



# Designing Transparency for Effective Human-AI Collaboration

Michael Vössing<sup>1</sup> · Niklas Kühl<sup>1</sup> · Matteo Lind<sup>1</sup> · Gerhard Satzger<sup>1</sup>

Accepted: 13 April 2022 / Published online: 26 May 2022  
© The Author(s) 2022

## Abstract

The field of artificial intelligence (AI) is advancing quickly, and systems can increasingly perform a multitude of tasks that previously required human intelligence. Information systems can facilitate collaboration between humans and AI systems such that their individual capabilities complement each other. However, there is a lack of consolidated design guidelines for information systems facilitating the collaboration between humans and AI systems. This work examines how agent transparency affects trust and task outcomes in the context of human-AI collaboration. Drawing on the 3-Gap framework, we study agent transparency as a means to reduce the information asymmetry between humans and the AI. Following the Design Science Research paradigm, we formulate testable propositions, derive design requirements, and synthesize design principles. We instantiate two design principles as design features of an information system utilized in the hospitality industry. Further, we conduct two case studies to evaluate the effects of agent transparency: We find that trust increases when the AI system provides information on its reasoning, while trust decreases when the AI system provides information on sources of uncertainty. Additionally, we observe that agent transparency improves task outcomes as it enhances the accuracy of judgemental forecast adjustments.

**Keywords** Agent transparency · Design science research · Human-AI collaboration · Human-AI interaction · Trust

## 1 Introduction

The widespread use of artificial intelligence (AI) and the adoption of AI-based information systems (AI-based IS) are radically transforming the way humans work (Buxmann et al., 2021; Maedche et al., 2019; Watson, 2017). In many industries, companies rely on a broad range of interactive systems (e.g., decision support systems, experts systems) that help decision-makers utilize data, models, and formalized knowledge to solve semi-structured and unstructured problems more efficiently (Phillips-Wren, 2013). In recent years, the increasing capabilities of AI, which have

originated from advances in computing power, increased availability of data, and improved algorithms, have fundamentally changed the nature of these systems and, more importantly, the way humans interact with technology (Morana et al., 2018). In contrast to previous generations of systems—which primarily relied on symbolic logic with clear rules and definitions—state-of-the-art AI systems utilize models far more complex than those incorporated by earlier generations. As a result, their inner workings are opaque, and their reasoning is challenging to comprehend even for experts. Systems that utilize AI to enhance decision-making are sometimes referred to as *intelligent decision support systems* (Phillips-Wren, 2013). As the underlying technology continues to evolve, established knowledge in the context of human-computer interaction has to be reevaluated. Notably, the ability of humans to comprehend and utilize these systems has not kept pace with the capabilities and functionalities of AI (Kagermann, 2015; Maedche et al., 2016). So far, the implications of this transformation are largely unknown.

There is evidence that more and more tasks will shift from humans to AI systems in the coming decades (Maniyaka, 2017). However, researchers agree that humans will

---

✉ Michael Vössing  
michael.voessing@kit.edu

Niklas Kühl  
niklas.kuehl@kit.edu

Matteo Lind  
matteo.lind@alumni.kit.edu

Gerhard Satzger  
gerhard.satzger@kit.edu

<sup>1</sup> Karlsruhe Institute of Technology, Kaiserstraße 89,  
76133 Karlsruhe, Germany

have to collaborate with these systems for the foreseeable future (Kamar, 2016). For instance, companies typically rely on their employees' domain knowledge to adjust forecasts (Fildes et al., 2007). As this situation will not change in the near future, researchers argue that systems are needed that augment humans' capabilities instead of replacing them (Seeber et al., 2019). Additionally, it has become apparent that humans need to understand these systems, trust them, and remain in control (Abdul et al., 2018). In this work, we contribute to the understanding of how AI-based IS should be designed to enable effective collaboration between humans and AI systems.

