Wanner, Jonas; Herm, Lukas-Valentin; Heinrich, Kai; Janiesch, Christian

The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study

Electronic Markets

Provided in Cooperation with:
Springer Nature

Suggested Citation: Wanner, Jonas; Herm, Lukas-Valentin; Heinrich, Kai; Janiesch, Christian (2022) : The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study, Electronic Markets, ISSN 1422-8890, Springer, Berlin, Heidelberg, Vol. 32, Iss. 4, pp. 2079-2102, https://doi.org/10.1007/s12525-022-00593-5

This Version is available at:
https://hdl.handle.net/10419/311817

Electronic Markets (2022) 32:2079–2102
https://doi.org/10.1007/s12525-022-00593-5

RESEARCH PAPER

The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study

Jonas Wanner[1] · **Lukas-Valentin Herm**[1] · **Kai Heinrich**[2] · **Christian Janiesch**[3]

Received: 29 March 2022 / Accepted: 7 September 2022 / Published online: 24 October 2022

Abstract

Contemporary decision support systems are increasingly relying on artificial intelligence technology such as machine learning algorithms to form intelligent systems. These systems have human-like decision capacity for selected applications based on a decision rationale which cannot be looked-up conveniently and constitutes a black box. As a consequence, acceptance by end-users remains somewhat hesitant. While lacking transparency has been said to hinder trust and enforce aversion towards these systems, studies that connect user trust to transparency and subsequently acceptance are scarce. In response, our research is concerned with the development of a theoretical model that explains end-user acceptance of intelligent systems. We utilize the unified theory of acceptance and use in information technology as well as explanation theory and related theories on initial trust and user trust in information systems. The proposed model is tested in an industrial maintenance workplace scenario using maintenance experts as participants to represent the user group. Results show that acceptance is performance-driven at first sight. However, transparency plays an important indirect role in regulating trust and the perception of performance.

Keywords User acceptance · Intelligent system · Artificial intelligence · Trust · System transparency

JEL classification C6 · C8 · M15

Introduction

Would you trust a superintelligent computer's recommendation on a critical decision such as turning off crucial machinery if it offered no transparency into the decision-making?

Intelligent systems with human-like cognitive capacity have been a promise of artificial intelligence (AI) research for decades. Due to the rise and sophistication of machine learning (ML) technology, intelligent systems are becoming a reality and can now solve complex cognitive tasks (Benbya et al., 2021). They are being deployed rapidly in practice (Janiesch et al., 2021). More recently, deep learning allows tackling even more compound problems such as playing Go (Silver et al., 2016) or driving autonomously in real traffic (Grigorescu et al., 2020). On the downside, the decision rationale of intelligent systems based on deep learning is not per se interpretable to humans and requires explanations. That is, while the decision is documented, its rationale is complex and essentially intransparent from the point of human perception constituting a perceived black box (Kroll, 2018).

Further, users tend to credit anthropomorphic traits to an intelligent system subconsciously to ascribe the system's AI a sense of efficacy (Epley et al., 2007; Pfeuffer et al., 2019). In this respect, intelligent systems are credited with the trait of agency (Baird & Maruping, 2021), creating a situation comparable to the principal-agent problem as their decision rationale is self-trained (self-interest) and intransparent to the principal. This results in an information asymmetry between the user (principal) and the intelligent system (agent). This information asymmetry constitutes a major barrier for intelligent system acceptance and initial trust in intelligent systems (McKnight et al., 2002; Shin

Responsible Editor: Veda Storey

✉ Lukas-Valentin Herm
lukas-valentin.herm@uni-wuerzburg.de

[1] Julius-Maximilians-Universität Würzburg, Sanderring 2, Postal Code 97070 Würzburg, Germany

[2] Otto-von-Guericke-Universität Magdeburg, Magdeburg, Germany

[3] TU Dortmund University, Dortmund, Germany

et al., 2020), because the system cannot provide credible, meaningful information about or affective bonds with the agent (Bigley & Pearce, 1998).

Altogether, this lack of transparency and, subsequently, trust can be a hindrance when delegating tasks or decisions to an intelligent system (Shin, 2020a, 2021). More specifically, the acceptance and adoption of AI currently remains rather hesitant (Chui & Malhotra, 2018; Milojevic & Nassah, 2018; Wilkinson et al., 2021). The result is observable user behavior, such as algorithm aversion, where the user will not accept an intelligent system in a professional context even though it outperforms human co-workers (Burton et al., 2020). While this can be attributed at least partially to lack of control and the information asymmetry due to its black-box nature, we also observe the inverse, algorithm appreciation, and, thus, acceptance and use of intelligent systems in other scenarios (Herm et al., 2021; Logg et al., 2019).

This is a crucial point, as intelligent systems can only be effective if users are willing to engage with them actively and have confidence in their recommendations. Consequently, it is of great importance to understand what the intended users of such systems expect, and which influences have to be considered for mitigation of algorithm aversion and successful acceptance (Mahmud et al., 2022).

While the factors of performance, trust, and transparency have been connected to user perception of technology, a rigorous study to connect them to intended usage behavior of intelligent systems is missing (Venkatesh, 2022). With our research, we expand the body of knowledge on the acceptance of intelligent systems by considering system transparency and trust in combination as pivotal factors (e.g., Adadi & Berrada, 2018; Mohseni et al., 2021; Rudin, 2019). Furthermore, we extend beyond the measurement of direct effects and investigate their mediating, indirect roles regarding the drivers of behavioral intention.

We build a theoretical model by synthesizing explanation theory, user trust theory, and the unified theory of acceptance (UTAUT) to fit the nature of intelligent systems and to understand the human attitude towards them.

Thereby, we offer three key contributions. First, we provide an explanatory model for the context of intelligent systems. It can serve as a starting point for research in distinct fields. Second, by validating established hypotheses, we provide a better understanding of the actual factors that influence the user's acceptance of intelligent systems and explain user behavior towards AI-based systems in general. This allows both the use of this knowledge for the (vendor's) design and implementation of intelligent systems and its use for the (customer's) process of software selection. Third, by establishing new hypotheses that regard the nature of trust and transparency in system acceptance, we take into account the unique attributes of intelligent systems related to the perceived black-box nature of their underlying rationales (Herm et al., 2022; Mohseni et al., 2021).

Our paper is structured as follows: In Section "Theoretical background", we introduce the theoretical background for our research. In Section "Methodological overview", we describe our research design. In Section "Research theorizing", we describe our research theorizing. This includes the review of existing UTAUT research on trust and system transparency as well as the hypothesis and items of the derived constructs and relationships. In Section "Study and results", we describe the empirical testing of the theoretical derivations and their results. Finally, we discuss the implications for theory and practice in Section "Discussion", before we summarize and offer an outlook on future research in "Conclusion and outlook".

Theoretical background

Artificial intelligence and intelligent systems

AI is an umbrella term for any technique that enables computers to imitate human intelligence and replicate or even surpass human decision-making capacity for complex tasks (Russell & Norvig, 2021). This entails that the meaning and scope of AI is constantly being refined as technology evolves while the reference point of human intelligence remains relatively static (Berente et al., 2021).

In the past, AI focused on handcrafted inference models known as symbolic AI or the knowledge-based approach (Goodfellow et al., 2016). While this approach is inherently transparent and enabled trust in the decision process, it is limited by the human's capability to explicate their tacit knowledge relevant to the task (Brynjolfsson & Mcafee, 2017). More recently, ML and deep learning algorithms have overcome these limitations by automatically building analytical models from training data (Janiesch et al., 2021). However, the resulting advanced analytical models often lack immediate (system) transparency constituting an information asymmetry to the user.

Intelligent systems are software systems that make use of AI technology. They exhibit at least two traits towards end-user that separate them from traditional commercial-off-the-shelf software with decision support such as accounting information systems or enterprise resource planning software. First, intelligent systems enable decision-making with human-like or even super-human cognitive abilities for certain tasks (McKinney et al., 2020). Second, the decision rationale of intelligent systems cannot be looked up conveniently.

That is, intelligent systems do not use handcrafted and thus traceable, deterministic rulesets to make decisions, but intelligent systems exhibit complex probabilistic behavior

with superior performance that was learned based on data input rather than explicitly programmed, for example using ML algorithms (Janiesch et al., 2021; Mohseni et al., 2021). While the underlying relations in the analytical models can be analyzed by experts given enough time and resources (and technically constitute white-box decision making), no end-user is capable of extracting explanations on the decision process or individual decisions. Rather, the model constitutes a black box from the perspective of the end-user (Savage, 2022).

This circumstance leads to an increased tension between human agency and machine agency during decision making (Sundar, 2020). In this context, intelligent systems inherit characteristics associated with new, revolutionary technologies, including technology-related anxiety and alienation of labor through a lack of comprehension and a lack of trust (Mokyr et al., 2015). Hence, when facing these properties, due to effecting motivation, the human has a "desire to reduce uncertainty and ambiguity, at least in part with the goal of attaining a sense of predictability and control in one's environment" (Epley et al., 2007).

Transparency and trust in intelligent systems

Trust in the context of technology acceptance has widely been studied and derived from organizational trust towards humans. Notably, besides the core construct of the cognition-based trust in the ability of the system, additional affect-based trust aspects like the general propensity to trust technology and the believed goodwill or benevolence of the trustee towards the trustor exist (von Eschenbach, 2021). While it can be argued that the system has no ill will by itself, in the case of black-box systems, we cannot observe whether it acts as intended, possibly hindering initial trust formation (Dam et al., 2018).

Building trust in new technologies is initially hindered by unknown risk factors and thus uncertainty, as well as a lack of total user control (McKnight et al., 2011; Shneiderman, 2020). The main factors in building initial trust are the ability of the system to show possession of the functionalities needed, to convey that they can help the user when needed, and to operate consistently (McKnight et al., 1998; Paravastu & Ramanujan, 2021).

For human intelligence, it is generally an important aspect to be able to explain the rationale behind one's decision, while simultaneously, it can be considered as a prerequisite for establishing a trustworthy relationship (Samek et al., 2017). Thus, observing a system's behavior in terms of transparency plays an important role. In IS research, it has been argued that transparency can increase the cognition-based part of trust towards the system (Shin et al., 2020). In addition, system transparency is assumed to have an indirect influence on IS acceptance via trust in the context of recommending a favorable decision to the user (Wilkinson et al., 2021).

While general performance indicators of ML models can be used to judge the recommendation performance of an intelligent system, the learning process and the inner view of the intelligent system towards the problem can be different from the human understanding, generating a dissonance, suggesting system performance by itself is not sufficient as a criterion (Miller, 2019).

Thus, the ML model underlying an intelligent system cannot address these factors itself. Therefore, it is widely suggested that this issue can be alleviated or resolved by providing an overall system transparency by offering explanations of the decision-making process (i.e., global explanations) as well as explanations of individual recommendations (i.e., local explanations) (Mohseni et al., 2021). That is, in recent AI-based IS literature the perceived explanation quality is defined as the level of explainability (Herm et al., 2021). The field of explainable AI (XAI) offers augmentations or surrogate models that can explain the behavior of intelligent systems based on black-box ML models (Injadat et al., 2021).

Altogether, the rise of design-based literature on explainable, intelligent systems suggests that the lack of transparency of deep learning algorithms poses a problem for user acceptance, rendering the systems inefficacious (Bentele & Seidenglanz, 2015; Sardianos et al., 2021). It is reasonable to assume that system transparency or its explainability, as well as trust, play a central role when investigating socio-technical aspects of technology acceptance. Furthermore, both seem to be interrelated to one another (Shin, 2021). Nevertheless, it is not evident to what extent an increase in the user's perceived system explainability improves the user's trust factor or how this affects the user's technology acceptance of intelligent systems (Shin et al., 2020; Wang & Benbasat, 2016).

Technology acceptance

Technology acceptance has been widely studied in the context of several theoretical frameworks. In its core idea, a behavioral study is used to draw conclusions regarding the willingness of a target group to accept an investigative object (Jackson et al., 1997; Venkatesh et al., 2012).

Davis (1989) utilized the *Theory of Reasoned Action* to propose the *Technology Acceptance Model* (TAM) that explains the actual use of a system through the perceived usefulness and perceived ease of use of that system. It was later updated to include other factors such as subjective norms (Marangunić & Granić, 2015). An extension of the theory that includes the additional determinant is the *Theory of Planned Behavior* by Taylor and Todd (1995). As a competing perspective of explanation, the *Model of PC Utilization* includes determinants that are less abstract

to the technology application environment, such as job-fit, complexity, affect towards use, and facilitating conditions that reflect on the actual objective factors from the application environment, it can differ largely from case to case. The *Innovation Diffusion Theory* by Rogers (2010) is specifically tailored to new technologies and the perception of several determinants like a gained relative advantage, ease of use, visibility, and compatibility. Furthermore, the *Social Cognitive Theory* was extended to explain individual technology acceptance by determinants like outcome expectancy, self-efficacy, affect, and anxiety (Bandura, 2001).

