Electronic Markets (2023) 33:17
https://doi.org/10.1007/s12525-023-00640-9

RESEARCH PAPER

Explanation matters: An experimental study on explainable AI

Pascal Hamm[1] · **Michael Klesel**[2] · **Patricia Coberger**[3] · **H. Felix Wittmann**[4]

Received: 31 May 2022 / Accepted: 17 November 2022

Abstract
Explainable artificial intelligence (XAI) is an important advance in the field of machine learning to shed light on black box algorithms and thus a promising approach to improving artificial intelligence (AI) adoption. While previous literature has already addressed the technological benefits of XAI, there has been little research on XAI from the user's perspective. Building upon the theory of trust, we propose a model that hypothesizes that post hoc explainability (using Shapley Additive Explanations) has a significant impact on use-related variables in this context. To test our model, we designed an experiment using a randomized controlled trial design where participants compare signatures and detect forged signatures. Surprisingly, our study shows that XAI only has a small but significant impact on perceived explainability. Nevertheless, we demonstrate that a high level of perceived explainability has a strong impact on important constructs including trust and perceived usefulness. A post hoc analysis shows that hedonic factors are significantly related to perceived explainability and require more attention in future research. We conclude with important directions for academia and for organizations.

Keywords Artificial intelligence · AI · XAI · Perception · Experiment

JEL Classification C91

Introduction

AI has shown great potential in various areas of the private and organizational life. However, research has shown that current AI implementations are flawed as they often obscure the underlying mechanisms and only present predictive results. This in turn can lead to undesirable effects such as "automation bias," which describes a tendency to rely too much on automated processes (Goddard et al., 2012). Problems related to current AI solutions have also been recognized in politics. In this context, it is particularly worth mentioning that the European Union (EU) has created uniform rules for dealing with AI with the introduction of the General Data Protection Regulation (GDPR) (European Union 2016). This provides, among other things, that users affected by an automated decision have a right to transparent information about the logic of the algorithm (Confalonieri et al., 2019).

From an academic perspective, XAI has emerged as an important research field, which seeks to develop and evaluate mechanisms that offer better insights into the underlying mechanisms of an AI algorithm. XAI is an interdisciplinary research field in the AI ecosystem that aims to make the results of AI systems understandable and comprehensible to humans (Adadi & Berrada, 2018; Förster et al., 2020). The research field focuses on trying to open up the "black box" associated with AI and thereby creates explainability for humans (Muddamsetty et al., 2020). For that reason, XAI applications can be found in many areas of daily life, such as healthcare (Jussupow et al., 2021),

Responsible Editor: Christian Meske

✉ Michael Klesel
m.klesel@utwente.nl

Pascal Hamm
pascal.hamm@myebs.de

Patricia Coberger
patricia.c.coberger@stud.h-da.de

H. Felix Wittmann
hfwittmann@gmail.com

[1] EBS University, Rheingaustraße 1, 65275 Oestrich-Winkel, Germany

[2] University of Twente, De Horst 2, 7522LW Enschede, Netherlands

[3] Darmstadt University of Applied Science, Haardtring 100, 64295 Darmstadt, Germany

[4] University of Cambridge, Trumpington St, Cambridge CB2 1RF, UK

finance (Mao & Benbasat, 2000), and autonomous driving (Muhammad et al., 2021). Modern software packages, including Shapley Additive Explanations (SHAP) (Bowen & Ungar, 2020) and Local Interpretable Model-Agnostic Explanations (LIME) (Ribeiro et al., 2016) can be used to explain algorithmic decisions with local post hoc explanations to users in non-technical terms so that they can understand, appropriately trust, and effectively manage AI results.

Previous research has shown that XAI can lead to positive effects on potential users, for instance, an increased perception of usefulness and ease of use (Meske & Bunde, 2022), a higher perceived explainability (Dominguez et al., 2020), or a higher observability (Schrills & Franke, 2020). However, there are also studies with non-significant or mixed results (e.g., Alufaisan et al., 2020; David et al., 2021; Druce et al., 2021; Schrills & Franke, 2020; van der Waa et al., 2021). As a consequence of incomplete knowledge about the effects of XAI, more research is needed to guide organizations in terms of how to use and implement XAI components. This gap is also reflected in the current literature, which calls for more research on XAI that investigates user perception (van der Waa et al., 2021).

We respond to this gap and extend existing literature by focusing on user perception. We focus on a common business task in which participants have to identify forged signatures. Detecting forged signatures is important in various areas of social and professional life, as signatures are a fundamental part of any contract and essential for most official documents. This is also a reason why there are increasing efforts to detect forged signatures (e.g., Zhou et al., 2021). For our study, during task completion, participants were assisted by an AI system that included an XAI module using SHAP (Adadi & Berrada, 2018) for individuals in the treatment group. To pursue our objective, the following overarching research question is used to guide this study (RQ):

> RQ: How does XAI influence a user's perception of usefulness, ease of use, trust, and performance?

The remainder of this paper is organized as follows: in the "Related work" section, we review related work on XAI and offer an overview of existing experimental studies. In the "Hypothesis development" section, we propose a research model that incorporates important perceptional variables based on prior literature. In the "Methodology" section, we describe the experimental design and the methodological procedure to test our hypotheses. The "Results" section summarizes the results of the experiment. We conclude the "Discussion" section with a discussion on the relevance of the collected results for theory and practice.

Related work

Despite impressive performance improvements of current AI algorithms, many modern AI algorithms lack inherent explainability due to the "black box" associated with AI, which makes AI decisions and predictions opaque and non-transparent to the user (Förster et al., 2020; A. Rai, 2020; Ribeiro et al., 2016). This can lead to undesirable or questionable results and, ultimately, to significant distrust of a particular system. A notable example in this regard is an AI system trained to support the hiring process that initially produced undetected and undesirable results such as gender bias. This was observed at Amazon, where an AI-based hiring tool favored male over female applicants rather than providing objective suggestions for new applicants (Hsu, 2020). Black box algorithms are particularly critical in such situations, because they prevent a solid understanding of the underlying mechanisms that is required to detect and avoid undesirable outcomes. Due to the increasing interest in understanding black box algorithms, governments have begun to enact regulations in this regard. The previously mentioned GDPR requires that AI systems be held accountable using techniques that can explain the underlying mechanisms (European Union 2016). While government regulations such as the GDPR are arguably a critical factor in boosting explanatory efforts, there are also emerging fields that benefit greatly from a better understanding of AI. For example, using "machine learning" to teach people (Schneider & Handali, 2019) requires a solid level of explanation to impart knowledge to people.

In this context, XAI has become an important research direction that attempts to shed light on the black box problem. XAI describes the type of AI systems that provides insight into how a decision or prediction is made and how the resulting actions are executed. However, there is currently no universally accepted definition of XAI. Rather than a formal technical concept, the term refers to the movement, initiatives, and efforts being made in response to concerns about transparency and trust in AI (Adadi & Berrada, 2018). Basically, two types can be distinguished: (1) transparent models and (2) black box models. Transparent models benefit from their basic design, which allows them to provide explanations based on them. For example, linear regression models, rule-based systems, or decision trees can be understood as transparent models because their design provides insights into their mechanism, and the models are fully explainable and understandable. In contrast, black box models do not provide such insights, so post hoc analysis is required to further illuminate the results of the algorithms. There are several post hoc techniques including textual explanations, visualizations, local explanations, explanations by example, explanations by simplification, and feature relevance (Arrieta et al., 2020). Currently, the

use of SHAP values is considered a promising technique for obtaining local explanations (Adadi & Berrada, 2018). Examples of black box models include feed-forward neural networks, convolutional neural networks, recurrent neural networks, and generative adversarial networks. Both agnostic and model-specific approaches are possible for black box models (Arrieta et al., 2020; Rai, 2020). Explanations of model-agnostic techniques are necessarily not model-specific. Model agnostic techniques use the inputs and predictions of black box models to generate explanations based on the data inputs (Rai, 2020). On the other hand, transparent models rarely require further post hoc analysis, although, agnostic approaches can of course be applied to transparent models, as well. A classification of XAI techniques is summarized in Table 1.

Besides the need for explanation from a social and technological perspective, previous literature has highlighted that XAI can have a positive impact on (non-technical) users. The fundamental idea is that individuals perceiving a higher degree of transparency by means of an explainable component will also perceive the outcome more positively. For this reason, the theory of trust (Gefen et al., 2003) has been used as a central theoretical lens to study the impact of XAI (e.g., Ribeiro et al., 2016; Sperrle et al., 2020; Thiebes et al., 2021).

There are several studies that provide evidence on the impact of XAI in experimental settings (an overview is shown in Table 2). For instance, Meske and Bunde (2022) observed a positive effect of XAI on perceived ease of use, perceived usefulness, intention to use, perceived informativeness, trustworthiness, and mental model. Concurrently, the authors identify a negative effect of XAI on perceived cognitive effort. The participants' task in the study was to identify hateful content via the user interface to detect hate speech. Shafti et al. (2022) observed that good explanations of XAI can lead to a significantly lower error rate, a higher human performance and higher user confidence in AI. In their experimental study, a grade prediction task of students was used based on tabular data about the student's background (e.g., parents' jobs or weekly study time). A positive impact of XAI on human performance was also observed by Lai et al. (2020) and Ray et al. (2019). Lai et al. (2020) asked participants to use an AI system to detect text-based fraudulent reviews and classify a total of 20 reviews as fraudulent or genuine. Ray et al. (2019) challenged participants to guess an image secretly selected by the AI system by asking the AI system questions in natural language (e.g., what kind of vehicle is in the picture?). In another study by Dominguez et al. (2020), the authors observed a positive effect of XAI on user satisfaction, perception of explainability, and relevance. These effects were observed in an experimental study in which participants provided feedback on image recommendations generated by an AI algorithm. The recommended images were based on a selection of images the participants "liked" via a dashboard before. Weitz et al., (2019, 2021) found a positive effect of XAI on user trust in the context of an experimental task on classification of audio keywords spoken by participants into an AI-based speech recognition system. In an experiment to evaluate the performance of an AI system on different types of image-based questions (e.g., what color is the man's phone?), Alipour et al. (2020a) detected a positive effect of XAI on the prediction of the users and competencies of the model. Furthermore, Alipour, Schulze, et al. (2020b) observed a positive impact of XAI on user prediction accuracy, user confidence, and user reliance. While the task in Alipour et al. (2020b) was to evaluate the AI's respective performance in answering four different types of image-based questions, the task in Alipour, Schulze, et al. (2020b) was to predict the answer accuracy of a visual question answering (VQA) agent.