To design AI-based IS that facilitate human-AI collaboration, researchers and practitioners need to understand how humans can augment AI systems, how AI systems can augment humans, and how business processes have to be designed to support this partnership (Rai et al., 2019; Wilson et al., 2018). In 2016, a study highlighted that even though 84% of the surveyed practitioners believe that AI systems can make their work more effective, only 18% would trust their advice when making business decisions (Kolbjornsrud et al., 2016). To address these issues, researchers have primarily focused on mediating factors such as acceptance and credibility (Giboney et al., 2015a; Jensen et al., 2010). Recently, Seeber et al. (2019) surveyed 65 researchers about the risks and benefits of collaboration between humans and AI systems to identify important research questions for the field. Their work also emphasizes that mechanisms that affect the trust of humans in "machine teammates" require further research. While a variety of factors are known to increase the trust of humans (Wang and Benbasat, 2008), one particular approach that has received significant attention in recent years emphasizes that the reasoning of the systems should be made transparent through explanations (Gilpin et al., 2018; Meske et al., 2020). We contribute to this discussion by examining how *agent transparency* (Stowers et al., 2016) affects the collaboration between humans and AI systems (Kioussis, 2002; Benyon, 2014; Patrick et al., 2021). In this work, we study how agent transparency influences the trust of domain experts as well as the collaborative task outcome (i.e., forecasting accuracy). Accordingly, we aim to contribute to the following research question: "*How should agent transparency in AI-based information systems be designed to increase trust and improve task outcomes in the context of human-AI collaboration?*"

To systematically approach this question, we conduct a design science research (DSR) project based on the methodology outlined by Sonnenberg and Brocke (2012). We work with an industry partner in the hospitality industry that has developed a system that utilizes artificial intelligence (AI) to help restaurant owners predict the revenue of their restaurants. Even though the currently used AI system is capable of modeling multiple drivers of a restaurant's revenue (e.g.,

trends and seasonality), restaurant owners often rely on additional factors (e.g., events or weather) that are, so far, difficult to incorporate into AI models. Unfortunately, though, they sometimes apply this information inappropriately to modify the provided forecast. However, in theory, combining the capabilities of humans and AI systems is known as a viable strategy to improve forecasting accuracy. Based on the theoretical concept of agent transparency (Chen et al., 2014; Stowers et al., 2016), we iteratively develop multiple artifacts that we use to evaluate how increasing the transparency of an AI system affects *trust* and *task outcomes* (Terveen, 1995; Crouser & Chang, 2012; Holzinger, 2016).

Our contribution is twofold: First, we formulate testable propositions (TP1 & TP2), derive design requirements, and formulate principles for facilitating human-AI collaboration through expert interviews and an exploratory focus group. Second, we develop and implement two design features (DF1 & DF2) that instantiate the design principles and evaluate them in a naturalistic setting as suggested by Dellermann et al. (2019b). The evaluation shows that the design features improve the collaborative task outcome. We further study how the design features influence the trust of the domain experts. The results reveal that higher transparency of the information that the AI bases its decision on increases the humans' trust in its capabilities. In contrast, communicating sources of uncertainty decreases trust in the AI's reliability. Our results are not only relevant for decision-makers in the hospitality industry—including restaurant owners, hotel managers, or even retail operators—but also provide novel design insights for researchers.

## 2 Fundamentals and Related Work

The following section introduces prior knowledge utilized in the conducted design activities. We outline and discuss relevant literature on human-AI collaboration, agent transparency, as well as trust.

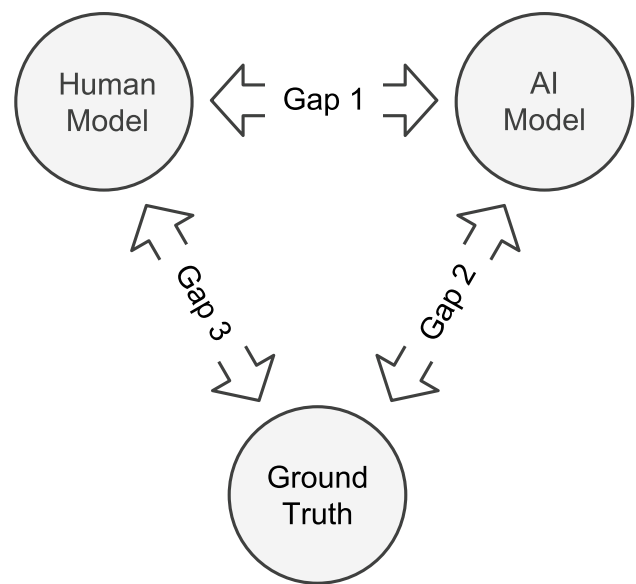
### 2.1 Human-AI Collaboration

In the last decades, information systems that support the decision-making of knowledge workers have been adopted across industries (Power, 2002). Traditionally, many systems rely on symbolic logic with clear rules and definitions, mathematical algorithms, and simulation techniques (Power et al., 2019). However, systems that incorporate artificial intelligence (AI) have become more accessible to companies in recent years. So far, no commonly accepted definition of the term AI has been established in the IS community (Collins et al., 2021; Russell and Norvig, 2016; Kurzweil, 1990). However, two contemporary definitions recently proposed in MIS Quarterly express a

similar understanding of the term: (Rai et al., 2019) define AI as “the ability of a machine to perform cognitive functions that we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem solving, decision-making, and even demonstrating creativity” (p. 1). Similarly, Berente et al. (2021) denote AI as “the frontier of computational advancements that references human intelligence in addressing ever more complex decision-making problems” (p. 1435).