Venkatesh et al. (2003) combined those theories in the *Unified Theory of Acceptance and Use of Technology* (UTAUT). It provides a holistic model that includes adoption theories for new technologies and approaches to computer usage that capture the actual factors of the implementation environment. Compared to ABM, UTAUT is favored due to its ability to explain the variance within the dependent variable more precisely (Demissie et al., 2021). Here, *behavioral intention* (BI) acts as an explanatory factor for the actual user behavior. Determinants of *BI* in the UTAUT model are, for example, *performance expectancy* (PE), *effort expectancy* (EE), or social influence. UTAUT has been used extensively to explain and predict acceptance and use in a multitude of scenarios (Williams et al., 2015).

Despite the fundamental theoretical foundation, it has become a common practice to form the measurement model for a specific use case given by multiple iteration cycles (e.g., Yao & Murphy, 2007). Thus, many authors modify their UTAUT model (e.g., Oliveira et al., 2014; Shahzad et al., 2020; Slade et al., 2015). Typically, an extension is applied in three different ways (Slade et al., 2015; Venkatesh et al., 2012): i) using UTAUT for the evaluation of new technologies or new cultural settings (e.g., Gupta et al., 2008); ii) adding new constructs to expand the investigation scope of UTAUT (e.g., Baishya & Samalia, 2020); and/or iii) to include exogenous predictors for the proposed UTAUT variables (e.g., Neufeld et al., 2007). Furthermore, many contributions such as Esfandiari and Sokhanvar (2016) or Albashrawi and Motiwalla (2017), combine multiple extension methods to construct a new model. Lastly, Blut et al. (2021) introduce four new broad predictors for future technology acceptance and use. However, they do not incorporate the idea of black-box systems common with contemporary AI.

Related work of technology acceptance research is primarily focused on e-commerce, mobile technology, and social media (Rad et al., 2018). The intersection with AI innovations is rather small yet. Despite contributions for autonomous driving (e.g., Hein et al., 2018; Kaur & Rampersad, 2018) or healthcare (e.g., Fan et al., 2018; Portela et al., 2013), only a few studies exist for industrial applications, such as on the acceptance of intelligent robotics in production processes (e.g., Bröhl et al., 2016; Lotz et al., 2019). Also, there is the intention to understand the acceptance of augmented reality (Jetter et al., 2018).

Consequently, knowledge about the technology acceptance of intelligent systems is still limited. In particular, *trust* and *system transparency* have not been considered in conjunction as potential factors for technology acceptance of intelligent systems.

Methodological overview

The focus of our research problem is the acceptance of an intelligent system from an end-user perspective. It is located at the intersection of two fields of interest: technology acceptance and AI, more specifically XAI.

Figure 1 presents our methodological frame to develop our UTAUT model for the context of intelligent systems. It corresponds to the procedure presented by Šumak et al.

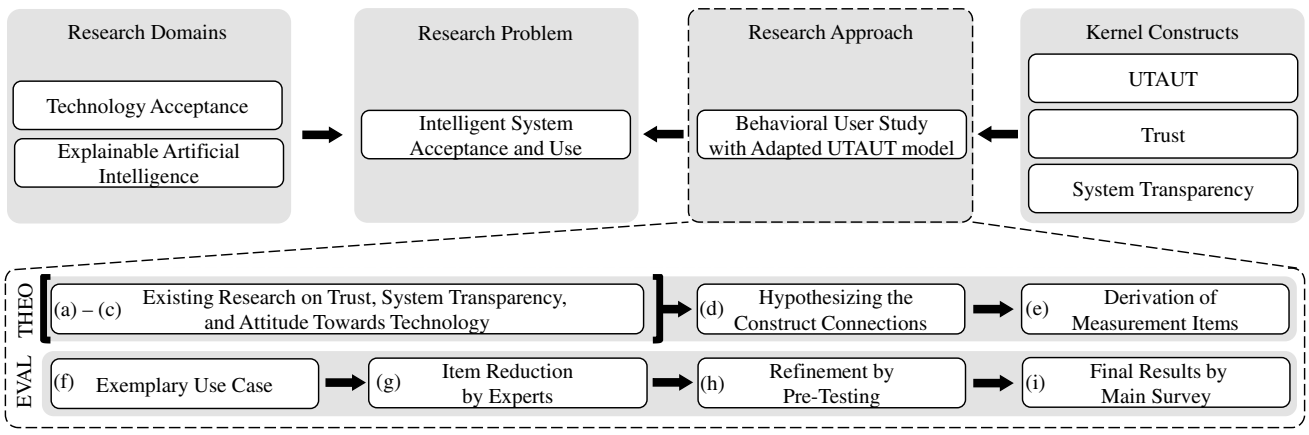


Fig. 1 Methodology overview

(2010), which we modified to suit our objective. We detail the steps in the respective sections.

The kernel constructs to form our model are derived from the related research on UTAUT, trust, and system explainability. Thus, in the theorizing section (THEO, see Section "Research theorizing"), we derive a suitable model from existing UTAUT research on (a-c) system transparency and attitude towards technology. We then (d) hypothesize the derived measurement model constructs and connections based on empirical findings, and we I collect potential measurement items.

In the evaluation section (EVAL, see Section "Study and results"), we (f) validate and modify our UTAUT model by using an exemplary application case in the field of industrial maintenance. Further, we (g, h) iteratively adapt it in empirical studies, perform the main study, and (i) discuss the results.

As scientific methods, we use empirical surveys (see e.g., Lamnek & Krell, 2010) in combination with a structural equations model (SEM) (see e.g., Weiber & Mühlhaus, 2014). For the analysis of the SEM, we apply the variance-based partial least squares (PLS) regression (see e.g., Chin & Newsted, 1999).

Research theorizing

Trust extensions in UTAUT

While trust has been widely recognized as an important factor in information system usage in ABM theory, the UTAUT model does not account for trust in its original form (Carter & Bélanger, 2005). While several extensions of the UTAUT model have been proposed to address this drawback, both the inclusion and definition vary among research contributions (Venkatesh et al., 2016). Table 1 depicts a summary of UTAUT extensions regarding the construct of trust.

We can characterize these extensions by inclusion type regarding the dependent variables, which are affected by the trust construct in the respective UTAUT model. Endogenous inclusion refers to a direct connection between trust and *BI*, while exogenous inclusion refers to an indirect relationship through other variables. Furthermore, we indicate which determinants are included for the trust variable itself.

In terms of trust-based model components, we found i) several theoretical approaches to describe trust itself; ii) multiple determinants of the embedded trust construct (determinants); iii) several different ways of embedding trust into existing technology acceptance models such as ABM or UTAUT (inclusion type/ dependent variable).

While a majority of contributions (e.g., Oh & Yoon, 2014) include trust as a single variable with no determinants in an endogenous manner, other studies (e.g., Cheng et al., 2008) adopted more complex theoretical frameworks. McKnight et al. (2002) present a frequently adopted framework. They define a model to determine the intention to trust a system by building upon trust perception theory by Mayer et al. (1995). Specifically, they determine the users' intention to trust as the willingness of the user to depend on the system. This intention is influenced by three variables: *disposition to trust/trust propensity, institution-based trust*, and *trusting beliefs*. Disposition to trust or trust propensity is the general tendency to trust others, in this case an intelligent system. Institution-based trust refers to the contextual propitiousness that supports trust, indicating an individual's belief in good structural conditions for the success of the system. Trusting beliefs indicate an individual's confidence in the system to fulfill the task as expected (Mayer et al., 1995; Vidotto et al., 2012).

Trusting beliefs itself is comprised of three determinants: *trust benevolence*, *trust integrity*, and *trust ability/ability beliefs*. *Ability beliefs* (AB) refers to the system's perceived competencies and knowledge base for solving a task, that is trust in the ability of the system. Trust integrity involves the

Table 1 Trust-based UTAUT extensions

Inclusion type	Dependent variables	Determinants	Example references
Endogenous	BI	None	Alaiad and Zhou (2013); Carter and Bélanger (2005); Oh and Yoon (2014)
		Personal propensity to trust	Oliveira et al. (2014)
		Trust integrity, Trust ability	Komiak and Benbasat (2006)
		Trust property, satisfaction	Kim (2014)
Exogenous	PE	Trust benevolence, trust integrity, trust ability	Cheng et al. (2008); Lee and Song (2013)
	Perceived usefulness	Perceived ease of use, consumer decision making	Xiao and Benbasat (2007)
	Perceived risk, perceived usefulness	System transparency, technical competence, situation management	Choi and Ji (2015)
Endogenous/ exogenous	Perceived risk, BI	None	Slade et al. (2015)
	PE, BI	Trust benevolence, trust integrity, trust ability	Cody-Allen and Kishore (2006)

user's perception that the system acts according to a set of rules that are acceptable to him or her. Trust benevolence is indicated to be the belief in the system to do good to the user beyond its own motivation (Cheng et al., 2008; Mayer et al., 1995; McKnight & Chervany, 2000).

Considering the problem of a complex, intelligent system that mimics human functions, we adopted the unified model of McKnight et al. (2002) but made several modifications. First, we adopt *trust propensity towards an AI system* (TP) as the indicator for an aggregation of prior beliefs that potentially allow a professional to become vulnerable to an intelligent system. Moreover, we include this measure as an important determinant as in comparison to other information systems, perception of intelligent systems is different since the belief is also formed by outside media and social influence in more drastic way (e.g., sentient AI) that can increase factors of fear and/or aversion leading to decreased trust. We adopt *AB* as the determinant for *TP*, since it is the component that directly measures trust in the system itself rather than environmental factors and personal factors that are covered by facilitating conditions and moderators of the core UTAUT model already. Following the discussion in the realm of algorithm aversion, we argue that the trust propensity will be changed by seeing the system perform. While this is contrary to related work on trust, we believe that based on findings rooted in the algorithm aversion theory that for intelligent systems, *TP* also encompasses the changeable beliefs regarding the ability of algorithms. As argued above trust propensity also reflects beliefs that reflect external sources like media. This renders its role more important than merely reflecting on a general trusting behavior but rather as an indicator of trusting an intelligent system specifically. Thus, we also used items that express a tendency for trust propensity that can be subject to change. Third, we model *TP* as a direct influence factor of *BI* and as an exogenous factor for *PE*. Following the discussion in Lankton et al. (2015) between human-like and system-like trust, we come to understand that with a system, *AB* reflects the system-like trust properties of reliability, functionality, and helpfulness, as a system is not able to exhibit behaviors on its own that would not adhere to the given rules (trust integrity) or exocentric motive (trust benevolence). Thus, in a bid to limit complexity, we omit the factors trust benevolence and trust integrity and adopt *AB* as the sole determinant for *TP* by assuming that no matter how many functions or tasks are assigned, the system has no hidden intention to extend its tasks beyond its programming and it cannot change its "promise" by itself. For a possible pre-existing perception of ill will towards the system, trust propensity will collect those prior beliefs reflected by the tendency of the user to trust the system prior to use. This is also confirmed by Jensen et al. (2018) who in the case of computer systems attribute the most influence on benevolent beliefs and perception to dispositional characteristics that are already reflected in our model by trust propensity. However, if we extend the definition of the system by including system providers, programmers, and other stakeholders that are involved in the creation and maintenance, this simplification will pose a limitation, since hidden, malicious behavior can then play a pivotal, especially with intelligent systems since they can be subject to manipulation for example via adversarial learning (Heinrich et al., 2020).

System transparency extensions in UTAUT

Especially in recent years, the transparency of a system, as the backbone of an XAI's system explainability, has been increasingly integrated into studies of technology acceptance of intelligent systems and seems to have a direct influence on the perceived trust of users (e.g., Nilashi et al., 2016; Peters et al., 2020). In the context of intelligence systems, we define *system transparency* (ST) as the ability of the system to explain and reveal its decision rationale to the user by visual means (e.g., a visual panel that shows based on which maintenance-related input variables the suggestion of imminent maintenance was made). Table 2 depicts a summary of UTAUT extensions regarding the construct of *ST*.

We can characterize these extensions regarding the dependent variables, which are affected by the *ST* construct in the respective UTAUT model. Again, we found only references for the exogenous/ endogenous inclusion type. Similarly, we indicate which determinants are included for the transparency variable itself.

Among others, Brunk et al. (2019) and Hebrado et al. (2013) define *ST* as a factor to increase the user's understanding of how a system works. It further entails an understanding of the system's inner working mechanisms. That is why specific recommendations were made according to different characteristics and assumptions for a single item (Nilashi et al., 2016; Peters et al., 2020) as well as the system's overall decision rationale. Furthermore, *ST* should be used for required justifications (Shahzad et al., 2020).