While there are several studies showing positive effects, there are also studies showing that XAI does not always lead to the desired outcome. For example, Alufaisan et al. (2020) found that XAI has a positive effect on the decision-making process. In contrast, no significant effect of XAI was found on decision accuracy, following the AI recommendation, and decision confidence. The authors detected these findings in two

Table 1 Classification of XAI techniques

	Transparent models	Black- box models
Definition	Models that can fully and understandably explain how an algorithm operates and, given an input, can tell what the output will be and why (Arrieta et al., 2020; Asatiani et al., 2020)	Models that create internal structures that determine outputs, but are opaque to external parties. Even the programmers cannot tell why a particular output was produced (Asatiani et al., 2020)
Agnostic	Possible	Possible
Model specific	Not required	Possible
Example algorithms	Linear regression model (Chatla & Shmueli, 2017), rule-based systems (Golding & Rosenbloom, 1991), decision trees (Quinlan, 1987)	Feed-forward neural networks (Bebis & Georgiopoulos, 1994), convolutional neural networks (Albawi et al., 2017), recurrent neural networks (Ghanvatkar & Rajan, 2019), generative adversarial networks (Wong et al., 2020)

Table 2 Existing experimental studies on XAI and the user's perception

Articles	Task	Sample/design	Dependent variables[1]	How is it explained?/what is explained?
(Meske & Bunde, 2022)	Identifying hateful content using a user interface for hate speech detection	$N=550$ 1×3 between-subject design Participants: CloudResearch, MTurk	Perceived ease of use (+), perceived usefulness (+), intention to use (+), perceived cognitive effort (−), perceived informativeness (+), mental model (+), trustworthiness (+)	Universal Language Model Fine-Tuning's (ULMFiTs) interpretation module (local explanations through colored marking of text sections)
(Shafti et al., 2022)	Grade prediction task that requires participants to predict the student's grade based on tabular data about the student's background (e.g., parents' jobs or weekly study time)	$N=167$ 1×1 within-subject design Participants: MTurk	Lower error ($+^2$), human performance ($+^2$), trust (+)	LIME visualizations (marking of positive and negative characteristics that can influence the result)
(Alufaisan et al., 2020)	Experiment 1: prediction of the likelihood that a criminal will reoffend within a given time period based on characteristics such as age, charge degree and number of priors Experiment 2: prediction of people's income (≤88 K; > 88 K) based on characteristics such as age, education and work hours per week	$N=300$ 2×3 between-subject design Participants: MTurk	Decision-making process (+), decision-making accuracy (−), follow AI recommendation (−), decision confidence (−)	LIME visualizations (marking of the characteristics on the basis of which the AI has made its decision)
(David et al., 2021)	Web-based game with the objective of generating maximum revenue through the production and sale of lemonade with real financial implications	$N=449$ 1×5 between-subject design Participants: MTurk	Readiness to adopt (+/−) Trust (+/−) Willingness to pay (+/−) Mixed results across different phases (longitudinal study)	Global explanation (general information about how the model operates), feature-based explanation, (display of features that were considered in the decision-making process), performance-based explanation (display of the accuracy of the decision)
(Druce et al., 2021)	Completion of system acceptance questionnaire based on 3 agents appearing in 12 video sequences either with or without explanations. Then, participants are additionally asked to predict system performance for each of the 3 agents based on 36 (freeze-frame) Amidar game states	$N=18$ 1×2 between-subject design Participants: not specified	Trust (+), AI acceptance (+), perception of prediction correctness (n.s.)	Threefold explanation including (1) graphical representation of the generalization and performance of the system in the current game state, (2) how well the agent would play in semantically similar environments, and (3) a narrative explanation of what the graphical information implies
(van der Waa et al., 2021)	The task of the participants is to decide whether to increase or decrease the dose of insulin based on the characteristics planned alcohol intake, water intake so far, and hours slept	$N=90$ 2×3 between-subject design Participants: participant database at TNO Soesterberg (Netherlands), social media	System understanding ($+^2$), persuasive power (+), task performance (n.s.)	Rule-based explanation (marking of the characteristics on the basis of which the AI has made its decision), example-based explanation (comparison of the current expression of the characteristics with the expression of past situations and the associated AI decisions)
(Weitz et al., 2021)	Participants have the task to speak 10 English keywords (e.g., dog or happy) into a speech recognition system, which in the second step are classified by the AI based on the audio signal	$N=60$ 1×4 between-subject design Participants: not specified	Trust (+)	LIME visualizations (visual representation of audio samples using highlighted spectrograms to show sound pressure levels as pixel values)

Table 2 (continued)

Articles	Task	Sample/design	Dependent variables[1]	How is it explained?/what is explained?
(Alipour, Ray, et al. 2020a)	The participants' task is to evaluate the AI's respective performance in answering four different types of image-based questions (e.g., what color is the man's phone?). To do this, participants can view the AI's top five answers along with the probability of the answers	$N=40$ 2×2 between-subject design Participants: not specified	User's prediction (+[2]), model's competencies (+[2])	Spatial attention VQA (SVQA); BERT VQA (SOBERT) (usage of an attention mechanism to select visual features generated by an image encoder)
(Alipour, Schulze, et al. 2020b)	The participants' task is to predict the answer accuracy of a VQA agent (e.g., what covers the ground? in the context of an image-based question)	$N=90$ 1×6 between-subject design Participants: not specified	User prediction accuracy (+), user confidence (+), user reliance (+)	XVQA model (combining the attention features generated in the VQA model with meaningful annotations from the injected data) The paper's VQA is powered by the BERT model
(Dominguez et al., 2020)	Participants had to "like" at least 10 images via a Pinterest-like interface. Based on this, the participants received image recommendations, on which they were asked to provide feedback regarding their match with the images they had selected. While study 1 only focuses on the desktop interface, study 2 examines the effects of explanations on mobile devices	Study 1: $N=121$ 1×3 between-subject design Participants: MTurk Study 2: $N=177$ 2×2 between-subject design Participants: MTurk	User satisfaction (+), perception of explainability (+), relevance (+)	Deep neural network visual feature (DNN visual feature) (marking of visual features that are **not** interpretable by humans) Attractiveness visual features (AVF) (marking of visual features that are interpretable by humans (e.g., brightness, saturation, and sharpness))
(Lai et al., 2020)	The task of the participants is to detect text-based fraudulent reviews and classify a total of 20 reviews as fraudulent or genuine. Three experiments will be conducted to observe (1) whether explanations improve human performance, (2) whether human performance can be further improved by real-time assistance, and (3) the relationship between model complexity/methods of deriving explanations and human performance	Experiment 1: $N=480$ 1×6 between-subject design Participants: MTurk Experiment 2: $N=480$ 1×6 between-subject design Participants: MTurk Experiment 3: $N=480$ 1×6 between-subject design Participants: MTurk	Human performance (+)	SP-LIMEBERT Lime (colored marking of the most important text passages in green if there is evidence for the authenticity of the evaluation and red if the AI indicate a fraud) BERT attention (colored marking of the most and least important text passages)
(Schrills & Franke, 2020)	Evaluation of three different visual explanation approaches, which consisted of (1) a table of classification values only, (2) a table of values with an omni-condition, and (3) a table of classifications values with counterfactual reasoning	$N=83$ 1×1 within-subject design Participants: E-mail, social networks	Perceived trustworthiness (n.s.), observability (+)	Layer-wise Relevance Propagation (LRP) (heat map, where pixels were colored depending on the intensity of the color deviation in both images)
(Ray et al., 2019)	The participants' task is to guess a picture secretly selected by the VQA agent by asking the AI questions in natural language (e.g., what kind of vehicle is in the picture?). The VQA agent generates answers to these questions	$N=60$ 1×6 between-subject design Participants: not specified	Human performance (+)	VQA (attention masks highlight relevant locations and objects in the image)

Table 2 (continued)

Articles	Task	Sample/design	Dependent variables[1]	How is it explained?/what is explained?
(Weitz et al., 2019)	The task of the participants is to speak a predefined and fixed sequence of ten keywords (e.g., seven) into a microphone, which are classified by the AI after each phrase	$N = 30$ 1×2 between-subject design Participants: not specified	Trust (+)	LIME visualizations (green highlighting of the areas of the spectrograms that are favorable for prediction, and red highlighting of the unfavorable areas)

[1] A significant, positive effect (+), a significant, negative effect (-), a non-significant, positive or negative effect [not specified (n.s.)]. We only included the hypothesis with XAI as an independent variable

[2] Only significant between specific groups

experimental studies in which participants were asked to (1) predict the likelihood that a criminal would reoffend within a given time period based on characteristics such as age, level of charge, and number of prior convictions and (2) predict individual's income (≤ 88 K; > 88 K) based on characteristics such as age, education, and work hours per week. Schrills and Franke (2020) observed a positive influence of XAI on the observability of the AI system. However, the hypothesis that XAI leads to higher perceived trustworthiness in an AI system could not be confirmed. The task for the participants was to evaluate three different visual explanation approaches, which consisted of either a table of classification values only or additionally one of two different backtracked visual explanations. In a study comparing rule-based explanations and example-based explanations, van der Waa et al. (2021) reported that only rule-based explanations have a positive effect on system understanding. For both rule-based explanations and example-based explanations, a positive effect on persuasive power could be observed. However, no effect was observed on the impact of XAI on task performance. In a further study, David et al. (2021) showed that XAI may have a positive influence on users' readiness to adopt and their willingness to pay and trust. The authors observed mixed results based on different phases of the longitudinal study. The experimental study was performed in a web-based game in which participants were asked to generate as much revenue as possible by producing and selling lemonade under real monetary conditions. The authors found that users were only willing to pay an average of 1.005 game coins for assistance from an AI. In contrast, participants were willing to pay an average of 1.774 game coins, for the options of human advisor, global explanation, feature-based explanations, and performance-based explanations. A positive effect of XAI on trust and AI acceptance was identified by Druce et al. (2021) in a two-part experimental study investigating system acceptance of video game playing agents via a questionnaire and asking participants to predict the system performance of the agents. However, XAI did not lead to a significant improvement in perception of prediction accuracy.

