several ways. First, we contribute to the research stream of AI explanations by showing how causability can be achieved in the medical domain. As shown in the research background, prior literature established the importance of causability for AI explanations (Holzinger et al. 2019, 2020; Shin 2021). In this context, prior studies also stated that a respective causal understanding is especially needed in the medical domain (Holzinger et al. 2019). However, to the best of our knowledge, to this date there wasn't a study conducted with medical experts that empirically showed how to achieve causability in the context of AI in medicine. Our study fills this gap, by indicating the importance of transparency and usefulness to achieve causability for medical experts. Hence, to achieve sufficient causability, when designing AI explanations, it should be payed attention to these factors, for example by ensuring that the given explanation is sufficiently transparent with regard to which medical information was decisive for the given recommendation. Thus, we enforce the importance of causability in the context of AI explanations and demonstrate with our study how such causability can be achieved in the medical domain.

Second, we contribute to the research stream of AI explanations by conceptualizing causability as an antecedent for usage intention. As the importance of causability has been shown, studies analyze how to achieve causability in different domains and contexts (Shin 2020, 2021). However, while we do not deny the importance of causability, we argue that the assessment of the effect of AI explanations need to go one step further. One of the core goals of XAI is to help users to build more trust into the systems they interact with (Meske et al. 2020). Here, especially in the medical domain there is criticism that a lack of explainability is hindering the adoption of such by medical experts (Alam and Mueller 2021; Elshawi et al. 2019). Therefore, we argue that it is only consequent to include the intention to use the system into the assessment of AI explanations. This need has been addressed by our study. Our findings show, that causability acts as a strong antecedent for usage intention. Thus, we combine prior literature on the importance of causability with the logical step of influence on the behavior in the form of usage intention. This combination and addition can act as a first step from a theoretical analysis in form of causability, into a more practice-oriented analysis, in the form of intention to use the AI system.

Third, we contribute to the research stream of AI explanations by understanding AI explanations through the lens of the signaling theory. As we analyzed prior literature on AI explanations in the research background section, it became evident that prior studies missed a clear underlying theory. This lack of theory causes issues, such as a lack of understanding of relationships in a context or a common understanding of what the actual goal of AI explanations should be. To fill this gap, we adopted the lens of the signaling theory as it gave us clear instruction on how the explanations can function as a successful signal of AI systems. These include the concepts of pre-purchase information scarcity and post-purchase information clarity, which highlighted that the explanation should function to minimize the information asymmetry existent. Further, the signaling theory offered us the change of attitude as the ultimate goal of a signal. Using this, we took the usage intention of the AI system into consideration of our analysis and ultimately were able to demonstrate how causability can be used to achieve usage intention of the AI system. These examples show, how signaling theory helped us to frame prior knowledge in AI explanation research more effectively (see pre-purchase information scarcity and post-purchase information clarity) and extend the scope to relationships that weren't addressed before (see goal of attitude change). Thus, it becomes apparent, that the signaling theory can be used as a powerful and fruitful theoretical lens for future AI explanation studies.

Practical Implications

Additional to the contributions to literature, our work has several managerial implications. First, our study highlights the importance for medical managers of acting now and *just* implementing a local explanation. As touched upon, there is a widespread discussion in academic literature on how to exactly design explanations, for example regarding the detail of information shown or the level of personalization ensured

(Kuhl et al. 2020; Schoonderwoerd et al. 2021). While these studies on the nuances of explanations are certainly valid and are proven to deliver an additional value, our study show that, given some basic requirements, multiple forms of explanations work in generating a causal understanding. These requirements consist in designing local explanations that address single outcomes, rather than global explanations, and including medical experts in the design process. Both factors, derived from prior literature, are given in our study and are proven to work. Hence, from a practical point of view, rather than waiting for the scientific discourse on the details of explanations for medical experts to settle, managers in the medical domain should act *now* and engage in implementing explanations, with respect to the stated requirements, that will add value to the work of medical experts immediately. Especially, when considering that the alternative is to continue using black-box models, which are proven to hinder the adaption in the medical domain through their lack of explainability.