Nevertheless, the influence of other factors on *ST* differs in these models. While many contributions, such as Brunk et al. (2019) and Peters et al. (2020), take no further factors into account, Nilashi et al. (2016) consider the type of explanation and the kind of presented information. They measure the factor of explanation through the level of explainability according to the user's perception and, thus, how, and why a recommendation was made and the interaction level within the recommendation process. For Shahzad et al. (2020), it is about characteristics of the information quality, such as for example accuracy and completeness, which influence *ST*.

Further, we noticed *ST* influences many factors: *BI*, *PE, EE,* and trust. As argued above, the factor of trust is modeled as *TP* and *AB*. Here, it is assumed that a highly transparent

Table 2 System-transparency-based UTAUT extensions

Inclusion type	Dependent variables	Determinants	Example references
Exogenous/Endogenous	Trust, BI	None	Brunk et al. (2019); Slade et al. (2015); Choi and Ji (2015); Hebrado et al. (2011); Hebrado et al. (2013)
	Trust, BI	Explanation	Nilashi et al. (2016)
	Trust, BI	Accuracy, completeness	Peters et al. (2020)
	ATT, trust, BI	None	Shahzad et al. (2020)
	Understanding, BI, users' privacy concerns	None	Zhao et al. (2019)
	Trusting beliefs, understanding, competence, acceptance	None	Cramer et al. (2008)
	AB, information satisfaction	Accuracy, completeness, time information currency	Cody-Allen and Kishore (2006)
	EE, trusting beliefs	None	Wang and Benbasat (2016)

decision-making process results in an increasing *TP* (Shin, 2020b; Vorm & Combs, 2022), while also increasing transparency results in a better *AB* of the user (Cody-Allen & Kishore, 2006). It is important to know that our assumption reflects a time-dependent use behavior were an introduction of the system takes place and through experiencing performing and through explanation of the decision rationale, the prior trust behavior (TP) changes through a change in the beliefs of the system's ability (AB) after seeing it perform (Dietvorst et al., 2015). *BI* is defined as the degree to how a user's intention changes through the level of *ST* (Peters et al., 2020). Lastly, an increase in *ST* results in a clearer assessment by the user, and thus the user's mental model assumes a higher performance of the system leading to increased *PE*. Likewise, a transparent system can reduce a user's efforts to understand the systems' inner working mechanisms (Wang & Benbasat, 2016).

Attitude towards technology extension in UTAUT

Consistent with the theory of planned behavior, an individual's attitude towards technology, in this case *attitude towards AI technology* (ATT), has been found to act as a mediating construct (Dwivedi et al., 2019; Kim et al., 2009; Yang & Yoo, 2004). People are said to be more likely to accept technology when they can form a positive attitude towards it. It is important to note that usually the construct is placed between the endogenous variables in the UTAUT context (e.g., *PE* and *EE*) and intention to use (e.g., *BI*). Furthermore, we believe that *ATT* is influenced the individual's pre-formed opinion about AI technology. The prevailing opinion that forms into attitude is not changed easily and depends on an individual's prior exposure to the technology (Ambady & Rosenthal, 1992). Factors accumulated in *ATT* can be religious beliefs, job security, attitude carried over from popular culture, as well as knowledge and familiarity and privacy, and relational closeness (Persson et al., 2021). Thus, it acts as a place to collect emotional attitude towards a technology, which in the case of AI is reinforced by its anthropomorphic and intransparent nature. While some studies show that not all of these factors are present in an individual's mind, general states of mind like fear towards the technology can influence and form the person's attitude (Dos Santos et al., 2019; Kim, 2019). Thus, we argue to include *ATT* and hypothesize that the mediation strength and thus indirect connections to *ATT* are increasingly present for intelligent systems.

Model and hypotheses

As a result of the above construct derivation, we present our UTAUT model for intelligent systems along with the hypotheses and their respective direction (− or +) in Fig. 2. The measurement model can be divided into three major parts: i) UTAUT core (*PE*, *EE*, and *BI*), ii) UTAUT AI (*AB*, *TP*, *ST,* and *ATT*), and iii) moderators (*gender*, *age*, *experience*).

The derivation of the hypotheses from i) UTAUT core research is primarily based on general research on UTAUT (e.g., Dwivedi et al., 2019; Venkatesh et al., 2003). Nevertheless, these construct interrelations can also be found in UTAUT studies on trust or system transparency (e.g., Lee & Song, 2013; Wang & Benbasat, 2016). Compared to the UTAUT core established by Venkatesh et al. (2003), the constructs of facilitating conditions and social influence are not included due to the results of our multistage reduction process (cf. Section "Study design").

The construct *BI* represents our target variable. It measures the strength of a user's intention to perform a specific behavior (Fishbein & Ajzen, 1977). Here, it is about the willingness of a user to adopt an intelligent system or more

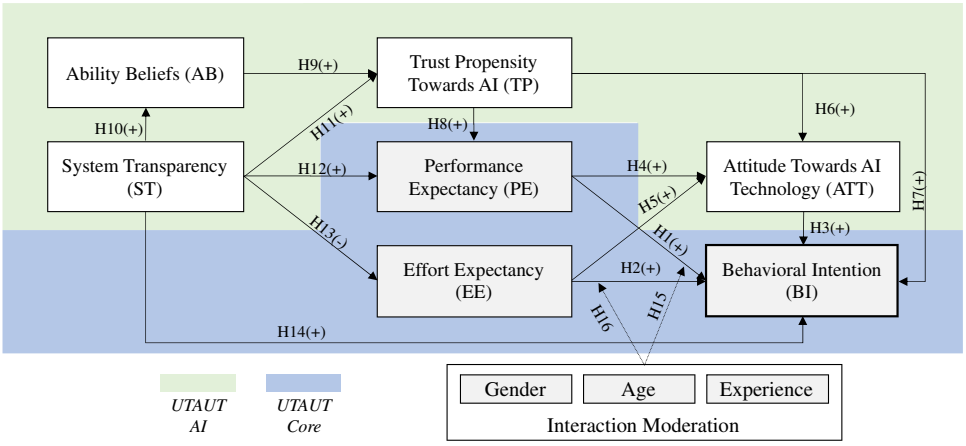


Fig. 2 Derived acceptance model for intelligent systems

specifically the willingness of the user to take advice recommended by the intelligent system. This is an important distinction as intelligent system use can be mandated in a professional setting. For example, in the case of intelligent systems for decision support, technically it is not about the intention of the user to adopt the system as such, but about the intention of the user to considers the system's output in his or her work processes.

The construct is influenced endogenously by the two basic UTAUT constructs of *PE* and *EE*. *PE* measures the degree to which an individual believes that using the system can increase their job performance. This includes factors such as perceived usefulness, job-fit, relative advantage, extrinsic motivation, and outcome expectation (Venkatesh et al., 2003). Whereas *EE* measures the degree of individual ease associated with the use of the system, including factors such as perceived ease of use and complexity (Venkatesh et al., 2003). For both constructs, we assume that they have a positive influence on *BI*. This correlation can be seen in UTAUT (e.g., Dwivedi et al., 2019; Venkatesh et al., 2003) as well as in UTAUT model studies on trust and on *ST* (Cheng et al., 2008; Cody-Allen & Kishore, 2006; Lee & Song, 2013; Wang & Benbasat, 2016). Thus, we state:

H1: *Performance expectancy positively affects behavioral intention.*
H2: *Effort expectancy positively affects behavioral intention.*

The hypotheses of ii) UTAUT AI, and thus, for *ATT*, *AB*, *TP*, and *ST* are primarily based on the references from Tables 1 and 2 as well as Venkatesh et al. (2003)'s considerations.

ATT is defined as a user's overall affective reaction to using AI technology or an AI system (Venkatesh et al., 2003). While the authors did not include the construct in his final model, it is regularly used in the context of decision support systems. UTAUT research, as well as ABM research on trust, indicate that *ATT* has a positive effect on *BI* (e.g., Chen, 2013; Hwang et al., 2016; Mansouri et al., 2011). That is, people form intentions to engage in behaviors to which they have a positive attitude (Dwivedi et al., 2019). Inversely, it is assumed that both *PE* and *EE* have a positive influence on a user's *ATT*. Suleman et al. (2019) derive this significantly positive influence from the ABM research (Hsu et al., 2013; Indarsin & Ali, 2017) and later confirm it in their own research. Dwivedi et al. (2019) and Thomas et al. (2013) confirm the connection. We summarize these findings by our next hypotheses:

H3: *Attitude towards AI technology positively affects behavioral intention.*
H4: *Performance expectancy positively affects attitude towards AI technology.*
H5: *Effort expectancy positively affects attitude towards AI technology.*

Trust is regarded as a necessary prerequisite to forming an effective intelligent information system (e.g., Dam et al., 2018) and, thus, it is a crucial construct to build our model. In Section "Trust extensions in UTAUT", we have explained that *TP* is influenced by AB and thus can be changed by observing the system behavior. *AB* measures the assumed technical competencies of the system to solve a task (Schoorman et al., 2007). *TP* is about the user's general disposition to trust an intelligent system with a task. For *TP*, we expect a positive effect on *ATT* and on *BI*, as this preformed trust is a major influence that has formed from user experienced and external exposure (e.g. through media) which is increasingly critical for intelligent system (Gherheş, 2018). For Suleman et al. (2019), trust in general was the most influential and significant factor affecting a participant's *ATT*. The positive influence of trust on *BI* is well proven by several UTAUT studies (Choi & Ji, 2015; Lee & Song, 2013). In addition, we argue that specifically for intelligent systems, while the assumed ability of the system is quite high, the black-box nature and skepticism or the presence of algorithm aversion

in humans can make it increasingly harder to form trust towards a system. As opposed to traditional software systems, intelligent systems make decisions based on a learning process and do not necessarily follow the same reasoning as the human decision-making process. Thus, building trust towards an intelligent system seems more important since the natural state usually assumes a rather critical view of aversion (Mahmud et al., 2022). We summarize this with the next hypotheses:

H6: *Trust propensity towards AI positively affects attitude towards AI technology.*
H7: *Trust propensity towards AI positively affects behavioral intention.*

In turn, we assume that *AB* has a positive influence on a user's *TP*, which in turn has a positive influence on *PE*. However, the direction of the latter influence is disputed in prior research. While Oliveira et al. (2014), Nilashi et al. (2016), and Wang and Benbasat (2016) assume that *PE* has a positive influence on trust, Cody-Allen and Kishore (2006), Lee and Song (2013), and Choi and Ji (2015) think that trust affects *PE*. We additionally argue that the propensity to trust an intelligent system will also result in increased expectance of future performances, while distrust in a system will also lower the expectations towards future high performance. As mentioned earlier with intelligent systems there is a general aversion on the one side, while there is also evidence of performance that exceeds human decision makes. Although intelligent systems can outperform humans, humans are sometimes preferred despite performing worse because of trust issues (Dietvorst et al., 2015). This is partially reflected by a person's trust propensity. Therefore, we assume a positive influence of *TP* on *PE*. The influence of *AB* on *TP* is also relying on the fact that humans change behavior towards algorithms once they observe their behavior, which even can result in switching from a state of aversion to a state of algorithm appreciation (Logg et al., 2019). In fact, we argue that the assumed effect is even stronger with the performance promise that is attributed to intelligent systems compared to a traditional software system. Accordingly, we formulate our hypotheses:

H8: *Trust propensity towards AI positively affects performance expectancy.*
H9: *Ability beliefs positively affect trust propensity towards AI.*

Trust – as a multifaced term – is assumed to have a strong correlation with *ST* (e.g., Dam et al., 2018). *ST* measures the user's understanding of the intelligent system's decision rationale (Hebrado et al., 2011). In other words, it represents how openly an intelligent system's inner decision rationale is working as well as how openly characteristics that determine why an intelligent system made a certain decision hare communicated (Mohseni et al., 2021). As the user of such an intelligent system decides whether or not to adopt the system recommendation, *ST* might influence his or her decision-making process. We expect a positive effect of *ST* on *AB* and *TP* based on the findings of Pu and Chen (2007) and Wang and Benbasat (2016) that rely on trust in general. The former found that users assign a recommender system a higher level of competence if the decision-making process is explained in a traceable manner. The latter is supported by the preliminary UTAUT research of Brunk et al. (2019), Hebrado et al. (2013), Nilashi et al. (2016), Peters et al. (2020), and Chen and Sundar (2018). For example, Peters et al. (2020) found that *ST* positively influenced trust in the intelligent system significantly in the context of testing a consumer's willingness to pay for transparency of such black-box systems. We further argue that the ability to experience the reasoning of the system as a form of self-disclosure will be accepted as some kind of honesty. In addition to the related work that deals with trust as a general construct, we argue that based on an existing state of aversion towards intelligent systems before seeing them perform and experiencing their decision rationale, a method to create transparency not only with regard to the performance but also with regard to the inner logic on a case-to-case basis we increase likelihood of mitigating the aversion (Herm et al., 2022; Mohseni et al., 2021). While a prior trust (measured through trust propensity) has already been made up, we argue that through the usually missing experience with the design of intelligent systems, this formed opinion can be changed through visual means and demonstrations such as how-to or why-explanations as presented in the XAI literature (Arrieta et al., 2020; Herm et al., 2022). We conclude with the hypotheses:

H10: *System transparency positively affects ability beliefs.*
H11: *System transparency positively affects trust propensity towards AI.*

Technology acceptance research supposes that *ST* also influences the residual UTAUT constructs of *PE*, *EE*, and *BI*. We derive the assumed positive effect of *ST* on *PE* from Zhao et al. (2019), who revealed that a higher level of a decision support system supports the user's perception of the performance of that system. If users understand how a system works and how calculations are performed, they will perceive that, in some cases, implementing and using the system requires more effort (Gretzel & Fesenmaier, 2006). This is a special case with black-box intelligent systems since not everything is transparent out-of-the-box and, thus, an associated effort cannot always be clearly derived. However, through *ST* the effort can be monitored and revealed.