In summary, research on XAI is flourishing and there are several studies highlighting a positive impact of XAI on outcome variables such as trust (David et al., 2021; Druce et al., 2021; Shafti et al., 2022; Weitz et al., 2019, 2021) or human performance (Lai et al., 2020; Ray et al., 2019; Shafti et al., 2022). However, there are also studies showing that the effectiveness of XAI is related to specific implementations (e.g., van der Waa et al., 2021) and that XAI does not per se lead to intended effects (e.g., Alufaisan et al., 2020). Therefore, more research is required to investigate how XAI should be designed and implemented in order to achieve the desired outcomes.

The studies not only reviewed contributed to a better understanding of XAI but also highlighted limitations that should be addressed to further advance XAI research. For example, the study by Alufaisan et al. (2020) mentioned the limited

transferability of their results to other datasets and explainable AI techniques. In addition, van der Waa et al. (2021) mentioned the type of the explanatory component used in the study as a limitation of their study. The focus on a single manifestation of trust (e.g. situational trust), rather than looking at trust as a whole (Weitz et al., 2019, 2021), and the lack of representativeness of the study group due to acquiring participants via MTurk (Lai et al., 2020) were also mentioned as limitations of previous studies. We seek to address some of these limitations and propose a research model that investigates the relationship between XAI and important dependent variables as explained in the following section.

Hypothesis development

This research focuses on the relationship between XAI and perceptional constructs to strengthen our understanding of how XAI influences potential users. We draw on previous literature on technology acceptance and use (Venkatesh et al., 2016) and trust theory (Gefen et al., 2003) as a theoretical lens for this study. Therefore, we include four fundamental variables from these streams of literature, namely, *perceived usefulness*, *perceived ease of use, trust intention*, and *performance*. We selected these variables because they are well established in the information systems (IS) literature (Venkatesh et al., 2016) and because these constructs have been used in prior literature on XAI (e.g., Druce et al., 2021; Meske & Bunde, 2022; Weitz et al., 2021). In addition, we include *performance* as this is pivotal from an organizational perspective (e.g., Goodhue & Thompson, 1995). We use these well-established constructs to draw specific conclusions on the role of XAI in the domain of IS research. *Perceived explainability* (Wang & Benbasat, 2016) is used as the corresponding latent construct for XAI. We do this in light of the fact that XAI aims to explain AI algorithms to users in an understandable way and is therefore perceived by users as having explanatory power. Our research model is shown in Fig. 1 and explained below. An overview of the key concepts used here is shown in Table 3.

A fundamental objective of XAI is to reveal the underlying mechanisms of an algorithm and to make the results and predictions understandable and transparent to humans (Adadi & Berrada, 2018). In general, humans tend to prefer things that are universally understandable and are reluctant to adopt technologies that are not comprehensible or trustworthy (Arrieta et al., 2020; Gefen et al., 2003). This is also supported by previous studies suggesting that effective explanations generate understandable explanations (Galhotra et al., 2021). In addition, this is consistent with the findings of Muddamsetty et al. (2020), who suggest that opening up the black box of AI algorithms can increase the degree of explainability. Consequently, it is reasonable to assume that XAI provides users with a higher degree of perceived explainability than AI modules without explanation. Accordingly, we state our first hypothesis (H) as follows:

> H_1: *XAI has a higher degree of perceived explainability compared to AI without a XAI component.*

Since Davis (1989) proposed the Technology Acceptance Model (TAM) and its well-known extensions (Venkatesh et al., 2016), perceived usefulness and perceived ease of use have been established constructs for predicting use behavior. According to them, useful and easy-to-use technologies are more likely to be used. Explanations and perceived explainability support users to use the AI more thoroughly with regard to the underlying task. AI algorithms are often used to support users' decision-making. For this reason, XAI can be expected to lead to an understanding of the information on which predictions are made. A better insight into the mechanism of an algorithm can help users to better understand the technology, which in turn has a positive impact on technology-related perceptual variables. In addition, explanations are also helpful to obtain new information and thus gain additional knowledge which in turn can help to justify an AI-based decisions (Adadi & Berrada, 2018). Therefore, a higher level of perceived explanation

Fig. 1 Research model

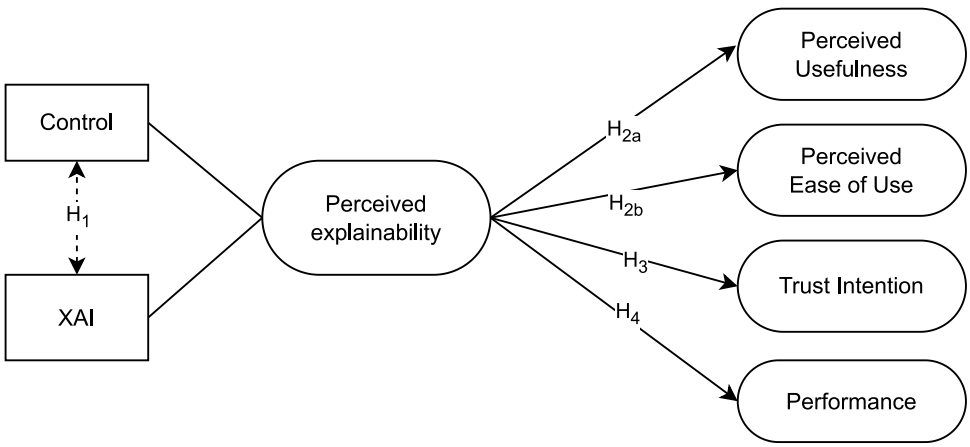

Table 3 Definition of concepts

Concept	Definition
Perceived explainability	"Explainability is associated with the notion of explanation as an interface between humans and a decision-maker that is, at the same time, both an accurate proxy of the decision maker and comprehensible to humans." (Arrieta et al., 2020, p. 85)
Perceived usefulness	"The degree to which a person believes that using a particular system would enhance his or her job performance." (Davis, 1989, p. 320; Venkatesh et al., 2003, p. 428)
Perceived ease of use	"The degree to which a person believes that using a particular system would be free of effort." (Davis, 1989, p. 320; Venkatesh et al., 2003, p. 428)
Trust	"[…] researchers view trust as (1) a set of specific beliefs dealing primarily with the integrity, benevolence, and ability of another party, (2) general belief that another party can be trusted, sometimes also called trusting intentions or "the 'willingness' of a party to be vulnerable to the actions of another", (3) affect reflected in "'feelings' of confidence and security in the caring response" of the other party, or (4) combination of these elements." (Gefen et al., 2003, p. 55)
Performance	"Higher performance implies some mix of improved efficiency, improved effectiveness, and/or higher quality." (Goodhue & Thompson, 1995, p. 218)

can also lead to a higher level of usefulness. Combining these arguments, we hypothesize that perceived explainability has a positive impact on perceived usefulness and perceived ease of use and propose the following two hypotheses:

H_{2a}: *A higher degree of perceived explainability results in a higher level of perceived usefulness.*
H_{2b}: *A higher degree of perceived explainability results in a higher level of perceived ease of use.*

Understanding of someone or something is crucial in building trust (Gefen et al., 2003; Gilpin et al., 2019). Therefore, a XAI component that helps users to better understand what the AI is doing has the potential to increase users' perceived trust. This is consistent with previous research that has shown that trust can only be built if humans understand the decisions of AI algorithms (Sperrle et al., 2020). Consequently, explanations help to verify predictions, improve models, and gain new insights into the problem which ultimately leads to an increase in trust in AI algorithms (Adadi & Berrada, 2018). In contrast, trust will be lost when users cannot understand the behavior or decisions of AI algorithms (Miller, 2018). In particular, when an AI algorithm's results and predictions do not match users' expectations, a lack of explainability can lead to a loss of trust on the part of users (Kizilcec, 2016). For example, in a study on intelligent systems, Holliday et al. (2016) found that trust among users of intelligent systems with explanation increased over the duration of the experiment, whereas trust among the user group of intelligent systems without explanation decreased. As a result, several authors agree that increasing trust is the primary goal of XAI (Arrieta et al., 2020; Kim et al., 2015; Ribeiro et al., 2016). In this sense, it can be assumed that the explanatory power of XAI leads to higher levels of trust. Accordingly, we hypothesize that:

H_3: *A higher degree of perceived explainability results in a higher level of trust intention.*

A higher degree of perceived explainability can also influence how individuals perform. For example, Ray et al. (2019) demonstrate that helpful explanations can improve participants' performance. Similarly, Lai et al. (2020) show that explanations lead to better human performance than pure deep learning models. Moreover, van der Waa et al.'s (2021) study of insulin dose prediction in the context of diabetes provided the first evidence that example-based explanations by AI algorithms can improve participants' task performance. Lai et al. (2020) also show that the methods used to derive explanations have a significant impact on human performance. The results suggest that human performance is better when, in the context of a text-based task, explanations highlight important words that contributed to the AI's result. Therefore, we propose our final hypothesis:

H_4: *A higher degree of perceived explainability results in a higher level of human performance.*

Methodology

Experimental design

To test our research model, we carried out an experimental study. We developed a dashboard with an AI component to support participants detecting forged signatures. Participants in the control group were able to use the dashboard as shown in Fig. 2. For the treatment groups, another dashboard was developed with an additional XAI component as shown in Fig. 3. The XAI module highlighted specific pixels in the images that had an effect on the decision of the AI. Areas highlighted in red represent a deviation from the reference signature and thus indicate a forged signature. In contrast, areas highlighted

in green indicate a similarity to the reference signature and thus indicate authenticity. As part of the XAI component, we added an additional slider that allowed participants to change the sensitivity of the explanation. In other words, each participant was able to vary the sensitivity of the XAI component. A higher value of the slider resulted in a higher sensitivity of the illustrated results and vice versa. Both dashboards provided a recommendation from the AI module at the top (e.g., "The AI identified the unknown signature as **forged** [or as **original**]). SHAP was used to color the pixels of the unknown signature that differed from the original signature in red and the similar pixels in green. Both dashboards and the subsequent questionnaires were operable in English as well as in German.

Before the experimental tasks were performed, there was a short introduction explaining the experimental task and the use of the dashboard in general to the participants. In particular, we provided a tutorial video illustrating the functions of the AI dashboard and its use in performing the experimental task. After watching the video, participants had the opportunity to familiarize themselves with the functions of the AI dashboard by completing a sample task. Once participants were familiar with the AI dashboard, they could start the experiment after confirming that they had understood the task.