Second, our findings highlight the importance of explanations on the intention to use the AI system. As the diffusion of AI promises great potential for the medical sector, there is a growing sentiment for medical managers to invest in AI to not miss the development towards a more digital and "intelligent" future of medicine. This is proven by the constant year-to-year growth in investments in the global medical AI market (ReportLinker 2021). However, while ever growing sums are invested in AI assets it remains largely unclear how to positively affect the adaption of such systems by the end-users, the medical experts. Here, our study shows that explanations are not only helpful to generate a causal understanding within the medical case at hand but can also positively influence the intention to use the AI system itself. Hence, explanations can be used by medical managers as a strategic asset, to influence the sentiment of medical experts towards AI systems. This promises to be a great strategic advantage for medical facilities that heavily invest in AI assets and hope to reap the benefits offered by such for their medical processes. Therefore, viewing explanations as a strategic asset and benefit from them for the adaption of AI systems, will be crucial for the future of medicine which will rely on a strong partnership between medical experts and AI systems.

Third, through the lens of the signaling theory we are able to view the AI system and the explanation in conjunction, that mutually reinforces each other. As illustrated, a lot of the most performant AI models and therefore most promising real-world applications are black-box models. Hence, if the goal is to benefit from such models, due to both regulatory measures as well as lack of user acceptance, there is no way around making such systems more explainable. Consequently, an increase in explainability is a *must-have* not a *nice-to-have*. In line with this assessment, we argue that AI systems and explanations shouldn't be viewed as separate entities, as it is present in current discourse, but rather as a conjunct system that needs but also strengthens each other. Here, the signaling theory and its explanatory powers come into play. We argue, to realize such a conjunct view, explanations should be viewed as signals for the AI system to reduce present information asymmetry and positively influence the perceived quality of the AI system. From a business and managerial perspective, this view opens multiple fruitful advancements. For example, rather than investing in a (black-box) AI asset and later wondering how to achieve higher levels of explainability, this described conjunct view would result in the realization that the AI system itself shouldn't be considered without a respective explanation mechanism. Thus, adopting such a conjunct view from a managerial perspective, would result in an attitude towards strategically investing in AI which is much more thoughtful and financially sustainable.

Limitations and Opportunities for Future Research

Besides the illustrated contributions, both to literature and practice, we note three areas in which future research could strengthen and extend our results. First, we recognize that in our experiment, the content type of explanations was limited. For example, it might be sufficient in another context to include the explanation type justification by Gregor and Benbasat (1999) or combine the content of the explanation types tested in this study. Future research could investigate the effects of such changes in the explanation content on the perception of medical experts. Second, we note that our study was only limited to changes in content type. However, as Gregor and Benbasat (1999) propose, explanations can also be adjusted regarding presentation format (e.g., text-based vs. multimedia) or the provision mechanism (e.g., user-invoked vs. automatic). Extending our understanding of AI explanations to these dimensions in future research can generate a more holistic view on how causal understanding can be achieved. Third, we recognize that our study was limited to a single medical case. However, prior research on explanations showed, that the given (medical) task can influence the type of explanation needed (Woodcock et al. 2021).

Hence, future research should extend our findings to a broader range of medical cases, e.g., in form of different diseases, to identify commonalities and differences.

Conclusion

In conclusion, our work provides valuable insights, both for literature and practice, into the perception of AI explanations by medical experts. In conducting a between-subject online experiment (n=204) with medical experts, we shed light on the needs of them, a perspective neglected in prior research, in AI explanations. We show that to achieve causability, i.e., higher levels of causal understanding, respectively transparency and usefulness must be considered in designing AI explanations. Additionally, we show the influence on the intention to use the system, as causability has been shown to be a strong antecedent of usage intention. Nevertheless, we also demonstrate that all explanations, irrespective of type, help to increase causability, pinpointing to a "*just implement any (local) explanation*"-approach rather than a "*doing-nothing*"-approach. These findings have important implications for literature as we reinforce the importance of causability in the context of AI explanations, by showing its explanatory powers for medical experts. For practice, our findings demonstrate how AI explanations can be used strategically for a better adoption of AI systems, which is particularly relevant when considering recently growing investments in medical AI.

References

Alam, L., and Mueller, S. 2021. "Examining the Effect of Explanation on Satisfaction and Trust in AI Diagnostic Systems," *BMC Medical Informatics and Decision Making* (21:1), BioMed Central, pp. 1–15.

Barda, A. J., Horvat, C. M., and Hochheiser, H. 2020. "A Qualitative Research Framework for the Design of User-Centered Displays of Explanations for Machine Learning Model Predictions in Healthcare," *BMC Medical Informatics and Decision Making* (20:1), BMC Medical Informatics and Decision Making, pp. 1–16.