Thus, we expect a positive influence of *ST* onto *EE*. We also expect a positive influence of *ST* for *BI*. Making the reasoning behind a recommendation transparent allows for an understanding of the recommendation process, significantly increasing acceptance (Bilgic & Mooney, 2005). This significant and strong influence is also reflected in further studies by Venkatesh et al. (2016) and Hebrado et al. (2011). Furthermore, the basic concept of seeing an algorithm perform well can, for some tasks, increase performance expectancy. As a prerequisite of making up one's mind about an algorithm, the ability to experience it is a fundamental necessity (Dietvorst et al., 2015; Logg et al., 2019). We address this through three hypotheses:

H12: *System transparency positively affects performance expectancy.*
H13: *System transparency positively affects effort expectancy.*
H14: *System transparency positively affects behavioral intention.*

The hypothesis for the iii) moderators is also part of the original UTAUT model according to Venkatesh et al. (2003). We assume that *gender*, *age*, and *experience* have a moderating effect on *PE*, *EE*, and *BI constructs*. We derive this assumption from the initial UTAUT model (Venkatesh et al., 2003). It has been confirmed in several other UTAUT studies (e.g., Alharbi, 2014; Esfandiari & Sokhanvar, 2016; Wang & Benbasat, 2016). In contrast, we do not consider voluntariness of use due to the obligatory use of intelligent systems in day-to-day business. From this, we derive the following hypotheses:

H15: *Gender, age, and experience moderate the effects of performance expectancy on behavioral intention.*
H16: *Gender, age, and experience moderate the effects of effort expectancy on behavioral intention.*

Study and results

Study use case

In the following, we offer an exploration of the theoretical constructs put forward. For this purpose, we defined a real-world use case and transferred it to the UTAUT model in a step-by-step procedure. In this way, we validate the applicability of our proposed model. Moreover, we gain first insights into the user's willingness to accept intelligent systems at their workplace.

We consider industrial machine maintenance to be a suitable scenario. Its focus is to maintain and restore the operational readiness of machinery to keep opportunity costs as low as possible. In contrast to reactive strategies, anomalous machine behavior can be identified and graded early on using statistical techniques to avoid unnecessary work. Given the technological possibilities to collect large and multifaceted data assets in a simplified manner, intelligent systems based on machine learning are a promising alternative for maintenance decision support (Carvalho et al., 2019).

In this context, rolling bearings are used in many production scenarios of different manufacturers. For example, they are often installed in conveyor belts for transport or within different engines and show signs of wear and tear over time that requires maintenance (Pawellek, 2016).

For our evaluation, we decided to use an automated production process to manufacture window and door handles, as these are common everyday items every respondent can relate to. In our scenario, there shall be several production sections connected by high-speed conveyor belts. Inside these conveyor belts, several bearings are installed. These are monitored by sensors to monitor change (e.g., noise sensor, vibration, and temperature). A newly introduced intelligent system evaluates this data automatically. In case of anomalous data patterns, a dashboard displays warnings and errors with concrete recommendations for action (cf. Appendix A).

The respondents of the survey(s) shall be confronted with a decision situation that tests whether or not the user adopts the system recommendation in his or her own decision-making process. That is, they need to decide for or against an active intervention in the production process as recommended by the system. In an extreme case, the optical condition of the conveyor belt bearings is perceived as good. However, the system recommends that the conveyor belt must be switched off immediately. This error does not occur regularly, and the message contradicts the previous experience of the service employee (here, the respondent) with this production section. As additional information, we provide the reliability of the system recommendations and hint at the high follow-up costs in case of a wrong decision.

Study design

Our design and conduct of the survey are based on Šumak et al. (2010), which we modified to our objective. We used five steps to obtain our study results: i) collection of established measurement items; ii) pre-selection by author team; iii) reduction by experts; iv) evaluation and refinement through pre-study; and v) execution of the main study. See Appendix B for the results of each step, as well as the primary and secondary source(s) for all measurement items. A more detailed result table of the validity and reliability

measures of the pre-study and main study is available in Appendix C.

Step i). First, we collected those measurement items that already exist for the respective constructs of interest and are, thus, empirically proven.

As we adopted *PE*, *EE*, and *BI* from Venkatesh et al. (2003), we built on their findings. Venkatesh et al. (2003) chose the measurement items for UTAUT by conducting a study and testing the measurement items for consistency and reliability. For the additional constructs *ATT*, *ST*, *AB*, and *TP*, we examined the source construct measurement items as well as examples of secondary literature and derived constructs. Initially, we used three items to form the construct of *ATT* – one was adopted from Davis et al. (1992) and two from Higgins and his co-authors (Compeau et al., 1999; Thompson et al., 1991). As we derived *ST* from the perceived local explainability of an intelligent system decision's result visualization as well as the perceived global explainability of the intelligent system's decision process, we initially included five items from Madsen and Gregor (2000) to address the global component and two items from Cramer et al. (2008) to address the local component, as noted in recent XAI-related research (Adadi & Berrada, 2018; Mohseni et al., 2021). The measurement items for *AB* and *TP* were derived from McKnight et al. (2002) (trust competence) and Lee and Turban (2001) (trust propensity). Lastly, the measurement items for facilitating conditions and social influence were adapted from Taylor and Todd (1995), Thompson et al. (1991), Moore and Benbasat (1991), and Davis (1989).

Step ii). Next, we discussed the appropriateness of each of the collected measurement items within the team of authors.

The team members merge knowledge in the respective domains of industrial maintenance, technology acceptance, and (X)AI research. Special attention was paid to the duplication of potential item questions and their feasibility for the use case. We reduced the total number of measurement items for the model's constructs from 71 to 24.

Step iii). Subsequently, we conducted an expert survey with practitioners from industrial maintenance regarding our intended main study.

The survey with ten experts had two goals: reducing the remaining measurement items and understanding the explainability of intelligent system dashboards. For the former, we briefly explained each of the model measurement constructs to the experts. Thereby, we removed the constructs of facilitating conditions and social influence completely. Subsequently, the experts selected the most appropriate remaining measurement items for the use case per measurement construct. They were given at least one vote and at most votes for half the items. Then, we selected the final measurement items based on a majority vote. For the latter, we presented the experts with four different maintenance dashboards of intelligent systems as snapshots adapted from typical software in the respective field (e.g., Aboulian et al., 2018; Moyne et al., 2013). Here, the experts rated their perceived level of explanation goodness on a seven-point Likert scale. Using the dashboard with the highest overall (median) explanation goodness, ensures that the dashboard for the quantitative survey has inherent explainability to the end-user and thus provides enhanced system transparency (cf Appendix A).

Step iv). Then, we conducted a quantitative pilot study to critically examine our questionnaire and research design (Brown et al., 2010). The testing includes checks for internal consistency, convergent reliability, indicator reliability, and discriminant validity.

The study contained 60 valid responses. Here, we ensured representative respondents, that is, maintenance professionals holding a position to use an intelligent system for their job-related tasks (e.g., experience in maintenance). See Appendix E for the demographics of the responses. We provided the participants with a description of the exemplary use case and screenshots of the prototype. We asked them to respond to their perceptions of each of the measurement items on a seven-point Likert scale. See Table 3 for the assessment of measurement items and Appendix D for a summary of our decisions on individual items.

Step v). Then, we conducted our main quantitative study. Table 4 comprises the final set of measurement items. We, again, checked for internal consistency, convergent reliability, indicator reliability, and discriminant validity. Further, while we did not include explicit control variable, we included control questions (CQ) following Meade and Craig (2012) and Oppenheimer et al. (2009) to increase result validity.

We acquired a total of 240 participants who completed the questionnaire via the academic survey platform Prolific. Out of this sample, 240 respondents answered CQ1 correctly. Twenty-three respondents failed CQ2. For CQ3 and CQ4, we decided to add a tolerance of ± 1 point. The scale for CQ3 was inverted, and answers compared to PE4, while answers for CQ4 were compared to TP1. The final

Table 3 Validation and reliability testing of pre-study

Construct	Assessment measurement items					
	CA	AVE	CR	FL criterion	Cross-loadings	Item loadings
PE	0.86	0.64	0.90	–	–	–
EE	0.74	0.57	0.84	–	–	EE1
ST	**0.67**	0.75	0.86	–	–	–
TP	**0.23**	0.56	**0.67**	–	TP4 (*BI*)	TP3, 4
AB	0.74	0.66	0.85	–	–	–
ATT	**0.67**	0.59	0.81	–	–	ATT2
BI	0.83	0.74	0.89	–	–	–

Internal consistency: Cronbach's alpha (CA) > 0.7; composite reliability (CR) > 0.7 (Gefen et al., 2000; Hair et al., 2011)

Convergence reliability: average variance extracted (AVE) > 0.5 (Hair et al., 2011)

Indicator reliability: item loadings $0.7 < x < 1$ (Hair et al., 2011)

Discriminant validity: cross-loadings; Fornell-Larcker (FL) criterion (Fornell & Larcker, 1981; Hair et al., 2011)

dataset consists of 160 samples. See the following Table for the demographics of the sample. The demographical data (*gender*, *age*, *experience*) was used for the interaction moderation of the UTAUT model's results presented in Section "Study results".

All constructs achieved reliability and validity across all measurements. Values for Cronbach's alpha, average variance extracted, and composite reliability are well above their respective thresholds. Item ATT2 was below this limit for item loadings ($0.59 < 0.7$) and was thus excluded from the measurement model, resulting in an overall good reliability of *ATT*. We did not observe any cross-loadings, and none of the constructs failed the Fornell-Larcker criterion (cf. Table 5). We additionally checked for collinearity-based indicators of common method bias following the suggestions of Kock (2015). We thus compared the variance inflation factor with the proposed threshold and found that no independent variable exhibits the variance inflation factor threshold of 3.30 and thus no common method bias was detected. See Appendix F for null validation and reliability testing results and Appendix G for the inner and outer variance inflation factor values. Lastly, see Appendix H for the median and standard deviation of the conducted measurement items.

Study results

In the following, we present the results from the main study. The estimated model with *direct effect* estimates is depicted in the upper part of Table 6, while the lower part contains the observed *indirect effects*.

In addition, we conducted a mediation analysis based on the indirect and direct effects in our SEM to further investigate the role of system transparency and the two trust constructs following the methodology described in Zhao et al. (2010) and Hair Jr et al. (2021). The type of the mediation effect was derived by comparing direct and indirect effects of the constructs and is subsequently given in Table 7. The effects are determined according to the common decision scheme of mediation roles in SEM that was suggested by Hair Jr et al. (2021).

The role of UTAUT core constructs

First, we examine the role of the initial exogenous UTAUT constructs *PE* and *EE*. In accordance with Venkatesh et al. (2016) and Dwivedi et al. (2019), *PE* is connected significantly to *BI*. While we observe this effect of *PE* with magnitude 0.313, we cannot confirm a significant effect of *EE* on *BI*. Thus, we can confirm H1 but reject H2. However, we can confirm a significant effect from *ATT* to *BI* in its exogenous role with an effect strength of 0.348. With the established confirmation of H3, we can observe a significant effect of magnitude 0.162 from *EE* to the construct *ATT* in its endogenous role, resulting in an indirect relationship to *BI*. Likewise, with a comparably more substantial effect than its direct connection (0.480), *PE* affects *ATT* significantly. We can therefore confirm H4 and H5, respectively. Comparing the bias-corrected confidence intervals of *EE* to *BI* (width 0.338, from −0.218 to 0.120) and *EE* to *ATT* (width 0.25, from 0.027 to 0.278) further strengthen the notion that *EE* affects *BI* rather indirectly through *ATT* in our context of intelligent systems, confirming results by Dwivedi et al. (2019) and Thomas et al. (2013).