The experimental procedure was divided into three parts: first, participants were asked to complete a pre-questionnaire, where we provided an introduction and asked demographic questions. Second, we directed participants to the AI dashboard where the experimental task took place. Third, we conducted a post-survey, in which we measured the perceived variables related to the experimental task. We did not set a time limit to allow participants to engage with the AI dashboard as much as they would like and to give the opportunity to find a solution without time pressure.

Experimental task

To make this research relevant to theory and practice, we used an experimental task involving some kind of sensitive data, as organizations commonly deal with sensitive information. In addition, the task should reduce language and cultural biases to increase the scope of the results. For these reasons, we chose an experimental task where participants had to verify signatures against a reference signature. This task is highly relevant as it is part of many business processes (e.g., account opening, buying an insurance policy, or contract amendment) and is of great interest to organizations to reduce fraud (e.g., Hussein et al., 2016). Moreover, the use of images can reduce any kind of language-related bias, as no understanding of a specific language is required to fulfill this task. Moreover, signatures used here are very similar within Western countries which also reduces language-specific bias effects. Also, in the case of signatures, it can be largely assumed that there are fewer differences in understanding between native speakers and foreign speakers.

Specifically, a pair of images depicting the same signature was shown. A total of 20 pairs of signatures were provided. The first signature served as a reference (on the left side of Figs. 2 and 3), while the second was a potential candidate for a forged signature. An AI component was available to assist in this process. As explained above, the participants of the XAI group had the option of using a slider that highlighted the impact of different pixels on the AI's decision. Furthermore, these highlighted pixels indicated whether they supported the decision for a faked signature.

The 20 different tasks contained equal numbers of originals and forgeries. In addition, the pairs were chosen so that the recommendation of the AI corresponds to the truth in 80% of the cases (c.f. Table 4). This applies to both original and forged signatures. This means that for 10 pairs, the forgery would be the correct choice. However, the AI classified only 8 of them correctly, two were wrongly classified as originals. This distribution also applied to the group of original pairs. In total, following the AI recommendation would lead to 16 correct answers.

The AI component in the background was a Siamese neural network. Siamese neural networks are tailor-made for the comparison of potentially identical entities and therefore outperform other deep learning-based approaches. For that reason, they are well-suited for our research setting. Two images are processed simultaneously by a convolutional neural network (CNN). Then, the distance between the resulting images is calculated, and an output of the classification problem is provided. The TensorFlow library was used to set up the neural network (Abadi et al., 2016). A training dataset from Kaggle (Rai, 2022) with 1320 pairs of signatures was

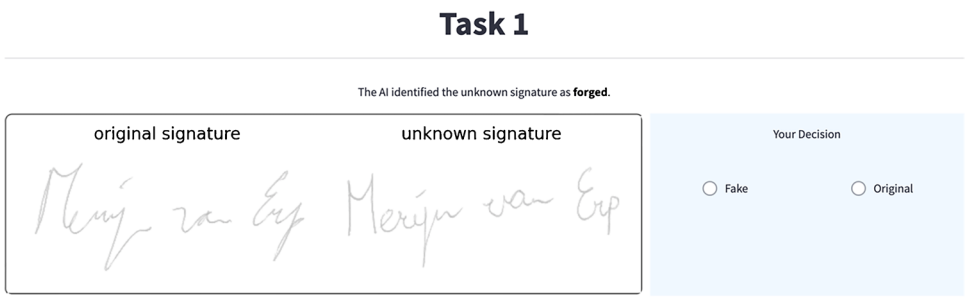

Fig. 2 AI dashboard (control group)

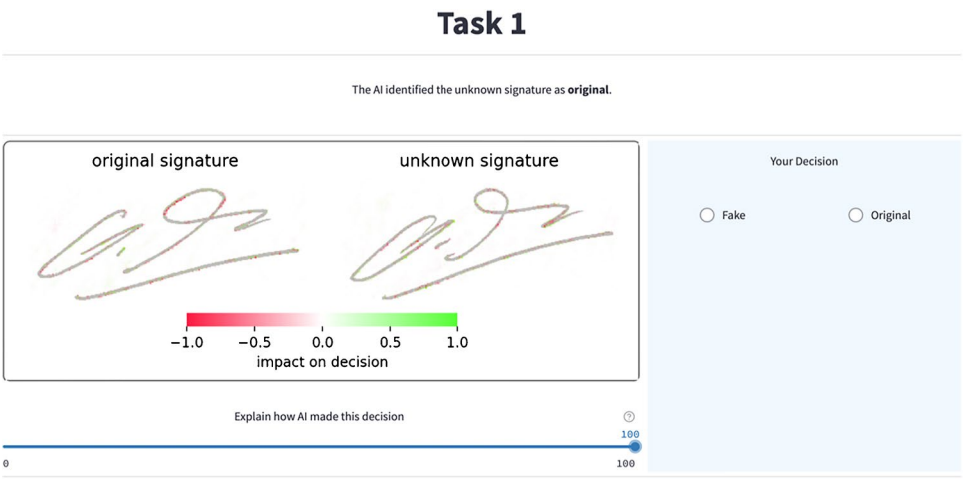

Fig. 3 XAI dashboard (treatment group)

used to train our model and evaluated on more than 500 different pairs. In addition, the explanation of the underlying AI decision process was provided using SHAP as it has higher performance compared to other approaches (e.g., Gramegna & Giudici, 2021). The dashboard was implemented using Streamlit (Streamlit, 2022).

Sample

According to Cohen (1988), a total sample size of 63 subjects is required to achieve sufficient statistical power of 0.8 for a mean effect size ($f=0.25$). We collected data from 106 participants (39 control group, 67 treatment group). Because we are trying to understand the effects of XAI, we excluded observations using the XAI slider five times or less for the entire experiment. Consequently, we had to exclude 27 observations from the treatment group, resulting in a total of 79 observations (a more detailed analysis of the exclusion of observations is attached in the appendix). As we were interested in a high degree of external validity, we recruited participants from different backgrounds and specialties using a snow-ball sampling strategy (Biernacki & Waldorf, 1981; Naderifar et al., 2017). Specifically, we started our strategy by recruiting staff from a large insurance organization and asking them to forward the study to other colleagues. In line with the snowball sampling strategy, we asked the initial participants to forward the participation link for the experiment only to people who had a similar profile to the participant him or herself (Naderifar et al., 2017). As a consequence of this strategy, the final sample includes participants from within and from outside the organization. Participation was voluntary and rewarded with non-monetary recognition in the form of a donation to a charitable organization. To avoid bias effects due to excessive engagement with current AI developments, we excluded individuals from the information technology department. We drew a random number to assign participants to one of the two experimental conditions. This random variable was drawn independently for each participant. The presentation of the 20 signature tasks was also randomized to avoid unwanted effects (e.g., learning effects) due to the order of the images. An overview of our sample is summarized in Table 5.

Measures

The purpose of this experiment was to increase the perceived explainability of AI via an explainable component. To investigate the effectiveness of our manipulation, we also collected a self-assessment of perceived explainability. We looked for an established measurement scale that has been used in a related context with similar constructs such as trust. Since there was no established measurement scale for perceived explainability, we looked for an established measurement scale that captures explainability to humans (c.f. Table 3). In this context, we chose Wang and Benbasat (2016) measurement scale for perceived transparency, which has been used to reflect knowledge-based reasoning and therefore aligns well with what we defined as perceived explainability. In summary, the close conceptual relationship between perceived explainability and perceived transparency allowed us to use an established measure for this study. We used this scale using a seven-point Likert scale with the endpoints labeled as "strongly disagree" and "strongly agree" to measure our manipulation (Wang & Benbasat, 2016). All five items are listed in Table 6. We adapted all items slightly to the context of this study e.g., the *AI dashboard* made its reasoning process clear to me.

We first examined convergent reliability using Cronbach's α and omitted the reverse item (MC3) to increase reliability

Table 4 Selected images

$N=20$ (tasks and 20 pairs of images respectively)		Predicted value	
		Forged	Real
True value	Forged	8	2
	Real	2	8

Table 5 Sample description

Dimension	Classification	Percentage	Percentage (control group)	Percentage (treatment group)
Age	18–34	83.5%	84.6%	82.5%
	35–44	6.3%	5.1%	7.5%
	45–54	2.5%	0.0%	5.0%
	Older than 55	7.6%	10.3%	5.0%
Role	No explicit role related to AI	60.8%	66.7%	55.0%
	Users	19.0%	20.5%	17.5%
	Researcher	11.4%	5.1%	17.5%
	Consultant	6.3%	5.1%	7.5%
	Developer	1.3%	0.0%	2.5%
	Other	1.3%	2.5%	0.0%
Experience	No or little experience	73.4%	76.9%	70.0%
	Experienced	15.2%	17.9%	12.5%
	Neutral	10.1%	2.6%	17.5%
	Other	1.3%	2.6%	0.0%

($\alpha = 0.83$). We then used a *t*-test to examine whether the two groups differed significantly. However, the 39 participants in the control group ($M = 4.24$, $SD = 1.50$) did not differ significantly from the treatment group ($M = 4.58$, $SD = 1.39$) $t\,(71) = -1.02$, $p = 0.31$. A further analysis of the individual variables (c.f. Figure 7) was conducted to investigate possible problems with content validity. In fact, it could be observed that the first item showed differences between the groups. In contrast, the other four items showed no differences. We conducted a content-validity check (MacKenzie et al., 2011) and conclude that the first item ("The AI dashboard made its reasoning process clear to me.") had a strong focus on the local transparency of the AI (i.e., it focuses on the reasoning process), while the remaining items reflect more of a global transparency of the AI. In our case, these questions focus more on the inner mechanisms of the overall dashboard rather than the AI proposal (e.g., "It was easy for me to understand the inner workings of this AI dashboard."). Therefore, instead of a multi-dimensional measurement, we continued to use a single-item (MC_1) for the consequent analysis. The results of a *t*-test indicate that there is a significant difference between the control group ($M = 3.74$, $SD = 2.0$) and the treatment group ($M = 4.65$, $SD = 1.64$) at the 5% level ($t\,(71) = -2.20$, $p = 0.03$).