Bhattacherjee, A., and Sanford, C. 2006. "Influence Processes for Information Technology Acceptance: An Elaboration Likelihood Model," *MIS Quarterly* (30:4), pp. 805–825.

Bussone, A., Stumpf, S., and O'Sullivan, D. 2015. "The Role of Explanations on Trust and Reliance in Clinical Decision Support Systems," *Proceedings - 2015 IEEE International Conference on Healthcare Informatics, ICHI 2015*, pp. 160–169.

Dhaliwal, J. S., and Benbasat, I. 1996. "The Use and Effects of Knowledge-Based System Explanations: Theoretical Foundations and a Framework for Empirical Evaluation," *Information Systems Research* (7:3), pp. 342–362.

Durcikova, A., and Gray, P. 2008. "How Knowledge Validation Processes Affect Knowledge Contribution," *Journal of Management Information Systems* (25:4), pp. 81–108.

Elshawi, R., Al-Mallah, M. H., and Sakr, S. 2019. "On the Interpretability of Machine Learning-Based Model for Predicting Hypertension," *BMC Medical Informatics and Decision Making* (19:1), BMC Medical Informatics and Decision Making.

Fombelle, P. W., Bone, S. A., and Lemon, K. N. 2016. "Responding to the 98%: Face-Enhancing Strategies for Dealing with Rejected Customer Ideas," *Journal of the Academy of Marketing Science* (44:6), pp. 685–706.

Fornell, C., and Larcker, D. F. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research* (18:1), p. 39.

Gefen, D., and Straub, D. 2000. "The Relative Importance of Perceived Ease of Use in IS Adoption: A Study of E-Commerce Adoption," *Journal of the Association for Information Systems* (1:1), pp. 1–30.

Goodhue, D. L., Lewis, W., and Thompson, R. 2012. "Does PLS Have Advantages for Small Sample Size or Non-Normal Data?," *MIS Quarterly* (36:3), pp. 981–1001.

Gregor, S., and Benbasat, I. 1999. "Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice," *MIS Quarterly: Management Information Systems* (23:4), pp. 497–530.

Gulati, S., Sousa, S., and Lamas, D. 2019. "Design, Development and Evaluation of a Human-Computer Trust Scale," *Behaviour and Information Technology* (38:10), Taylor & Francis, pp. 1004–1015.

Hair, J. F., Hult, G. T. M., Ringle, C., and Sarstedt, M. 2021. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, (3rd ed.), SAGE.

Hair, J. F., Ringle, C. M., and Sarstedt, M. 2011. "PLS-SEM: Indeed a Silver Bullet," *Journal of Marketing Theory and Practice* (19:2), pp. 139–152.

Hair, J. F., Risher, J. J., Sarstedt, M., and Ringle, C. M. 2019. "When to Use and How to Report the Results of PLS-SEM," *European Business Review* (31:1), pp. 2–24.

Henseler, J., Ringle, C. M., and Sarstedt, M. 2015. "A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling," *Journal of the Academy of Marketing Science* (43:1), pp. 115–135.

Henseler, J., Ringle, C. M., and Sinkovics, R. R. 2009. "The Use of Partial Least Squares Path Modeling in International Marketing," *Advances in International Marketing* (20), pp. 277–319.

Holzinger, A., Biemann, C., Pattichis, C. S., and Kell, D. B. 2017. *What Do We Need to Build Explainable AI Systems for the Medical Domain?*, pp. 1–28.

Holzinger, A., Carrington, A., and Müller, H. 2020. "Measuring the Quality of Explanations: The System Causability Scale (SCS): Comparing Human and Machine Explanations," *KI - Kunstliche Intelligenz* (34:2), Springer Berlin Heidelberg, pp. 193–198.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., and Müller, H. 2019. "Causability and Explainability of Artificial Intelligence in Medicine," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (9:4), pp. 1–13.

Horsky, J., Schiff, G. D., Johnston, D., Mercincavage, L., Bell, D., and Middleton, B. 2012. "Interface Design Principles for Usable Decision Support: A Targeted Review of Best Practices for Clinical Prescribing Interventions," *Journal of Biomedical Informatics* (45:6), Elsevier Inc., pp. 1202–1216.