The role of the user's attitude towards intelligent systems

Since *ATT* is defined as an affective reaction, we conclude that this construct has increased presence in the case of intelligent systems, resulting in its role as a transitory connection

Table 4 Final set of measurement items for main study

Construct		Measurement item	Reference(s)
PE	PE1	Using this system in my job would enable me to accomplish tasks more quickly.	Davis (1989)
	PE2	Using this system would improve my job performance.	
	PE3	Using this system would make it easier to do my job.	
	PE4	I would find this system useful in my job.	
	PE5	Using this system would increase my productivity.	Moore and Benbasat (1991)
EE	EE1	Learning to operate this system would be easy for me.	Davis (1989)
	EE2	I would find it easy to get this system to do what I want it to do.	
	EE3	My interactions with this system would be clear and understandable.	
	EE4	I would find this system easy to use.	
ATT	ATT1	The actual process of using this system would be pleasant.	Davis et al. (1992)
	ATT2	This system would make work more interesting.	Thompson et al. (1991)
	ATT3	I would like to work with this system.	Compeau et al. (1999)
	ATT4	Using the system would be a bad/good idea.	Peters et al. (2020); Taylor and Todd (1995)
	ATT5	Using the system would be foolish/wise move.	
BI	BI1	If this system was available to me, I would intend to use this system in the future.	Venkatesh et al. (2003)
	BI2	If this system was available to me, I predict I would use this system in the future.	
	BI3	If this system was available to me, I would plan to use this system in the future.	
ST	ST1	I would understand how this system will assist me with decisions I have to make.	Madsen and Gregor (2000)
	ST2	I would understand why this system provided the decision it did.	Cramer et al. (2008)
	ST3	I would understand what this system bases its provided decision on.	
AB	AB1	This system would be competent in providing maintenance decision support.	Cheng et al. (2008); McKnight et al. (2002)
	AB2	This system would perform maintenance decision support very well.	
	AB3	In general, this system would be proficient in providing maintenance decision support.	
TP	TP1	It would be easy for me to trust this system.	Cheng et al. (2008); Lee and Turban (2001); Wang and Benbasat (2007)
	TP2	My tendency to trust this system would be high.	
	TP3	I would tend to trust this system, even though I have little or no knowledge of it.	
CQ	CQ1	I would not find this system easy to use.	–
	CQ2	Although I may would not know exactly how this system works, I would know how to use it to make decision regarding the quality of its output. Please do not rate this statement and please choose scale point one instead to ensure the data quality of this survey. This only applies to this question.	Meade and Craig (2012)
	CQ3	I have read all questions carefully and answered truthfully.	
	CQ4	Thank you for taking the time to participate in this survey. We end the survey by capturing data about the demographics of the participants. As such, data about gender, age, and experience in the topic of the survey is being collected. In addition, we want to make sure the collected data is reliable. Please select the option "No answer" for the next question that asks about the length of the survey and simply write "I've read the instructions" in the box labeled "Additional remarks".	Oppenheimer et al. (2009)

of *EE* and *PE* to *BI*. It is reasonable to assume that a user is less affectionate about AI technology when it seems to be complicated to use. However, since intelligent systems are attributed with black-box properties, the ease of use can be difficult to determine beforehand. Hence, a direct connection between *EE* and *BI* seems less likely in the case of intelligent systems. The strong indirect relationship of *PE* to *BI* via *ATT* can be strengthened further by the notion of algorithm appreciation. Logg et al. (2019) found that an algorithmic system that is perceived as complex is expected to have high performance, preferable to that of humans. Thus, the increased *PE* will positively influence their *ATT* before an intention to use is formed. Contrary, the notion of algorithm aversion, as expressed by Castelo et al. (2019) can cause the *PE* to drop if it is observed or expected that the system errs, resulting in a transitory decrease of positive attitude towards the system and making it less likely for the intelligent system to be used. This can be explained by the feeling of missing control over the (partially) autonomous intelligent system (Dietvorst et al., 2015).

Table 5 Demographics of main study sample

Characteristic	Attribute	Value		Characteristic	Attribute	Value	
		Freq.	Percent.			Freq.	Percent.
Gender	Male	110	68.75	Experience with intelligent systems (EXP2)	None	87	54.38
	Female	50	31.25		<1 year	22	13.75
	Others	0	0.00		1–3 years	24	15.00
Age	<=20	2	1.25		3–5 years	12	7.50
	21–30	62	38.75		5–10 years	10	6.25
	31–40	52	32.50		>10 years	5	5.00
	41–50	31	19.38	Experience with AI (EXP3)	None	47	29.38
	51–60	12	0.75		<1 year	39	24.38
	>61	1	0.06		1–3 years	43	26.88
Experience in industrial maintenance (EXP1)	None	59	36.88		3–5 years	16	0.10
	<1 year	31	19.38		5–10 years	11	6.88
	1–3 years	30	18.75		>10 years	4	2.50
	3–5 years	17	10.06				
	5–10 years	15	9.38				
	>10 years	8	5.00				

Note: Gender, Age, EXP1, EXP2, and EXP3 were used as interaction moderation

Table 6 Validation and reliability testing of main study

Assessment measurement items							
Construct	CA	AVE	CR	Constr.	CA	AVE	CR
PE	0.91	0.74	0.93	AB	0.89	0.82	0.93
EE	0.85	0.70	0.90	ATT	0.86	0.71	0.91
ST	0.87	0.80	0.92	BI	0.95	0.90	0.97
TP	0.85	0.77	0.91				

Note: No FL criterion, cross-loadings, and item loadings measured

Internal consistency: Cronbach's alpha (CA) > 0.7; composite reliability (CR) > 0.7 (Gefen et al., 2000; Hair et al., 2011)

Convergence reliability: average variance extracted (AVE) > 0.5 (Hair et al., 2011)

Indicator reliability: item loadings $0.7 < x < 1$ (Hair et al., 2011)

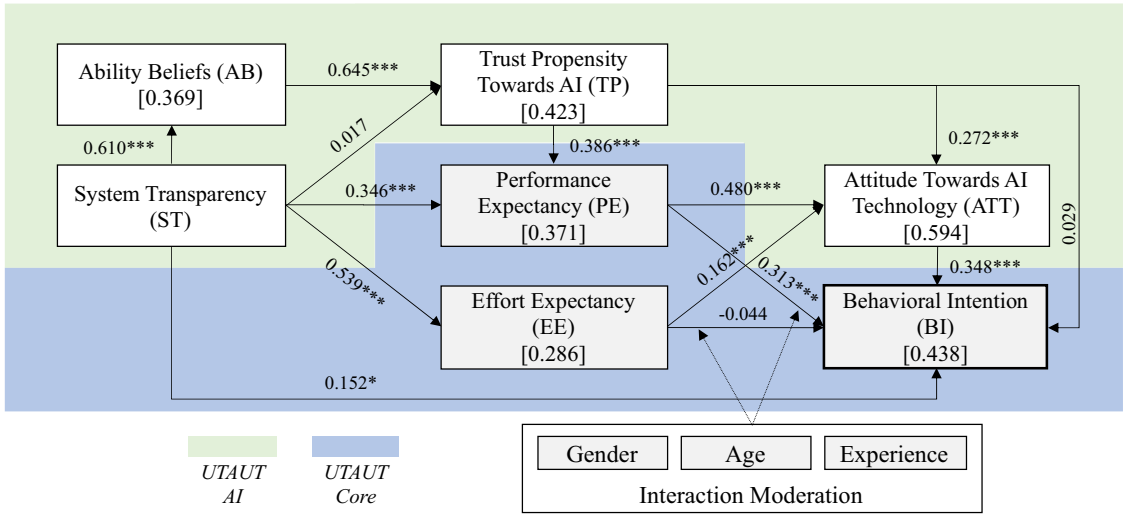
Discriminant validity: cross-loadings; Fornell-Larcker (FL) criterion (Fornell & Larcker, 1981; Hair et al., 2011)

The role of trust towards intelligent systems

Further, we examine the role of the trust-related exogenous constructs. We observe a significant effect from *TP* on *ATT* with a strength of 0.272. Again, we assume an indirect relation to *BI* through *ATT*, since the direct effect of *TP* on *BI* is not significant. This confirms that, especially in the context of intelligent systems, an a priori formed trust influences the affection towards technology and, in a transitory fashion, the intention to use said technology. This is further supported by the mediation analysis that found a purely mediating role for TP with regard to BI. Furthermore, we argue that there is a certain order that is important towards forming a decision. While trust is an important catalyst and mediator, it is not the sole determinant and it seems, from the results, inferior to actual performance. Similar findings are confirmed by Wanner et al. (2020) where in a choice experiment, the performance of an intelligent system played the most pivotal part. For some tasks it has also been found that trust is not a necessary condition for actual use (Logg et al., 2019). Especially for less critical scenarios such as maintenance, one can imagine that while important, pure performance can override trust. However, for tasks that have a more ethical and/or critical nature like healthcare, this might be different. Regarding H6, *TP* and *ATT* are highly affection-based constructs, and thus a connection between them seems highly appropriate. We can therefore confirm H6 and reject H7. We also found that *TP* has an effect on *PE* with a magnitude of 0.345. The confidence interval (width 0.269, from 0.239 to 0.508) confirms a strong effect along with hypothesis H8. The observations are in accordance with the findings of Cody-Allen and Kishore (2006), Lee and Song (2013),

Table 7 Results of main study

UTAUT results



Indirect effects

Construct	Indirectly effecting	Strength	*p*-Value	Mediation
EE	BI	0.056	0.059	Indirect-only (full mediation)
PE	BI	0.167	0.003	Complementary (partial mediation)
ST	ATT	0.441	0.000	Indirect-only (full mediation)
ST	BI	0.278	0.000	Complementary (partial mediation)
ST	PE	0.159	0.000	Indirect-only (full mediation)
ST	TP	0.394	0.000	Indirect-only (full mediation)
AB	ATT	0.295	0.000	Indirect-only (full mediation)
AB	BI	0.198	0.000	Indirect-only (full mediation)
AB	PE	0.249	0.000	Indirect-only (full mediation)
TP	ATT	0.185	0.000	Complementary (partial mediation)
TP	BI	0.268	0.000	Indirect-only (full mediation)

Confidence levels: * 0.10, ** 0.05, *** <0.001

Choi and Ji (2015) and thus confirm H8. Drawing from the findings of Logg et al. (2019), we can explain the increase in trust through algorithm appreciation that occurs with increasing algorithm performance. Thus, if a user experiences a well-performing intelligent system, the user is more likely to subsequently change the initial propensity to trust with regard to the system. To no surprise, also the user's trust in the algorithm's ability to perform well, *AB*, has a very strong effect on *TP* with a magnitude of 0.645, confirming H9. This is also expressed by the mediation analysis that shows no sign of a possible omitted mediator and identifying the role as full mediation.

The role of transparency of intelligent systems

Finally, we investigate the role of system transparency. Regarding the trust constructs *AB* and *TP*, we can confirm H10 since we observe a very strong effect of *ST* on *AB* with a magnitude of 0.610. However, we cannot confirm a significant direct connection from *ST* to *TP* and, thus, reject H11. This is not surprising since we expect the user to partially form a pre-existing opinion within trust propensity based on the pre-existing trust in the system's ability that can be better assessed when the user has access to an explanation of the system or the underlying algorithms. This is also supported by the full mediation role of *TP* that was added for this purpose and worked as expected. Furthermore, having a competitive or complementary mediation, while befitting our proposed hypothesis could imply the presence of other trust-related, yet unexplored mediation constructs as suggested in Zhao et al. (2010).

Regarding the initial UTAUT indicators, we find that an understanding of the system also affects the expected performance, as we observe a strong effect of 0.346 of *ST* on *PE*, confirming H12. The effect can be explained by the influences of explanations on perceived performance and decision towards an intelligent system as described by Wanner et al. (2020) through the means of local and global explanations. Through a global explanation of the intelligent system,

the user is made aware of its complexity, leading to increased performance expectancy because intelligent systems based on deep learning models are expected to outperform other systems. Likewise, local explanations that explain a single prediction enable a consensus between the mental model of the user and the system resulting in increased *PE*. An even more potent effect of *ST* was observed regarding *EE* with magnitude 0.539 and confirming H13. Revealing the system's complexity through global explanations also enables the user to realize the effort required to implement an intelligent system, thus increasing *EE*. Besides, we observe a direct effect of *ST* on *BI*, confirming H14 at the 0.10 significance level with a magnitude of 0.152. These results are in line with Wanner et al. (2020), who indicate that explainability plays a key role when deciding on an intelligent system. The direct effect is rather low compared with the indirect effect via *PE*, which is also in accordance with their findings, where explainability was not as strong a decision factor as performance. While the complementary mediation effect of *ST* regarding *BI* could indicate omitted mediators, we rather suspect the variety of functions of explanations in intelligence systems pose different influences that affect the behavioral intention in an either indirect or direct way. One can imagine the sheer presence of a self-disclosing explanation of how the system forms a decision will positively influence the *BI*, in addition with positively influencing trust in the system. In summary, we find that *ST* poses a strong influential factor concerning the attitude and intention to use an intelligent system either indirectly through previously introduced constructs or as a minor direct effect.