Reliability was measured using Cronbach's alpha (α). Items were deleted to further increase reliability. Based on α, all constructs are reliable and can therefore be used for hypothesis testing. "Percieved explainability" was measured based on previous work by Wang and Benbasat (2016). An established measurement instrument for perceived usefulness and trust intentions was used as suggested by Lankton et al. (2015). We also measured perceived ease of use using the measurement scale by Wang and Benbasat (2005). To measure performance, we used two objective measures: (1) the number of correct answers and (2) the time taken to complete the tasks.

Results

Analysis of group-wise differences

We tested our hypothesis using a multivariate analysis of variance (MANOVA). We conducted two MANOVAs separately: (1) the first used the experimental group as the independent variable; (2) the second used *perceived explainability* as the independent variable. *Perceived usefulness*, *perceived ease of use*, *trust intention*, and *performance (quality and time)* were used as dependent variables for both models. The Wilks test statistic (Bray et al., 1985) was used to test the first MANOVA model (see Table 7), which yielded a significant test statistic ($p < 0.000$) (Bray et al., 1985). Follow-up analysis of variance (ANOVA) revealed no significant differences in *perceived usefulness* ($p = 0.894$), *perceived ease of use* ($p = 1.000$), *trust intention* ($p = 1.000$), and in terms of *performance (quality)* ($p = 0.892$). *Performance (time)* showed a significant difference ($p < 0.000$). Effect sizes (partial η^2) (Bray et al., 1985) ranged from a low of 0.00 (*perceived ease of use*) to a high value of 0.19 (*performance (time)*).

In relation to model (2), the Wilks test statistic using perceived explainability was significant ($p < 0.004$). As a follow-up test, we conducted a series of ANOVA. There was also a significant difference in *perceived usefulness* ($p < 0.001$), *perceived ease of use* ($p < 0.006$), and *trust intentions* ($p < 0.004$). There was no significant relationship between *perceived explainability* and *performance (quality)* ($p = 1.000$) and *performance (time)* ($p = 1.000$). The analysis shows different effect sizes (partial η^2) (Bray et al., 1985) ranging from a low of 0.00 (*perfromance (time)*) to a high value of 0.19 (*perceived ease of use*). A summary of the follow-up ANOVAs is shown in Table 8.

Table 6 Measurement instrument

Construct	ID	Original item	Adapted item	German translation
Perceived explainability ($\alpha = 0.83$) (Wang & Benbasat, 2016)	MC1	This virtual adviser made its reasoning process clear to me	The AI dashboard made its reasoning process clear to me	Das KI-Dashboard machte mir seinen Entscheidungsprozess deutlich
	MC2	It was readily apparent to me how this virtual adviser generates its recommendations	It was easy for me to see how this AI dashboard generates its recommendations	Es war für mich leicht nachzuvollziehen, wie das KI-Dashboard seine Empfehlungen generiert
	MC3	I could not understand how this virtual adviser performs its job	I could **not** understand how this AI dashboard performs its job	Ich konnte **nicht** nachvollziehen, wie das KI-Dashboard seine Arbeit verrichtet
	MC4	It was easy for me to understand the inner workings of this virtual adviser	It was easy for me to understand the inner workings of this AI dashboard	Es fiel mir leicht, die Funktionsweise des KI-Dashboards zu verstehen
	MC5	I could understand why and how this virtual adviser recommends the products to me	I could understand why and how this AI dashboard makes the recommendations	Ich konnte verstehen, warum und wie dieses KI-Dashboard die Empfehlungen ausspricht
Perceived usefulness ($\alpha = 0.91$) (Lankton et al., 2015)	PU1	Using [Microsoft Access/MySNW.com3] improves my performance in [database work/online social networking]	Using the AI dashboard improves my performance	Die Nutzung des KI-Dashboards verbessert meine Performance
	PU2	Using [Microsoft Access/MySNW.com] increases my productivity in [database work/online social networking]	Using the AI dashboard increases my productivity	Die Nutzung des KI-Dashboards verbessert meine Produktivität
	PU3	Using [Microsoft Access/MySNW.com] enhances my effectiveness in [database work/online social networking]	Using the AI dashboard enhances my effectiveness	Die Verwendung des KI-Dashboards steigert meine Effektivität
Perceived ease of use ($\alpha = 0.84$) (Wang & Benbasat, 2005)	EOU1	My interaction with the virtual advisor is clear and understandable	The interaction with the AI dashboard is clear and understandable	Die Interaktion mit dem KI-Dashboard ist klar und verständlich
	EOU2	It is easy to get the virtual advisor to do what I want it to do	It is easy to get the AI dashboard to do what I want it to do	Es ist einfach, das KI-Dashboard dazu zu bringen, das zu tun, was ich will
	EOU3	Learning to use the virtual advisor was easy	Learning to use the AI dashboard was easy	Die Nutzung des KI-Dashboards war einfach zu erlernen
	EOU4	It was easy for me to find a suitable digital camera using the virtual advisor	It was easy for me to find a decision using the AI dashboard	Mithilfe des KI-Dashboards war es für mich einfach, eine Entscheidung zu treffen
	EOU5	Overall, I found that the virtual advisor is easy to use	Overall, I found that the AI dashboard is easy to use	Insgesamt finde ich, dass das KI-Dashboard einfach zu bedienen ist
Trust intentions ($\alpha = 0.90$) (Lankton et al., 2015)	TRI1	When I [do a class assignment/ network socially online], I feel I can depend on [Microsoft Access/MySNW.com]	When I use the AI dashboard, I feel I can depend on Artificial Intelligence	Wenn ich das KI-Dashboard benutze, habe ich das Gefühl, dass ich mich auf die künstliche Intelligenz verlassen kann
	TRI2	I can always rely on [Microsoft Access/MySNW.com] for [a tough class assignment/ online social networking]	I can rely on Artificial Intelligence while working with the AI dashboard	Bei der Nutzung des KI-Dashboards kann ich mich auf die künstliche Intelligenz verlassen
	TRI3	I feel I can count on [Microsoft Access/MySNW.com] when [doing my assignments/ networking online]	I feel I can count on Artificial Intelligence while working with the AI dashboard	Ich habe das Gefühl, dass ich mich auf künstliche Intelligenz verlassen kann, wenn ich das KI-Dashboard benutze
Performance (quality)	QUAL	Objectively measures (number of correct answers)		
Performance (time)	TIME	Objectively measures (overall task completion time)		

α, Cronbach's alpha

Table 7 Model (1): ANOVA results using the grouping variable as predictor

Dependent variables	SS	*df*	MS	*F*	*p*	$_{\text{partial}}\,\eta^2$	$_{\text{partial}}\,\eta^2$ 90% CI [LL, UL]	p_{adj}
Perceived usefulness	1.53	1	1.53	1.10	0.298	0.02	[0.00, 0.09]	0.894
Perceived ease of use	0.10	1	0.10	0.08	0.774	0.00	[0.00, 0.04]	1.000
Trust intention	0.12	1	0.12	0.09	0.766	0.00	[0.00, 0.04]	1.000
Performance (quality)	5.34	1	5.34	1.51	0.223	0.02	[0.00, 0.09]	0.892
Performance (time)	6.16	1	6.16	18.45	0.000	0.19	[0.08, 0.31]	0.000

LL and UL represent the lower-limit and upper-limit of the partial η^2 confidence interval, respectively. SS, sum of squares; MS, mean squares; p_{adj}, we used p value adjustment as suggested by Holm (1979)

Perceived explainability

So far, we have found that *perceived explainability* has a significant influence on important use-related variables such as *perceived usefulness* or *perceived ease of use*. However, we did not find these correlations using the experimental treatment. Since the XAI module was not the major determinant of *perceived explainability*, we conducted an additional analysis to gain further insights. Specifically, we conducted another ANOVA using *perceived explainability* as the dependent variable and included several independent variables. First, we included *group* as a dummy variable for each experimental group and *age* to examine the role of demographic differences. Second, we included *disposition to trust*, to examine predispositions to trust, and finally, we included *enjoyment* to examine the influence of hedonic motives in this context. The measurements of these items are summarized in the Appendix in Table 12.

Surprisingly, the group variable has a significant effect at the 10% level ($p = 0.010$). However, perceived enjoyment is highly significant ($p = 0.006$) and has a slightly larger effect size ($\eta^2 = 0.11$) compared to the effect size of the group variable ($\eta^2 = 0.09$). Furthermore, there is no significant effect on *disposition to trust* ($p = 0.188$) or *age* ($p = 0.164$). The ANOVA results are summarized in Table 9.

Discussion

Discussion of the results

The overall goal of this study is to investigate the effectiveness of XAI on user perception to better understand how XAI can be leveraged. This study is one of the first to examine the effects of post hoc explanations using Shapley values (Lundberg & Lee, 2017; Štrumbelj & Kononenko, 2014) using a randomized controlled experimental design. This study provides novel insights into the relationship between design manipulation and corresponding perceptual variables (i.e., perceived explainability). Most importantly, we demonstrate that providing XAI has only a small but significant effect on dependent variables. However, individuals who report a high level of perceived explainability also report higher levels in our dependent variables. We interpret these results as a strong indication that a design manipulation of XAI is important but needs to be enriched with other measures that increase perceived explainability. To the best of our knowledge, this is one of the first XAI studies to identify and assess this aspect in detail. In the following, we summarize (c.f. Table 10) and discuss the implications for theory and practice.