Kock, N. 2015. "Common Method Bias in PLS-SEM: A Full Collinearity Assessment Approach," *International Journal of E-Collaboration* (11:4), pp. 1–10.

Kuhl, Ni., Lobana, J., and Meske, C. 2020. *Do You Comply with AI? -- Personalized Explanations of Learning Algorithms and Their Impact on Employees' Compliance Behavior*, pp. 1–6.

Mao, J. Y., and Benbasat, I. 2000. "The Use of Explanations in Knowledge-Based Systems: Cognitive Perspectives and a Process-Tracing Analysis," *Journal of Management Information Systems* (17:2), pp. 153–179.

McKinney, V., Yoon, K., and Zahedi, F. 2002. "The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach," *Information Systems Research* (13:3), pp. 296–315.

Mehta, N., and Pandit, A. 2018. "Concurrence of Big Data Analytics and Healthcare: A Systematic Review," *International Journal of Medical Informatics* (114:March), Elsevier, pp. 57–65.

Meske, C., Bunde, E., Schneider, J., and Gersch, M. 2020. "Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities," *Information Systems Management*, Taylor & Francis, pp. 1–11.

Nunnally, J. C., and Bernstein, I. H. 1994. "The Assessment of Reliability," in *Psychometric Theory* (3rd ed.), New York, NY: McGraw-Hill, pp. 248–292.

Ochmann, J., Zilker, S., Laumer, S., Ochmann, J., Zilker, S., and Laumer, S. 2021. "The Evaluation of the Black Box Problem for AI-Based Recommendations: An Interview-Based Study," *Wirtschaftsinformatik 2021*.

Plumb, G., Molitor, D., and Talwalkar, A. 2018. "Model Agnostic Supervised Local Explanations," *Advances in Neural Information Processing Systems* (2018-Decem:NeurIPS), pp. 2515–2524.

ReportLinker. 2021. "Artificial Intelligence in Healthcare Market by Offering, Technology, Application, End User and Geography - Global Forecast to 2027." (https://www.reportlinker.com/p04897122/Artificial-Intelligence-in-Healthcare-Market-by-Offering-Technology-Application-End-User-Industry-and-Geography-Global-Forecast-to.html).

Ribera, M., and 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., and Andrews, T. A. 2014. "Creating Value in Health Care through Big Data: Opportunities and Policy Implications," *Health Affairs* (33:7), pp. 1115–1122.

Schnackenberg, A. K., Tomlinson, E., and Coen, C. 2021. "The Dimensional Structure of Transparency: A Construct Validation of Transparency as Disclosure, Clarity, and Accuracy in Organizations," *Human Relations* (Vol. 74).

Schneider, J., and Handali, J. P. 2020. "Personalized Explanation for Machine Learning: A Conceptualization," *27th European Conference on Information Systems*.

Schoonderwoerd, T. A. J., 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), Elsevier Ltd, p. 102684.

Shin, D. 2020. “How Do Users Interact with Algorithm Recommender Systems? The Interaction of Users, Algorithms, and Performance,” *Computers in Human Behavior* (109:March), Elsevier Ltd, p. 106344.

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), Elsevier Ltd, p. 102551.

Song, J., and Zahedi, F. M. 2007. “Trust in Health Infomediaries,” *Decision Support Systems* (43:2), pp. 390–407.

Wang, S., Beatty, S. E., and Foxx, W. 2004. “Signaling the Trustworthiness of Small Online Retailers,” *Journal of Interactive Marketing* (18:1), Elsevier, pp. 53–69.

Wells, J., Valacich, J., and Hess, T. 2011. “What Signal Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions,” *MIS Quarterly* (35:2), pp. 373–396.

Woodcock, C., Mittelstadt, B., Busbridge, D., and Blank, G. 2021. “The Impact of Explanations on Layperson Trust in Artificial Intelligence–Driven Symptom Checker Apps: Experimental Study,” *Journal of Medical Internet Research* (23:11), p. e29386.

Ye, L. R., and Johnson, P. E. 1995. “The Impact of Explanation Facilities on User Acceptance of Expert Systems Advice,” *MIS Quarterly: Management Information Systems* (19:2), pp. 157–172.

Zhang, Z., Genc, Y., Wang, D., Ahsen, M. E., and Fan, X. 2021. “Effect of AI Explanations on Human Perceptions of Patient-Facing AI-Powered Healthcare Systems,” *Journal of Medical Systems* (45:6), Springer US.