The role of user characteristics

Lastly, we look at the moderating effects of *age*, *gender,* and *experience* on *PE* and *EE*. We found no significant moderating effects of either variable or construct, contradicting the findings of Alharbi (2014) and Esfandiari and Sokhanvar (2016). We assume that this is because our prescreening sets boundary conditions that do not allow for a great deal of variance within the participants. Thus, we observed mostly minor experiences and age gaps. In addition, due to the application domain, the sample was skewed towards men (68.75%), barely allowing for reliable variation.

Discussion

Theoretical implications

Performance is crucial (when looking at direct effects)

We extended the modified UTAUT model by Dwivedi et al. (2019), which itself is based on the UTAUT model of Venkatesh et al. (2003), and derived additional constructs and connections in the context of intelligent systems acceptance and use. The direct and indirect effects of *PE* play a major role and are comparable to the findings of Dwivedi et al. (2019). The findings of Wanner et al. (2020) confirm the dominating role of the expected performance. Contrary, we found that the expected effort is not of major concern when looking at the direct effects since it only delivers impact via indirect connections. We consider this as a first indication of the increased difficulty to build a direct intention to use in the case of intelligent systems, since the intention relies on the affection towards the system more heavily as expressed by the extended UTAUT model of Dwivedi et al. (2019). Thus, while performance is king, it is insufficient to focus only on direct effects when evaluating an intelligent system's acceptance.

Human attitude and trust steer acceptance as latent indirect factors

As mentioned previously, the strength of indirect effects delivered through the more affectionate construct of *ATT* is substantial and shows the necessity for recognizing the deviation from a purely performance- and effort-centered model. Following that thought of increased affection constructs, we found that initial *TP* plays an essential role in determining the *PE* regarding the system. Thus, we revealed a significant indirect influence of *PE* so that we assume it is more likely that a user thinks the system will perform well when he or she trusts the system.

This transitory connection reveals the importance of trust in the context of intelligent system acceptance. The strong effect of *AB* also reveals that a prior belief in the system's problem-solving capability is fundamental. Especially when discussing algorithm appreciation vs. algorithm aversion, this particular construct plays a central role in building up *TP* towards the system. We theorize that the observation of algorithm appreciation or aversion is connected to *AB* and *TP* since they determine what to expect from a system. Trusting a system and expecting super-human performance in the case of algorithm appreciation can turn into mistrust when an aversion is built up due to the individuality of a single task or erratic system behavior (Dietvorst et al., 2015; Logg et al., 2019). However, as argued in XAI literature, an explanation of some sort can help to increase trust in the system (Adadi & Berrada, 2018; Páez, 2019).

System transparency enables trust building and contributes to performance expectancy in both ways

Including *ST*, we found that revealing the system's internal decision structure (global explanation) and explaining how it decides in individual cases (local explanation) positively affects almost all constructs. First, we can confirm that an

understanding of or at least visibility into the system's decision process has a powerful effect on the user's (initial) trust in the system, confirming the often-postulated connection that motivates much XAI research (Ribeiro et al., 2016). Second, we find that *ST* also has substantial effects on *PE* and in terms of usability (i.e., *EE*). We expected the strong connection of *ST* to *EE* as a global system explanation is usually required to determine the effort it takes to efficiently train and subsequently use an intelligent system (Wanner et al., 2020). It is reasonable to assume that the presence of an explanation in a psychological sense reduces uncertainty and thus technological anxiety towards the system (Miller, 2019). Therefore, we theorize that the presence of local and global explanations lets the user shift to a more rational behavior since he or she can make more informed decisions rather than relying on their gut when dealing with black-box intelligent systems.

When comparing the observed effects with related literature such as Wanner et al. (2020), which deals with determining the decision factors for adopting intelligent systems, we find that the relationship between explanation performance and using a system is a more complex one. While we cannot draw conclusions regarding a trade-off, as stated in Wanner et al. (2020), we found that the presence of an explanation indirectly influences the expected performance of a system, which is often the dominant influence factor. Therefore, we argue that while performance remains an essential factor for the actual intention to use, *ST* should be attributed a more critical role than current findings suggest since it can significantly increase the *PE* (or lower it depending on the revealed information through the explanation).

Additionally, taking temporal factors into account, we argue that initial trust factors and subsequently expected performance and attitude towards the system are formed by the information that is revealed before the system is used. That is, the availability of *ST* can steer those factors in one direction or another before the user sets his or her *PE*. Thus, we argue that it is less of a situational trade-off and more of a decision process that is repeated with each use and thereby manifesting in the user's attitude toward the system and AI technology in general.

Intelligent systems are a broad concept and may require contextualization

There is a plethora of research on trust in and transparency of technology – considering each aspect separately for the most part – as pointed out in the theorizing sections. We have focused our theorizing on UTAUT-related literature, but we have found that the relating constructs have been discussed similarly without relation to UTAUT. Our core contribution is twofold in that we propose to consider the combination of transparency and trust as well as their latent indirect effects to explain a user's intention to use a system. So far, research on AI acceptance has primarily focused on direct effects, where performance stands out (cf. e.g., Wanner et al., 2020) or considered trust or transparency separately (see Section "Research theorizing").

Our contribution further distinguishes itself from prior art as we focus on intelligent systems as any IT system that can make decisions indistinguishable in performance from or better than a human being based on analytical models that are opaque to the end-user. This definition is independent of the decision task. Consequently, our UTAUT model is designed as a broad model. Hence, its contribution is that it is applicable to multiple types of intelligent systems. As a consequence of this breadth, our model may lack precision for some applications. There may be factors that further affect intention to use in one case but not in another. We do not cover these domain-specific factors. We provide a base model that has merits of its own and can be extended with further constructs such as, for example, facilitating conditions or social influence if the scenario necessitates this.

Much of the extant literature has focused on domain-specific applications such as recommendation systems to support selection processes or human-computer interaction with AI agents that exhibit physical anthropomorphic demeanor. Our model can be used in these contexts but will not measure demarcating aspects such as the effect of physical interaction with AI agents.

Practical implications

Use expectation management to form attitude towards the system

In order to avoid disappointment and algorithmic aversion, managing the expectations towards performance can increase subsequent intention to use, even if the problem field for application is limited in the process since hesitation is build up through the system's self-signaling of suboptimal performance. In line with Dietvorst et al. (2016), it is important to manage expectations and show the user control opportunities of the system. This can be done with a pre-deployment introductory course involving users in the configuration state while using their knowledge in training the algorithms at the base of the intelligent system (Nadj et al., 2020).

Besides providing support for managing expectations and learning to use the system (Dwivedi et al., 2019), overcoming initial hesitation has a high priority in the case of intelligent systems.

Control the level of system transparency based on the target audience's capabilities and requirements

Global explanations depict the inner functioning and complexity of an intelligent system. They are suitable to manage the expected effort when procuring an intelligent system, specifically through either outsourcing or in-house development. In addition, global explanations can provide a problem/system-fit perspective in that the user can observe whether the complexity of the model is suitable for the task. For example, using a complex deep learning model for an intelligent system to detect simple geometric shapes such as cracks might even decrease performance.

Local explanations can assist with explaining single predictions of intelligent systems, helping the user to compare the decision process by i) visualizing the steps towards the decision (e.g., by creating images of the intermediate layers of the artificial neural network) and by ii) attributing the input data importance regarding the output decision (e.g., by creating a heatmap of input pixels that caused the intelligent system's decision).

Explanations can also prove useful as a communication bridge between developers of the intelligent system who are not domain experts and the domain experts who are AI novices. This helps to diagnose the model and create a common understanding of the decision process from a human point of view enabling all stakeholders to jointly avoid false system behavior that can lead to algorithm aversion, such as learning a wrong input-output relation.

However, disclosing too much information about the principal rationale of the intelligent system can lead to the opposite effect (Hosanagar & Jair, 2018; Kizilcec, 2016). Especially for the stakeholder group of domain experts that are the users of the system, as opposed to developers who are required a global explanation to diagnose system failure.

Implement trust management independent of transparency efforts

Our results also show that while being influenced by transparency, trust is not solely explained by it. In accordance with Madsen and Gregor (2000), the pre-existing propensity to trust that is reflected by *TP* requires extra treatment that goes beyond simply providing explanations. Thus, trust issues need to be addressed head-on by implementing guidelines for trustworthy AI (Thiebes et al., 2021). Furthermore, companies should think about introducing trust management. For similar reasons, the standard and idea of risk management were introduced decades ago: identify uncertainty roots and trust concerns and create trust policies (Müller et al., 2021).

The uncertainty regarding *PE* and *EE* could be reduced proactively by offering training to the users to experience the intelligent system to form a feeling of beneficence (Thiebes et al., 2021). Using the system in a training session in a non-critical context can support the acceptance of the system and provide a solution to the initial uncertainty about the performance. According to Miller (2019), this could provide partial transparency, in this case, as an indicator of ability and performance.

Limitations

In our study, we presented a use case based on a medium-stake scenario. Here, wrong decisions have consequences such as machine breakdown or downtimes within the production plant. This can result in high monetary loss. Nevertheless, wrong decisions do not endanger human lives. We used this scenario for two reasons. First, for the sake of generalization, and second, we tried to replicate a typical industrial medium-stake maintenance use case. However, following Rudin (2019), we need to keep in mind that user behavior may differ in high-stake use cases resulting in bodily harm due to the potential consequences of wrong decisions. Inversely, this also applies to low-stake use cases. Further, using a work system scenario entails that users cannot opt to not use the system. In the consumer space, where consumers can decide to choose a non-intelligent system or use no system at all, the results and necessary constructs may differ.

Further, we focused on user perception. Consequently, we cannot verify if the user's perception corresponds to the actual user behavior. This is especially related to the following: *PE* on whether the system can increase the user's productivity, *EE* on whether the user finds the system easy to use, and *ST* on whether the user understands why the system made the decision it did. The latter is closely related to findings from Herm et al. (2021), who address the knowledge gap on the perceived explainability of intelligent system explanations and user task solving performance.

Lastly, within our use case, we provided a textual and graphical explanation for intelligent system predictions and did not impose time limits for decision. While many different XAI augmentation techniques have been developed in XAI research, further evaluation of these techniques seems necessary. Similarly, the results may differ when different XAI augmentation techniques are applied. Hereby, inappropriate explanations can cause an overload of the user's cognitive capacity (Grice, 1975). Furthermore, a personalized explanation can increase the behavior intention (Schneider & Handali, 2019).

Conclusion and outlook

By extending the UTAUT model with factors of attitude, trust, and system transparency, we were able to explain better the factors that influence the willingness to accept intelligent systems in the workplace.

Our extension centers on affection constructs such as *ATT, TP,* and *AB* while simultaneously integrating *ST* as an opportunity to steer both to address the information asymmetry between black-boxed, anthropomorphic agents and their human principal. This combination as well as the consideration of latent indirect factors provides the community with a means to look beyond performance as the dominating decision factor for intelligent system efficacy.

In summary, on the one hand, our model enables researchers to understand the influence of this human factor for intelligent systems and in more general for analytical AI models. On the other hand, our findings can help to create measures to reduce acceptance barriers in practice and thus better leverage AI capabilities. Since our research is based on the UTAUT model and established extensions, we assume that our model is of general nature and generally transferable to or contextualizable in other domains. The results of our model application may be more specific to work system in maintenance as discussed in the limitations.

Since our research results clearly indicate how behavioral intention is influenced by this human factor, we aspire to develop design principles for intelligent systems that contribute to the user's willingness to accept and use these systems in their daily work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12525-022-00593-5.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Aboulian, A., Green, D. H., Switzer, J. F., Kane, T. J., Bredariol, G. V., Lindahl, P., Donnal, J. S., & Leeb, S. B. (2018). NILM dashboard: A power system monitor for electromechanical equipment diagnostics. *IEEE Transactions on Industrial Informatics, 15*(3), 1405–1414. https://doi.org/10.1109/TII.2018.2843770

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

Alaiad, A., & Zhou, L. (2013). Patients' behavioral intention toward using healthcare robots. In *Proceedings of the19th Americas conference on information systems (AMCIS)*. Illinois.