The first hypothesis (H_1) assumed that XAI had a higher degree of *perceived explainability* than an AI without an XAI component. We found support for this hypothesis,

Table 8 Model (2): ANOVA results using perceived explainability as predictor

Dependent variables	SS	*df*	MS	*F*	*p*	$_{\text{partial}}\,\eta^2$	$_{\text{partial}}\,\eta^2$ 90% CI [LL, UL]	p_{adj}
Perceived usefulness	18.22	1	18.22	15.85	0.000	0.19	[0.07, 0.31]	0.001
Perceived ease of use	11.35	1	11.35	10.51	0.002	0.14	[0.03, 0.26]	0.006
Trust intention	13.73	1	13.73	11.54	0.001	0.13	[0.04, 0.25]	0.004
Performance (quality)	0.27	1	0.27	0.07	0.787	0.00	[0.00, 0.04]	1.000
Performance (time)	0.01	1	0.01	0.03	0.868	0.00	[0.00, 0.03]	1.000

LL and UL represent the lower-limit and upper-limit of the partial η^2 confidence interval, respectively. SS, sum of squares; MS, mean squares; p_{adj}, we used p value adjustment as suggested by Holm (1979)

Table 9 ANOVA result using perceived explainability as dependent variable

Predictor	SS	*df*	MS	*F*	*p*	$_{\text{partial}}\,\eta^2$	$_{\text{partial}}\,\eta^2$ 90% CI [LL, UL]
(Intercept)	0.64	1	0.64	0.23	0.634		
Perceived enjoyment	22.81	1	22.81	8.20	0.006	0.11	[0.02, 0.23]
Disposition to trust	4.91	1	4.91	1.77	0.188	0.02	[0.00, 0.11]
Age	5.49	1	5.49	1.97	0.164	0.03	[0.00, 0.12]
Group (treatment)	19.47	1	19.47	7.00	0.010	0.09	[0.01, 0.21]
Error	191.84	69	2.78				

LL and UL represent the lower-limit and upper-limit of the partial η^2 confidence interval, respectively. *SS*, sum of squares; *MS*, mean squares

albeit only at 5% alpha level ($p < 0.03$). The second hypothesis (H_{2a} and H_{2b}) predicted higher levels of *perceived usefulness* and *perceived ease of use* through higher levels of *perceived explainability* which was supported in this study. This is consistent with the findings of previous studies (e.g., Meske & Bunde, 2022) that reported similar findings. We conclude that XAI indeed leads to higher levels of *perceived usefulness* and *perceived ease of use*. The third hypothesis (H_3) predicted higher levels of *trust intention* through higher levels of *perceived explainability* which was also confirmed in our study. This is in line with previous literature including the study by David et al. (2021) which reported that participants have higher levels of trust when AI systems have feature- and performance-based explanations.

The fourth hypothesis (H_4) predicted higher levels of *performance* through higher levels of *perceived explainability* which we were unable to confirm. Based on the fact that the literature to date has produced mixed results, we concluded that context-specific (confounding) factors have been overlooked so far that may explain why some studies find significant results and others report non-significant relationships. One reason could be the actual or the perceived performance of the AI (Shafti et al., 2022). Another factor could be the type of explanation (e.g., David et al., 2021; Lai et al., 2020; van der Waa et al., 2021) that leads participants to follow the system's advice more often. Even though it was not always correct, it led to a higher level of performance overall. In this study, we operationalized performance in two ways: first, by the number of correct answers, and second, by the time taken to answer the tasks. While no significant effect was measured for the number of correct answers of the treatment group compared to the control group, we found a significant effect for the time taken to complete the 20 individual tasks. From this, we can conclude that XAI gave the participants a kind of certainty to answer the question for the type of task used in the study, which ultimately led to a saving of time. We assume that this certainty is due to the perceived explainability triggered by XAI.

This study suggests that design components (e.g., XAI) without a strong influence on perceptual variables do not lead to significant effects on outcome variables. This finding is not novel, as a large body of previous literature has shown that technology design has a significant impact on users through facilitation or mediation (e.g., Wang & Benbasat, 2005). However, this highlights the relevance in the context of XAI. We therefore went one step further and analyzed how perceived explainability can be influenced by other factors other than design manipulation and included hedonic and demographic variables. The results show that the manipulation itself has an influence, but perceived enjoyment, which is an important construct from research on hedonic IS (Lowry et al., 2013) can be as important as the manipulation itself. This result is important because it shows that there are several potential moderating constructs that influence the relationships between design manipulations and outcome variables. In addition, the study assesses the extent to which the use of perceived explainability can eliminate undesirable effects such as automation bias. Perceived explainability can lead to users

Table 10 Overview of the inferential statistics

Hypothesis	Result
H_1: XAI has a higher degree of perceived explainability compared to AI without an XAI component	Supported ($p < 0.03$)
H_{2a}: A higher degree of perceived explainability results in a higher level of perceived usefulness	Supported ($p < 0.000$)
H_{2b}: A higher degree of perceived explainability results in a higher level of perceived ease of use	Supported ($p = 0.002$)
H_3: A higher degree of perceived explainability results in a higher level of trust intention	Supported ($p = 0.001$)
H_4: A higher degree of perceived explainability results in a higher level of human performance	Rejected ($p = 0.787$)

not blindly trusting AI algorithms, but questioning the extent to which the AI's decision seems understandable and correct to users. As a result, perceived explainability not only increases trust in AI, but also reduces blind trust, which has a positive impact on eliminating the undesirable effect of automation bias.

Implications for theory

So far, no specific factor has been identified in the literature that has a perfect correspondence with XAI. In this study, we used perceived explainability as the latent construct that corresponds with our manipulation. However, it is well-known that there can be a conceptual distance between the design of an artefact and the latent variable used (Niehaves & Ortbach, 2016). The measurement scale of perceived transparency that has been proposed for the context of recommendation agents (Wang & Benbasat, 2005) does not measure local explainability as pursued here. This leads to the assumption that the distance between an XAI component and the corresponding latent variables can be further reduced in the future. For example, a finer scale of measurement that distinguishes between global and local explainability might be useful. The former is arguably more relevant to a technical audience (e.g., data scientists), while local explainability is likely to be more relevant to end-users (e.g., case workers). Indeed, we argue that the target audience plays a central role in theorizing about the role of XAI. We have used a heterogeneous sample representing the end-user perspective. However, there are many scenarios where XAI is used for experts and power users. In these scenarios, the results can be very different.

This study also provides initial evidence that the XAI component is not massively used within the boundaries of our study. Rather, the participants in the XAI group used the XAI slider moderately. It can be assumed that the actual use of XAI components (i.e., components that enable some kind of user interaction) depends on the underlying task. In other scenarios (e.g., text editing tasks), the interaction with an XAI component may be different. We believe that moderate use of the XAI slider does not mean that XAI is irrelevant. In fact, we assume that the opposite is true. The fact that a participant has the opportunity to obtain additional information may be sufficient to positively influence the user's perception in general. This assumption is supported by the fact that perceived explainability has a significant antecedent function for important variables such as trust. This could be analogous to a phenomenon known from the open source field, where individuals have a higher level of trust even though a large number of potential users never actually examine the underlying code. This is mainly due to the fact that a significant number of users do not have the technical skills to review the software code. This aspect also underlines the central role of the target group using an XAI module.

This study shows that it is important to acknowledge boundary conditions (Gregor & Hevner, 2013) when theorizing about XAI. This includes the differences between specific tasks and how the XAI is implemented. Previous literature has already shown that different explanation strategies (i.e., rule-based vs. example-based) have different effects on users (van der Waa et al., 2021). Similarly, the underlying task may also look very different in terms of the target group (e.g., normal user vs. data scientists) which should be taken into account. This is consistent with previous IS theories that focus on the fit between a task and its supporting technology (Goodhue & Thompson, 1995). We argue that a strong fit is also central to XAI research to get the most out of XAI in terms of user perception.

Implications for practice

In addition to theoretical contributions, this research also has important implications for organizations. Most importantly, we have shown that the use of AI with the additional explanations (e.g., using SHAP (Bowen & Ungar, 2020) or LIME (Ribeiro et al., 2016)) can lead to positive effects. For XAI to lead to positive effects on objective measures (i.e., performance) and on user-related measures, it is not sufficient to use XAI, but to improve perceived explainability.

With this in mind, managers should support the implementation of XAI components and accompanying measures (e.g., trainings) to reap the benefits in terms of user perception of the XAI dashboard. Since users perceive XAI as more useful, user-friendly, and trustworthy than AI algorithms without an explanation, this is an important factor in improving the acceptance of these systems. Especially when AI systems are used in sensitive areas such as medical diagnosis decisions (Jussupow et al., 2021) or autonomous driving (Muhammad et al., 2021), it is (even) more important that users can trust the systems. By providing explanations, not only users but also programmers gain better insight into how the algorithms work and enable more effective debugging. This offers the opportunity not only to program more robust and advanced algorithms but also to identify and eliminate potential biases (e.g., gender biases in the hiring process) (Hsu, 2020). More generally, explainability has also been identified as an important success factor for the adoption of artificial intelligence in organizations (Hamm & Klesel, 2021), making it an important aspect from a strategic perspective as well.

It is also worth noting that organizations are encouraged and legally compelled to fulfil ethical guidelines for trustworthy AI from the independent High-Level

Expert Group on Artificial Intelligence (AI HLEG) commissioned by the European Commission, which stipulate that AI algorithms must be transparent and explainable (HLEG-AI, 2019). Considering that there are desirable outcomes (such as higher degrees in user perception), the addition of XAI components becomes a necessity, so organizations are well-advised to implement XAI components.

Limitations

As with any academic study, the results of this study have their limitations. First, because AI applications are not well established yet, there is a high number of respondents who have little experience with AI. 60.8% of participants have no explicit role related to AI in a professional setting (c.f. Table 5). Second, the experimental study was conducted in the context of a classification task of signatures by participants. We chose to focus this study on a classification task because it is a typical business task and highly relevant. However, this may limit the generalizability of the results. In others' tasks, participants may have a higher need to consult an XAI module than was observed here. This can have a significant impact on perceptual and behavioral outcomes. In addition, the choice of the post hoc explanation method SHAP may also be a limitation, as it may change the way users perceive an XAI component. Future studies can therefore extend this work by testing our hypothesis with other tasks and explanation methods (e.g., using text heavy tasks and/ or by using other explanation techniques). Third, it should be noted that the study is limited in terms of the incentive of the participants. This limitation is due to the lack of consequences from correctly or incorrectly classified signatures. Future research can address this issue by conducting a field experiment.

Appendix

XAI slider usage

We analyzed how often the participants used the XAI slider. Figure 4 provides an overview of the cumulated use of the XAI slider in the treatment group. To ensure that we only include observations that used the XAI slider, we dropped all responses where the slider was used in less than 5 times within the experimental task. This threshold is close to the lower 40% quantile and leaves a considerable amount of observations in the dataset which used the slider but only to a minor extend. We did not delete observations on the upper limit ($\text{slideruse}_{max} = 86$) because using the slider 4-times on every task on average is still in a reasonable range in operative systems.