Albashrawi, M., & Motiwalla, L. (2017). When IS success model meets UTAUT in a mobile banking context: A study of subjective and objective system usage. In *Swedish artificial intelligence society (SAIS) Proceedings, 1*. http://aisel.aisnet.org/sais2017/1

Alharbi, S. T. (2014). Trust and acceptance of cloud computing: A revised UTAUT model. In *International conference on computational science and computational intelligence, Las Vegas, NV, USA* (pp. 131–134). https://doi.org/10.1109/CSCI.2014.107

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*(2), 256. https://doi.org/10.1037/0033-2909.111.2.256

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (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

Baird, A., & Maruping, L. M. (2021). The next generation of research on is use: A theoretical framework of delegation to and from agentic is artifacts. *MIS Quarterly, 45*(1b), 315–341. https://doi.org/10.25300/MISQ/2021/15882

Baishya, K., & Samalia, H. V. (2020). Extending unified theory of acceptance and use of technology with perceived monetary value for smartphone adoption at the bottom of the pyramid. *International Journal of Information Management, 51*, 102036. https://doi.org/10.1016/j.ijinfomgt.2019.11.004

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. *Journal of the Association for Information Systems, 22*(2), 10. https://doi.org/10.17705/1jais.00662

Bentele, G., & Seidenglanz, R. (2015). Vertrauen und Glaubwürdigkeit. Begriffe, Ansätze, Forschungsübersicht und praktische Relevanz. In *Handbuch der Public Relations. Wissenschaftliche Grundlagen und berufliches Handeln. Mit Lexikon* (Vol. 3, pp. 411–430). Springer.

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

Bigley, G. A., & Pearce, J. L. (1998). Straining for shared meaning in organization science: Problems of trust and distrust. *Academy of Management Review, 23*(3), 405–421. https://doi.org/10.5465/amr.1998.926618

Bilgic, M., & Mooney, R. J. (2005). Explaining recommendations: Satisfaction vs. promotion. In *Beyond Personalization Workshop, IUI*.

Blut, M., Chong, A., Tsiga, Z., & Venkatesh, V. (2021). Meta-analysis of the unified theory of acceptance and use of technology (UTAUT): Challenging its validity and charting a research agenda in the Red Ocean. *Journal of the Association for Information Systems, forthcoming, 23*(1), 13–95. https://doi.org/10.17705/1jais.00719

Bröhl, C., Nelles, J., Brandl, C., Mertens, A., & Schlick, C. M. (2016). *TAM reloaded: A technology acceptance model for human-robot cooperation in production systems international conference on human-computer interaction*. Canada.

Brown, S. A., Dennis, A. R., & Venkatesh, V. (2010). Predicting collaboration technology use: Integrating technology adoption and collaboration research. *Journal of Management Information Systems, 27*(2), 9–54. https://doi.org/10.2753/MIS0742-1222270201

Brunk, J., Mattern, J., & Riehle, D. M. (2019). Effect of transparency and trust on acceptance of automatic online comment moderation systems. In *2019 IEEE 21st conference on business informatics (CBI), Russia*.

Brynjolfsson, E., & Mcafee, A. (2017). The business of artificial intelligence. *Harvard Business Review, 7*, 3–11. https://starlab-alliance.com/wp-content/uploads/2017/09/AI-Article.pdf

Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*(2), 220–239. https://doi.org/10.1002/bdm.2155

Carter, L., & Bélanger, F. (2005). The utilization of e-government services: Citizen trust, innovation and acceptance factors. *Information Systems Journal, 15*(1), 5–25. https://doi.org/10.1111/j.1365-2575.2005.00183.x

Carvalho, T. P., Soares, F. A., Vita, R., da P. Francisco, R., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering, 137*, 106024. https://doi.org/10.1016/j.cie.2019.106024

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research, 56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Chen, T.-W., & Sundar, S. S. (2018). This app would like to use your current location to better serve you: Importance of user assent and system transparency in personalized mobile services. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Paper No.: 537* (pp. 1–13). https://doi.org/10.1145/3173574.3174111

Chen, X.-B. (2013). Tablets for informal language learning: Student usage and attitudes. *Language Learning & Technology, 17*(1), 20–36. http://dx.doi.org/10125/24503

Cheng, D., Liu, G., Qian, C., & Song, Y.-F. (2008). Customer acceptance of internet banking: Integrating trust and quality with UTAUT model. In *IEEE International Conference on Service Operations and Logistics, and Informatics* (pp. 383–388). https://doi.org/10.1109/SOLI.2008.4686425

Chin, W. W., & Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. *Statistical Strategies for Small Sample Research, 1*(1), 307–341.

Choi, J. K., & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction, 31*(10), 692–702. https://doi.org/10.1080/10447318.2015.1070549

Chui, M., & Malhotra, S. (2018). *AI adoption advances, but foundational barriers remain*. McKinsey&Company. https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain.

Cody-Allen, E., & Kishore, R. (2006). An extension of the UTAUT model with e-quality, trust, and satisfaction constructs. In *Proceedings of the 2006 ACM SIGMIS CPR conference on computer personnel research: Forty four years of computer personnel research: achievements, challenges & the future* (pp. 82–89). https://doi.org/10.1145/1125170.1125196

Compeau, D., Higgins, C. A., & Huff, S. (1999). Social cognitive theory and individual reactions to computing technology: A longitudinal study. *MIS Quarterly*, 145–158. https://doi.org/10.2307/249749

Cramer, H., Evers, V., Ramlal, S., Van Someren, M., Rutledge, L., Stash, N., Aroyo, L., & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction, 18*(5), 455. https://doi.org/10.1007/s11257-008-9051-3

Dam, H. K., Tran, T., & Ghose, A. (2018). Explainable software analytics. In *Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results* (pp. 53–56). https://doi.org/10.1145/3183399.3183424

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

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology, 22*(14), 1111–1132. https://doi.org/10.1111/j.1559-1816.1992.tb00945.x

Demissie, D., Alemu, D., & Rorissa, A. (2021). An investigation into user adoption of personal safety devices in higher education using the unified theory of acceptance and use of technology (UTAUT). *The Journal of the Southern Association for Information Systems, 8*(1), 1–18. https://doi.org/10.17705/3JSIS.00017

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Dos Santos, D. P., Giese, D., Brodehl, S., Chon, S., Staab, W., Kleinert, R., Maintz, D., & Baeßler, B. (2019). Medical students' attitude towards artificial intelligence: A multicentre survey. *European Radiology, 29*(4), 1640–1646. https://doi.org/10.1007/s00330-018-5601-1

Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2019). Re-examining the unified theory of acceptance and use of technology (UTAUT): Towards a revised theoretical model. *Information Systems Frontiers, 21*(3), 719–734. https://doi.org/10.1007/s10796-017-9774-y

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864. https://doi.org/10.1037/0033-295X.114.4.864

Esfandiari, R., & Sokhanvar, F. (2016). Modified unified theory of acceptance and use of technology in investigating Iranian language learners' attitudes toward mobile assisted language learning (MALL). *Interdisciplinary Journal of Virtual Learning in Medical Sciences, 6*(4), 93–105. https://doi.org/10.5812/ijvlms.12010

Fan, W., Liu, J., Zhu, S., & Pardalos, P. M. (2018). Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). *Annals of Operations Research, 1-26*. https://doi.org/10.1007/s10479-018-2818-y

Fishbein, M., & Ajzen, I. (1977). Belief, attitude, intention, and behavior: An introduction to theory and research. *Philosophy and Rhetoric, 10*(2), 178–188.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems, 4*(1), 7. https://doi.org/10.17705/1CAIS.00407

Gherheş, V. (2018). Why are we afraid of artificial intelligence (Ai)? *European Review Of Applied Sociology, 11*(17), 6–15. https://doi.org/10.1515/eras-2018-0006

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.

Gretzel, U., & Fesenmaier, D. R. (2006). Persuasion in recommender systems. *International Journal of Electronic Commerce, 11*(2), 81–100. https://doi.org/10.2753/JEC1086-4415110204

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics, 37*(3), 362–386. https://doi.org/10.1002/rob.21918

Gupta, B., Dasgupta, S., & Gupta, A. (2008). Adoption of ICT in a government organization in a developing country: An empirical study. *The Journal of Strategic Information Systems, 17*, 140–154. https://doi.org/10.1016/j.jsis.2007.12.004

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice, 19*(2), 139–152. https://doi.org/10.2753/MTP1069-6679190202

Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage publications.

Hebrado, J., Lee, H. J., & Choi, J. (2011). The role of transparency and feedback on the behavioral intention to reuse a recommender system. In *Proceedings of the International Conference on Information Resources Management (CONF-IRM), 8*.https://aisel.aisnet.org/confirm2011/8

Hebrado, J. L., Lee, H. J., & Choi, J. (2013). Influences of transparency and feedback on customer intention to reuse online recommender systems. *Journal of Society for e-Business Studies, 18*(2), 279–299.

Hein, D., Rauschnabel, P., He, J., Richter, L., & Ivens, B. (2018). What drives the adoption of autonomous cars? In *International Conference on Information Systems (ICIS), San Francisco, USA*.

Heinrich, K., Graf, J., Chen, J., Laurisch, J., & Zschech, P. (2020). Fool me once, shame on you, fool me twice, shame on me: A taxonomy of attack and defense patterns for AI security. In *28th European Conference on Information Systems (ECIS), Virtual Conference*.

Herm, L.-V., Heinrich, K., Wanner, J., & Janiesch, C. (2022). Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability. *International Journal of Information Management, 102538*. https://doi.org/10.1016/j.ijinfomgt.2022.102538

Herm, L.-V., Wanner, J., Seubert, F., & Janiesch, C. (2021). I Don't get it, but it seems valid! The connection between Explainability and comprehensibility in (X)AI research. In *29th European Conference of Information systems (ECIS), Marrakech, Morocco*.

Hosanagar, K., & Jair, V. (2018). We need transparency in algorithms, but too much can backfire. *Harvard Business Review, 25*, 2018. https://hbr.org/2018/07/we-need-transparency-in-algorithms-but-too-much-can-backfire

Hsu, C. L., Lin, J. C. C., & Chiang, H. S. (2013). The effects of blogger recommendations on customers' online shopping intentions. *Internet Research, 23*(1), 69–88. https://doi.org/10.1108/10662241311295782

Hwang, W.-Y., Shih, T. K., Ma, Z.-H., Shadiev, R., & Chen, S.-Y. (2016). Evaluating listening and speaking skills in a mobile game-based learning environment with situational contexts. *Computer Assisted Language Learning, 29*(4), 639–657. https://doi.org/10.1080/09588221.2015.1016438

Indarsin, T., & Ali, H. (2017). Attitude toward using m-commerce: The analysis of perceived usefulness perceived ease of use, and perceived trust: Case study in Ikens wholesale trade, Jakarta–Indonesia. *Saudi Journal of Business and Management Studies, 2*(11), 995–1007. https://doi.org/10.21276/sjbms.2017.2.11.7

Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2021). Machine learning towards intelligent systems: Applications, challenges, and opportunities. *Artificial Intelligence Review, 54*(5), 3299–3348. https://doi.org/10.1007/s10462-020-09948-w

Jackson, C. M., Chow, S., & Leitch, R. A. (1997). Toward an understanding of the behavioral intention to use an information system. *Decision Sciences, 28*(2), 357–389. https://doi.org/10.1111/j.1540-5915.1997.tb01315.x

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets, 31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jensen, T., Albayram, Y., Khan, M. M. H., Buck, R., Coman, E., & Fahim, M. A. A. (2018). Initial trustworthiness perceptions of a drone system based on performance and process information. In *Proceedings of the 6th International Conference on Human-Agent Interaction (HAI)* (pp. 229–237). https://doi.org/10.1145/3284432.3284435

Jetter, J., Eimecke, J., & Rese, A. (2018). Augmented reality tools for industrial applications: What are potential key performance indicators and who benefits? *Computers in Human Behavior, 87*, 18–33. https://doi.org/10.1016/j.chb.2018.04.054

Kaur, K., & Rampersad, G. (2018). Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. *Journal of Engineering and Technology Management, 48*, 87–96. https://doi.org/10.1016/j.jengtecman.2018.04.006

Kim, D. J. (2014). A study of the multilevel and dynamic nature of trust in e-commerce from a cross-stage perspective. *International Journal of Electronic Commerce, 19*(1), 11–64. https://doi.org/10.2753/JEC1086-4415190101

Kim, J. (2019). Fear of artificial intelligence on people's attitudinal & behavioral attributes: An exploratory analysis of AI phobia. *Global Scientific Journal, 7*(10), 9–20. https://www.ieeesem.com/researchpaper/Fear_of_Artificial_Intelligence_on_People_s_Attitudinal_Behavioral_Attributes_An_Exploratory_Analysis_of_A_I_Phobia.pdf

Kim, Y. J., Chun, J. U., & Song, J. (2009). Investigating the role of attitude in technology acceptance from an attitude strength perspective. *International Journal of Information Management, 29*(1), 67–77. https://doi.org/10.1016/j.ijinfomgt.2008.01.011

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

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration (ijec), 11*(4), 1–10. https://doi.org/10.4018/ijec.2015100101

Komiak, S. Y., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly, 941-960*. https://doi.org/10.2307/25148760

Kroll, J. A. (2018). The fallacy of inscrutability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133), 20180084. https://doi.org/10.1098/rsta.2018.0084

Lamnek, S., & Krell, C. (2010). *Qualitative Sozialforschung* (6th ed.). Beltz.