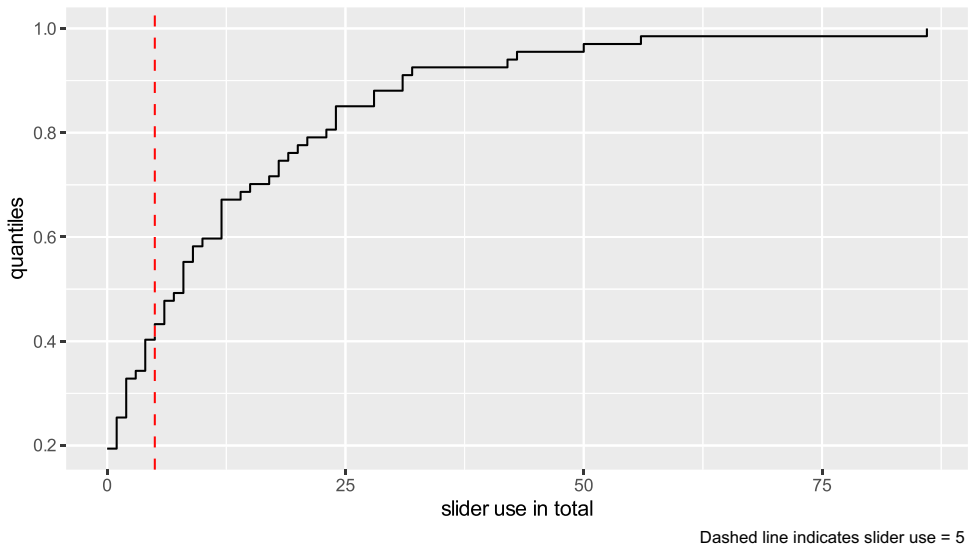

Fig. 4 Cumulated use of the XAI slider ($N = 67$)

An overview of the slider use is summarized in Table 11. In the original sample, participants in the treatment group used the slider 12.72 times on average (SD = 15.9). After excluding observations that used the slider less than 5 times in total, the average slider use is 20.48 (SD = 16.53).

Table 11 Slider use before and after sample reduction

ID	group	*n*	Mean	sd	Median	Min	Max	Skew	Kurtosis
1	**0 (control)**	**39**	**0**	**0**	**0**	**0**	**0**	**NA**	**NA**
2	1 (treatment)	67	12.72	15.9	8	0	86	2.11	5.63
3	Observations for exclusion[1]	27	1.22	1.48	1	0	4	0.8	-0.82
4	**1 (treatment) without #3[2]**	**40**	**20.48**	**16.53**	**16**	**5**	**86**	**1.9**	**4.29**

[1]We excluded all observations that used the XAI slider less than 5 times within the experiment

[2]The remaining 40 observations are used for the subsequent analysis

Bold: final data

We also analyzed how often the slider is used based on the task—which was randomized for every participant—and based on the pair of images shown. Figure 5 shows how often the slider was used based on the task. Figure 6 shows how often the slider was used based on the pair of images.

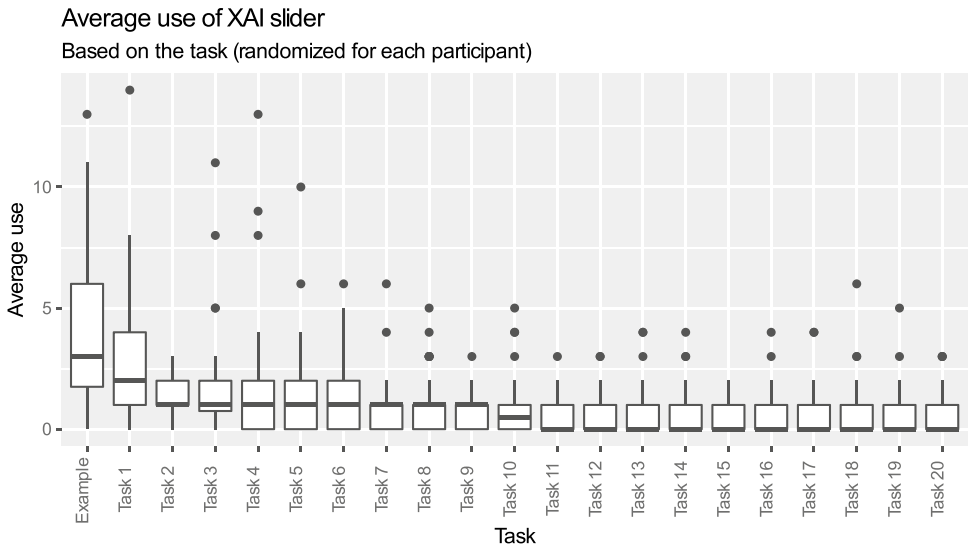

Fig. 5 Slider use based on the task

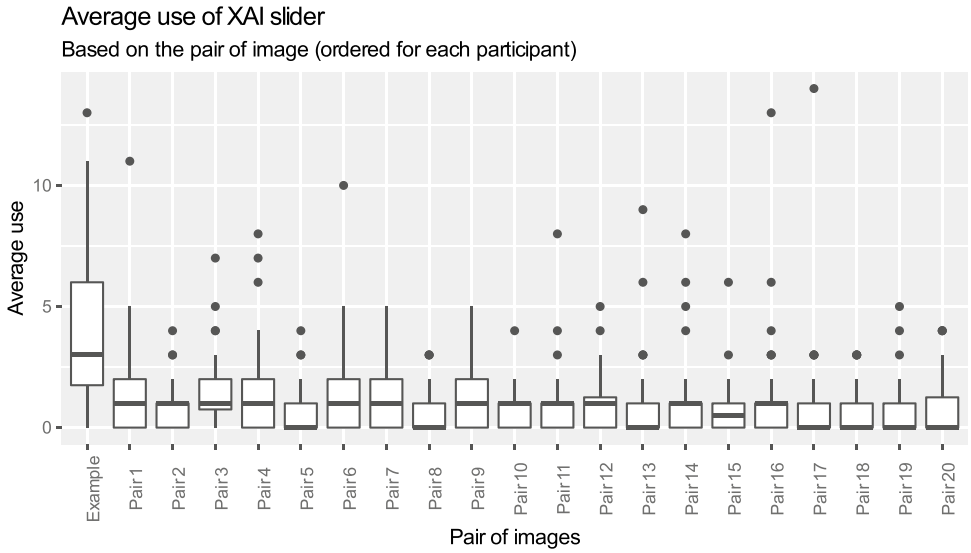

Fig. 6 Slider use based on the pair of images

Manipulation check

The following Fig. 7 shows the mean-wise differences between each indicator of *perceived explainability*.

Fig. 7 Manipulation check

Extension of the measurement instrument

Table 12 Extension of the measurement instrument

Construct	ID	Original item	Adapted item	German translation
Disposition to trust technology ($\alpha = 0.69$) (Lankton et al., 2015)	DPT1	My typical approach is to trust new information technologies until they prove to me that I shouldn't trust them	My typical approach is to trust AI dashboards until they prove to me that I shouldn't trust them	Mein typischer Ansatz ist es, KI-Dashboards zu vertrauen, bis sie mir beweisen, dass ich ihnen nicht vertrauen sollte
	DPT2	I usually trust in information technology until it gives me a reason not to	I usually trust AI dashboards until they give me a reason not to	Normalerweise vertraue ich KI-Dashboards, bis sie mir einen Grund geben, es nicht zu tun
	DPT3	I generally give an information technology the benefit of the doubt when I first use it	I generally give AI dashboards the benefit of the doubt when I first use it	Wenn ich KI-Dashboards zum ersten Mal verwende, bin ich im Allgemeinen skeptisch
Enjoyment ($\alpha = 0.86$) (Lankton et al., 2015)	ENJ1	I find using [Microsoft Access /MySNW.com] to be enjoyable	I find using the AI dashboard to be enjoyable	Ich finde die Verwendung des KI-Dashboards sehr angenehm
	ENJ2	The actual process of using [Microsoft Access/MySNW.com] is pleasant	The actual process of using the AI dashboard is pleasant	Die eigentliche Nutzung des KI-Dashboards ist angenehm
	ENJ3	I have fun using [Microsoft Access/MySNW.com]	I have fun using the AI dashboard	Ich habe Spaß an der Nutzung des KI-Dashboards

Data Availability The data and analysis for this study have been made openly available on the Open Science Framework (OSF) and can be accessed via the following link: https://osf.io/msqzy/. (https://doi.org/10.17605/OSF.IO/MSQZY).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., & Zheng, X. (2016). *TensorFlow: Large-scale machine learning on heterogeneous distributed systems*. https://doi.org/10.48550/arXiv.1603.04467.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *International Conference on Engineering and Technology (ICET)* (pp. 1–6). Antalya, Turkey. https://doi.org/10.1109/ICEngTechnol.2017.8308186

Alipour, K., Ray, A., Lin, X., Schulze, J. P., Yao, Y., & Burachas, G. T. (2020a). *The impact of explanations on AI competency prediction in VQA*. arXiv:2007.00900.

Alipour, K., Schulze, J. P., Yao, Y., Ziskind, A., & Burachas, G. (2020b). *A study on multimodal and interactive explanations for visual question answering*. arXiv:2003.00431.

Alufaisan, Y., Marusich, L. R., Bakdash, J. Z., Zhou, Y., & Kantarcioglu, M. (2020). Does explainable artificial intelligence improve human decision-making? *Conference on Artificial Intelligence (AAAI)* (pp. 6618–6626). Virtual Conference. https://doi.org/10.31234/osf.io/d4r9t

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Asatiani, A., Malo, P., Nagbøl, P., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2020). Challenges of explaining the behavior of black-box AI systems. *MIS Quarterly Executive, 19*(4), 259–278.

Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *IEEE Potentials, 13*(4), 27–31. https://doi.org/10.1109/45.329294

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods & Research, 10*(2), 141–163. https://doi.org/10.1177/004912418101000205

Bowen, D., & Ungar, L. (2020). *Generalized SHAP: Generating multiple types of explanations in machine learning*. arXiv:2006.07155.

Bray, J. H., Maxwell, S. E., & Maxwell, S. E. (1985). *Multivariate analysis of variance*. SAGE Publications.

Chatla, S., & Shmueli, G. (2017). An extensive examination of regression models with a binary outcome variable. *Journal of the Association for Information Systems, 18*(4), 340–371. https://doi.org/10.17705/1jais.00455

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Routledge. https://doi.org/10.4324/9780203771587

Confalonieri, R., Weyde, T., Besold, T. R., & Martín, F. M. del P. (2019). *Trepan reloaded: A knowledge-driven approach to explaining artificial neural networks*. arXiv:1906.08362.