Lankton, N. K., 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

Lee, J.-H., & Song, C.-H. (2013). Effects of trust and perceived risk on user acceptance of a new technology service. *Social Behavior and Personality: An International Journal, 41*(4), 587–597. https://doi.org/10.2224/sbp.2013.41.4.587

Lee, M. K., & Turban, E. (2001). A trust model for consumer internet shopping. *International Journal of Electronic Commerce, 6*(1), 75–91. https://doi.org/10.1080/10864415.2001.11044227

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Lotz, V., Himmel, S., & Ziefle, M. (2019). You're my mate–acceptance factors for human-robot collaboration in industry. In *International Conference on Competitive Manufacturing*. Stellenbosch.

Madsen, M., & Gregor, S. (2000). *Measuring human-computer trust Australasian conference on information systems*. Brisbane.

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change, 175*, 121390. https://doi.org/10.1016/j.techfore.2021.121390

Mansouri, S., Kaghazi, B., & Khormali, N. (2011). A survey the views of the students of Gonbad Payam Noor University to mobile learning. In *The first Conference of mobile value-added services in Iran* (pp. 23–24).

Marangunić, N., & Granić, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society, 14*(1), 81–95. https://doi.org/10.1007/s10209-014-0348-1

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., Ledsam, J.R., Melnick, D., Mostofi, H., Peng, L., Reicher, J. J., Romera-Paredes, B., Sidebottom, R., Suleyman, M., Tse, D., Young, K. C., De Fauw J., & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature, 577*(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems (TMIS), 2*(2), 1–25. https://doi.org/10.1145/1985347.1985353

McKnight, D. H., & Chervany, N. L. (2000). What is trust? A conceptual analysis and an interdisciplinary model. In *American Conference on Information Systems (AMCIS), Long Beach, California, USA*.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review, 23*(3), 473–490. https://doi.org/10.5465/amr.1998.926622

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

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

Milojevic, M., & Nassah, F. (2018). *Digital industrial revolution with predictive maintenance*. CXP Group. https://www.ge.com/digital/sites/default/files/download_assets/PAC_Predictive_Maintenance_GE_Digital_Executive_Summary_2018_1.pdf.

Mohseni, S., Zarei, N., & Ragan, E. D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS), 11*(3–4), 1–45. https://doi.org/10.1145/3387166

Mokyr, J., Vickers, C., & Ziebarth, N. L. (2015). The history of technological anxiety and the future of economic growth: Is this time different? *Journal of Economic Perspectives, 29*(3), 31–50. https://doi.org/10.1257/jep.29.3.31

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2*(3), 192–222. https://doi.org/10.1287/isre.2.3.192

Moyne, J., Iskandar, J., Hawkins, P., Walker, T., Furest, A., Pollard, B., & Stark, D. (2013). Deploying an equipment health monitoring dashboard and assessing predictive maintenance. In *Advanced Semiconductor Manufacturing Conference (ASMC 2013 SEMI)* (pp. 105–110). https://doi.org/10.1109/ASMC.2013.6552784

Müller, M., Ostern, N., Koljada, D., Grunert, K., Rosemann, M., & Küpper, A. (2021). Trust mining: Analyzing trust in collaborative business processes. *IEEE Access, 9*, 65044–65065. https://doi.org/10.1109/ACCESS.2021.3075568

Nadj, M., Knaeble, M., Li, M. X., & Maedche, A. (2020). Power to the oracle? Design principles for interactive labeling systems in machine learning. *KI-Künstliche Intelligenz, 34*, 1–12. https://doi.org/10.1007/s13218-020-00634-1

Neufeld, D. J., Dong, L., & Higgins, C. (2007). Charismatic leadership and user acceptance of information technology. *European Journal of Information Systems, 16*(4), 494–510. https://doi.org/10.1057/palgrave.ejis.3000682

Nilashi, M., Jannach, D., & bin Ibrahim, O., Esfahani, M. D., & Ahmadi, H. (2016). Recommendation quality, transparency, and website quality for trust-building in recommendation agents. *Electronic Commerce Research and Applications, 19*, 70–84. https://doi.org/10.1016/j.elerap.2016.09.003

Oh, J.-C., & Yoon, S.-J. (2014). Predicting the use of online information services based on a modified UTAUT model. *Behaviour & Information Technology, 33*(7), 716–729. https://doi.org/10.1080/0144929X.2013.872187

Oliveira, T., Faria, M., Thomas, M. A., & Popovič, A. (2014). Extending the understanding of mobile banking adoption: When UTAUT meets TTF and ITM. *International Journal of Information Management, 34*(5), 689–703. https://doi.org/10.1016/j.ijinfomgt.2014.06.004

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Páez, A. (2019). The pragmatic turn in explainable artificial intelligence (XAI). *Minds and Machines, 29*(3), 441–459. https://doi.org/10.1007/s11023-019-09502-w

Paravastu, N. S., & Ramanujan, S. S. (2021). Interpersonal trust and technology trust in information systems research: A comprehensive review and a conceptual model. *International Journal of Information Systems and Social Change, 12*(4), 1–18. https://doi.org/10.4018/IJISSC.287832

Pawellek, G. (2016). *Integrierte Instandhaltung und Ersatzteillogistik: Vorgehensweisen, Methoden, Tools, 2*, Springer-Verlag.

Persson, A., Laaksoharju, M., & Koga, H. (2021). We mostly think alike: Individual differences in attitude towards AI in Sweden and Japan. *The Review of Socionetwork Strategies, 15*(1), 123–142. https://doi.org/10.1007/s12626-021-00071-y

Peters, F., Pumplun, L., & Buxmann, P. (2020). Opening the black box: Consumer's willingness to pay for transparency of intelligent systems. In *28th European conference on information systems (ECIS)*. Marrakesch.

Pfeuffer, N., Benlian, A., Gimpel, H., & Hinz, O. (2019). Anthropomorphic information systems. *Business & Information Systems Engineering, 61*(4), 523–533. https://doi.org/10.1007/s12599-019-00599-y

Portela, F., Aguiar, J., Santos, M. F., Silva, Á., & Rua, F. (2013). Pervasive intelligent decision support system–technology acceptance in intensive care units. In *Advances in Information Systems and Technologies* (pp. 279–292). Springer.

Pu, P., & Chen, L. (2007). Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems, 20*(6), 542–556. https://doi.org/10.1016/j.knosys.2007.04.004

Rad, M. S., Nilashi, M., & Dahlan, H. M. (2018). Information technology adoption: A review of the literature and classification. *Universal Access in the Information Society, 17*(2), 361–390. https://doi.org/10.1007/s10209-017-0534-z

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD)* (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778

Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint.* https://arxiv.org/abs/1708.08296

Sardianos, C., Varlamis, I., Chronis, C., Dimitrakopoulos, G., Alsalemi, A., Himeur, Y., Bensaali, F., & Amira, A. (2021). The emergence of explainability of intelligent systems: Delivering explainable and personalized recommendations for energy efficiency. *International Journal of Intelligent Systems, 36*(2), 656–680. https://doi.org/10.1002/int.22314

Savage, N. (2022). Breaking into the black box of artificial intelligence. *Nature*. https://doi.org/10.1038/d41586-022-00858-1

Schneider, J., & Handali, J. (2019). Personalized explanation in machine learning: A conceptualization. *arXiv preprint.* https://arxiv.org/abs/1901.00770

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review, 32*(2). https://doi.org/10.5465/amr.2007.24348410

Shahzad, F., Xiu, G., Khan, M. A. S., & Shahbaz, M. (2020). Predicting the adoption of a mobile government security response system from the user's perspective: An application of the artificial neural network approach. *Technology in Society, 62*, 101278. https://doi.org/10.1016/j.techsoc.2020.101278

Shin, D. (2020a). How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. *Computers in Human Behavior, 109*, 106344. https://doi.org/10.1016/j.chb.2020.106344

Shin, D. (2020b). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media, 64*(4), 541–565. https://doi.org/10.1080/08838151.2020.1843357

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

Shin, D., Zhong, B., & Biocca, F. A. (2020). Beyond user experience: What constitutes algorithmic experiences? *International Journal of Information Management, 52*, 102061. https://doi.org/10.1016/j.ijinfomgt.2019.102061

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS), 10*(4), 1–31. https://doi.org/10.1145/3419764

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., & Lanctot, M. (2016). Mastering the game of go with deep neural networks and tree search. *Nature, 529*(7587), 484–489. https://doi.org/10.1038/nature16961

Slade, E. L., Dwivedi, Y. K., Piercy, N. C., & Williams, M. D. (2015). Modeling consumers' adoption intentions of remote mobile payments in the United Kingdom: Extending UTAUT with innovativeness, risk, and trust. *Psychology & Marketing, 32*(8), 860–873. https://doi.org/10.1002/mar.20823

Suleman, D., Zuniarti, I., Sabil, E. D. S., Yanti, V. A., Susilowati, I. H., Sari, I., Marwansyah, S., Hadi, S. S., & Lestiningsih, A. S. (2019). Decision model based on technology acceptance model (tam) for online shop consumers in Indonesia. *Academy of Marketing Studies Journal, 23*(4), 1–14. https://www.abacademies.org/articles/decision-model-based-on-technology-acceptance-model-tam-for-online-shop-consumers-in-indonesia-8624.html

Šumak, B., Polancic, G., & Hericko, M. (2010). An empirical study of virtual learning environment adoption using UTAUT. In *2nd International Conference on Mobile, Hybrid, and On-Line Learning* (pp. 17–22). https://doi.org/10.1109/eLmL.2010.11

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication, 25*(1), 74–88. https://doi.org/10.1093/jcmc/zmz026

Taylor, S., & Todd, P. (1995). Decomposition and crossover effects in the theory of planned behavior: A study of consumer adoption intentions. *International Journal of Research in Marketing, 12*(2), 137–155. https://doi.org/10.1016/0167-8116(94)00019-K

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

Thomas, T., Singh, L., & Gaffar, K. (2013). The utility of the UTAUT model in explaining mobile learning adoption in higher education in Guyana. *International Journal of Education and Development using ICT, 9*(3).

Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: Toward a conceptual model of utilization. *MIS Quarterly, 15*(1), 125–143. https://doi.org/10.2307/249443

Venkatesh, V. (2022). Adoption and use of AI tools: A research agenda grounded in UTAUT. *Annals of Operations Research, 308*(1), 641–652. https://doi.org/10.1007/s10479-020-03918-9

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (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. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178. https://doi.org/10.2307/41410412

Venkatesh, V., Thong, J. Y., & 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

Vidotto, G., Massidda, D., Noventa, S., & Vicentini, M. (2012). Trusting beliefs: A functional measurement study. *Psicologica: International journal of methodology and experimental. Psychology, 33*(3), 575–590.

von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy and Technology, (34)*, 1607–1622. https://doi.org/10.1007/s13347-021-00477-0

Vorm, E., & Combs, D. J. (2022). Integrating transparency, trust, and acceptance: The intelligent systems technology model (ISTAM). *International Journal of Human–Computer Interaction, 1-18*. https://doi.org/10.1080/10447318.2022.2070107

Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems, 23*(4), 217–246. https://doi.org/10.2753/MIS0742-1222230410

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

Wanner, J., Heinrich, K., Janiesch, C., & Zschech, P. (2020). How much AI do you require? Decision factors for adopting AI technology. In *41st International Conference on Information Systems (ICIS), Hyderabad, India*.

Weiber, R., & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS*. Springer.

Wilkinson, D., Alkan, Ö., Liao, Q. V., Mattetti, M., Vejsbjerg, I., Knijnenburg, B. P., & Daly, E. (2021). Why or why not? The effect of justification styles on chatbot recommendations. *ACM Transactions on Information Systems (TOIS), 39*(4), 1–21. https://doi.org/10.1145/3441715

Williams, M. D., Rana, N. P., & Dwivedi, Y. K. (2015). The unified theory of acceptance and use of technology (UTAUT): A literature review. *Journal of Enterprise Information Management, 28*(3), 443–488. https://doi.org/10.1108/JEIM-09-2014-0088

Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly, 31*(1), 137–209. https://doi.org/10.2307/25148784

Yang, H.-D., & Yoo, Y. (2004). It's all about attitude: Revisiting the technology acceptance model. *Decision Support Systems, 38*(1), 19–31. https://doi.org/10.1016/S0167-9236(03)00062-9

Yao, Y., & Murphy, L. (2007). Remote electronic voting systems: An exploration of voters' perceptions and intention to use. *European Journal of Information Systems, 16*(2), 106–120. https://doi.org/10.1057/palgrave.ejis.3000672

Zhao, R., Benbasat, I., & Cavusoglu, H. (2019). Transparency in advice-giving systems: A framework and a research model for transparency provision. In *IUI Workshops'19, Los Angeles, USA*.

Zhao, X., Lynch Jr., J. G., & Chen, Q. (2010). Reconsidering baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research, 37*(2), 197–206. https://doi.org/10.1086/651257