David, D. B., Resheff, Y. S., & Tron, T. (2021). Explainable AI and adoption of financial algorithmic advisors: an experimental study. *Conference on AI, Ethics, and Society (AAAI/ACM)* (pp. 390–400). Virtual Conference. https://doi.org/10.1145/3461702.3462565

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Dominguez, V., Donoso-Guzmán, I., Messina, P., & Parra, D. (2020). Algorithmic and HCI aspects for explaining recommendations of artistic images. *ACM Transactions on Interactive Intelligent Systems, 10*(4), 1–31. https://doi.org/10.1145/3369396

Druce, J., Harradon, M., & Tittle, J. (2021). *Explainable artificial intelligence (XAI) for increasing user trust in deep reinforcement learning driven autonomous systems*. arXiv:2106.03775.

European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC* (General Data Protection Regulation), OJ 2016 L 119/1, 2016., 1–88.

Förster, M., Klier, M., Kluge, K., & Sigler, I. (2020). Fostering human agency: A process for the design of user-centric XAI systems. *International Conference on Information Systems (ICIS)* (pp. 1–17). Hyderabad, India.

Galhotra, S., Pradhan, R., & Salimi, B. (2021). *Explaining black-box algorithms using probabilistic contrastive counterfactuals*. arXiv:2103.11972.

Gefen, D., Karahanna, E., & Straub, D. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly, 27*(1), 51–90. https://doi.org/10.2307/30036519

Ghanvatkar, S., & Rajan, V. (2019). Deep recurrent neural networks for mortality prediction in intensive care using clinical time series at multiple resolutions. *International Conference on Information Systems (ICIS)* (pp. 1–9). Munich, Germany.

Gilpin, L. H., Testart, C., Fruchter, N., & Adebayo, J. (2019). *Explaining explanations to society*. arXiv:1901.06560.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association, 19*(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

Golding, A., & Rosenbloom, P. (1991). Improving rule-based systems through case-based reasoning. *National Conference on Artificial Intelligence (AAAI)* (pp. 22–27). Anaheim, United States.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly, 19*(2), 213–236. https://doi.org/10.2307/249689

Gramegna, A., & Giudici, P. (2021). SHAP and LIME: An evaluation of discriminative power in credit risk. *Frontiers in Artificial Intelligence, 4*, 1–6. https://doi.org/10.3389/frai.2021.752558

Gregor, S., & Hevner, A. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly, 37*(2), 337–356. https://doi.org/10.25300/MISQ/2013/37.2.01

Hamm, P., & Klesel, M. (2021). Success factors for the adoption of artificial intelligence in organizations: A literature review. *Americas Conference on Information Systems (AMCIS)* (pp. 1–10). Montreal, Canada.

HLEG-AI. (2019). *Ethics guidelines for trustworthy artificial intelligence. Brussels: independent high-level expert group on artificial intelligence set up by the European Commission*. FUTURIUM - European Commission. Text. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1. Accessed 26 April 2022

Holliday, D., Wilson, S., & Stumpf, S. (2016). User trust in intelligent systems: A journey over time. *International Conference on Intelligent User Interfaces* (pp. 164–168). Sonoma, United States. https://doi.org/10.1145/2856767.2856811

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Hsu, J. (2020). Can AI hiring systems be made antiracist? Makers and users of AI-assisted recruiting software reexamine the tools' development and how they're used. *IEEE Spectrum, 57*(9), 9–11. https://doi.org/10.1109/MSPEC.2020.9173891

Hussein, W., Salama, M. A., & Ibrahim, O. (2016). Image processing based signature verification technique to reduce fraud in financial institutions. *International Conference on Circuits, Systems, Communications and Computers (CSCC)* (pp. 1–5). Corfu Island, Greece. https://doi.org/10.1051/matecconf/20167605004

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Information Systems Research, 32*(3), 713–735. https://doi.org/10.1287/isre.2020.0980

Kim, B., Glassman, E., Johnson, B., & Shah, J. (2015). *iBCM: Interactive Bayesian case model empowering humans via intuitive interaction*. Computer Science and Artificial Intelligence Laboratory Technical Report, 1–10.

Kizilcec, R. F. (2016). How much information?: effects of transparency on trust in an algorithmic interface. *Conference on Human Factors in Computing Systems (CHI)* (pp. 2390–2395). San Jose, United States. https://doi.org/10.1145/2858036.2858402

Lai, V., Liu, H., & Tan, C. (2020). "Why is 'Chicago' deceptive?" Towards building model-driven tutorials for humans. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, (pp. 1–13). https://doi.org/10.1145/3313831.3376873

Lankton, N., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems, 16*(10), 880–918. https://doi.org/10.17705/1jais.00411

Lowry, P., Gaskin, J., Twyman, N., Hammer, B., & Roberts, T. (2013). Taking "fun and games" seriously: Proposing the hedonic-motivation system adoption model (HMSAM). *Journal of the Association for Information Systems, 14*(11), 617–671. https://doi.org/10.17705/1jais.00347

Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. arXiv:1705.07874.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly, 35*(2), 293–334. https://doi.org/10.2307/23044045

Mao, J., & Benbasat, I. (2000). The use of explanations in knowledge-based systems: Cognitive perspective and a process-tracing analysis. *Journal of Management Information Systems, 17*(2), 153–179. https://doi.org/10.1080/07421222.2000.11045646

Meske, C., & Bunde, E. (2022). Design principles for user interfaces in AI-based decision support systems: The case of explainable hate speech detection. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10234-5

Miller, T. (2018). *Explanation in artificial intelligence: insights from the social sciences*. arXiv:1706.07269.

Muddamsetty, S. M., Jahromi, M. N. S., & Moeslund, T. B. (2020). SIDU: Similarity difference and uniqueness method for explainable AI. *IEEE International Conference on Image Processing (ICIP)* (pp. 3269–3273). IEEE: Abu Dhabi, United Arab Emirates.

Muhammad, K., Ullah, A., Lloret, J., Ser, J. D., & de Albuquerque, V. H. C. (2021). Deep learning for safe autonomous driving: Current challenges and future directions. *IEEE Transactions on Intelligent Transportation Systems, 22*(7), 4316–4336. https://doi.org/10.1109/TITS.2020.3032227

Naderifar, M., Goli, H., & Ghaljaie, F. (2017). Snowball sampling: A purposeful method of sampling in qualitative research. *Strides in Development of Medical Education, 14*(3), 1–4. https://doi.org/10.5812/sdme.67670

Niehaves, B., & Ortbach, K. (2016). The inner and the outer model in explanatory design theory: The case of designing electronic feedback systems. *European Journal of Information Systems, 25*(4), 303–316. https://doi.org/10.1057/ejis.2016.3

Quinlan, J. R. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies, 27*(3), 221–234. https://doi.org/10.1016/S0020-7373(87)80053-6

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

Rai, D. (2022). *Handwritten signatures*. https://www.kaggle.com/datasets/divyanshrai/handwritten-signatures. Accessed 24 Apr 2022.

Ray, A., Yao, Y., Kumar, R., Divakaran, A., & Burachas, G. (2019). *Can you explain that? Lucid explanations help human-AI collaborative image retrieval*. arXiv:1904.03285.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why should I trust you?": Explaining the predictions of any classifier*. arXiv:1602.04938.

Schneider, J., & Handali, J. (2019). *Personalized explanation in machine learning: A conceptualization*. arXiv:1901.00770.

Schrills, T., & Franke, T. (2020). Color for characters - Effects of visual explanations of AI on trust and observability. International Conference on Human-Computer Interaction (HCI) (pp. 121–135). Copenhagen, Denmark.In H. Degen & L. Reinerman-Jones (Eds.), *Artificial intelligence in HCI* (Vol. 12217, pp. 121–135). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-50334-5_8

Shafti, A., Derks, V., Kay, H., & Faisal, A. A. (2022). *The response shift paradigm to quantify human trust in AI recommendations*. arXiv:2202.08979.

Sperrle, F., El-Assady, M., Guo, G., Chau, D. H., Endert, A., & Keim, D. (2020). *Should we trust (X)AI? Design dimensions for structured experimental evaluations*. https://doi.org/10.48550/arXiv.2009.06433.

Streamlit. (2022). *Streamlit - The fastest way to build and share data apps*. https://streamlit.io/. Accessed 24 Apr 2022.

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems, 41*, 647–665. https://doi.org/10.1007/s10115-013-0679-x

Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets, 31*(2), 447–464. https://doi.org/10.1007/s12525-020-00441-4

van der Waa, J., Nieuwburg, E., Cremers, A., & Neerincx, M. (2021). Evaluating XAI: A comparison of rule-based and example-based explanations. *Artificial Intelligence, 291*, 103404–103423. https://doi.org/10.1016/j.artint.2020.103404

Venkatesh, M., Davis, & Davis. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J., & Xu, X. (2016). Unified theory of acceptance and use of technology: A synthesis and the road ahead. *Journal of the Association for Information Systems, 17*(5), 328–376. https://doi.org/10.17705/1jais.00428

Wang, W., & Benbasat, I. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems, 6*(3), 72–101. https://doi.org/10.17705/1jais.00065

Wang, W., & Benbasat, I. (2016). Empirical assessment of alternative designs for enhancing different types of trusting beliefs in online recommendation agents. *Journal of Management Information Systems, 33*(3), 744–775. https://doi.org/10.1080/07421222.2016.1243949

Weitz, K., Schiller, D., Schlagowski, R., Huber, T., & André, E. (2019). "Do you trust me?": Increasing user-trust by integrating virtual agents in explainable AI interaction design. *International Conference on Intelligent Virtual Agents (IVA)* (pp. 7–9). Paris, France. https://doi.org/10.1145/3308532.3329441

Weitz, K., Schiller, D., Schlagowski, R., Huber, T., & André, E. (2021). "Let me explain!": Exploring the potential of virtual agents in explainable AI interaction design. *Journal on Multimodal User Interfaces, 15*(2), 87–98. https://doi.org/10.1007/s12193-020-00332-0

Wong, M. H., LEE, L. H., & Hui, P. (2020). GANStick: US stock forecasting with GAN-generated candlesticks. *International Conference on Information Systems (ICIS)* (pp. 1–17). Hyderabad, India.

Zhou, Y., Zheng, J., Hu, H., & Wang, Y. (2021). Handwritten signature verification method based on improved combined features. *Applied Sciences, 11*(13), 5867–5881. https://doi.org/10.3390/app11135867