While systems that incorporate AI are built upon the same concepts adopted by previous generations of systems (Ford, 1985; Peter, 1986), they differ fundamentally in terms of the complexity of the underlying models. Today, state-of-the-art models rival the accuracy of domain experts (Ardila et al., 2019). However, to achieve high levels of accuracy, the inner workings of these systems have become increasingly opaque. As a result, users struggle to comprehend their recommendations. In recognition of the increasing capabilities of these systems, researchers have started reconsidering whether these systems are simply tools that support human decision-making and have begun studying how humans can collaborate them closely (Dellermann et al., 2019a; Grønsund and Aanestad, 2020). As a result, established assumptions on how humans interact with technologies need to be reevaluated.

We understand *human-AI collaboration* as problem-solving in which “two or more agents work together to achieve shared goals [...] involving at least one human and at least one computational agent” (Terveen, 1995, p. 67). More and more, researchers envision a future where AI “augments what people want to do rather than substitutes it” (Harper, 2019, p. 1341). In several domains, the feasibility of collaboration between humans and AI systems has already been demonstrated. However, empirical evaluations—both qualitative and quantitative—are rare. Notably, few researchers have studied *how* collaboration works in organizational settings (Fildes & Goodwin, 2013) and *how* these systems should be designed. As outlined by Seeber et al. (2019), collaboration between humans and AI systems entails several challenges and opportunities for the design of socio-technical systems. They emphasize that additional research is needed to conceptualize design features for its facilitation and for their systematic evaluation in naturalistic settings. As of now, there is a lack of consolidated design knowledge regarding how information systems can support collaboration between humans and AI systems. While humans have utilized machines to relieve themselves from mundane or repetitive work for a while, the way humans use technologies to augment their decision-making has changed significantly in recent years. Most notably, the capabilities of information systems have become more sophisticated and harder to understand. Nowadays, the inner workings of these systems are



**Fig. 1** Three gaps in human-AI collaboration, adapted from Kayande et al. (2009)

often opaque (Lee and See, 2004; Söllner et al., 2016b; Rai et al., 2019).

On a conceptual level, we utilize the *3-Gap framework*—introduced by Kayande et al. (2009) and extended by Martens and Provost (2017)—to study human-AI collaboration. Kayande et al. highlight that humans often struggle to understand the recommendations of automated systems, which affects their acceptance of these systems. As depicted in Fig. 1, they describe three gaps between the mental model of humans, the decision model embedded in the system (i.e., the AI model), and reality (i.e., the ground truth). They show that information systems benefit from the alignment of these models and should facilitate it. Our work is motivated by the assumption that complex decision problems require humans to utilize their domain knowledge to complement and support the AI system. We use the term “AI model” as a reference to any element of an information system that utilizes artificial intelligence or machine learning. The three gaps refer to the difference between the AI model and the human model (Gap 1), as well as the gap between these models and reality (Gap 2 & 3). As outlined in the following section, we examine how agent transparency—for example, communicating the reasoning of the AI—affects the first and third gap in human-AI collaboration.

## 2.2 Agent Transparency in Human-AI Collaboration

Transparency is an essential requirement for the design of information systems (Fleischmann & Wallace, 2005; Hosseini et al., 2017; Street & Meister, 2004; Vössing et al., 2019). However, as outlined by Hosseini et al. (2017, p.

3), designing “mechanisms to implement transparency is more complex than deciding whether to make information available”. In recent years, the topic has received considerable attention and has been recognized as an important requirement for the design of information systems (Leite & Cappelli, 2010; Hosseini et al., 2018). We adopt *agent transparency* as a kernel theory for our work. The term refers to the ability of an information system to afford humans’ “comprehension about an intelligent agent’s intent, performance, future plans, and reasoning process” (Chen et al., 2014, p. 2). Chen et al. understand agent transparency as a means to increase the effectiveness of collaboration by increasing humans’ situational awareness. Drury et al. (2003, p. 912) similarly highlight that the “information that collaborators have about each other in coordinated activities” is important for their awareness. Building on Endsley’s (1995) work on situational awareness in dynamic systems, Chen et al. (2014) and Stowers et al. (2016) discuss three transparency requirements that highlight the importance of communicating information about (1) the goals and proposed actions, (2) the reasoning process behind those actions, and (3) the uncertainty associated with those actions. The three requirements are discussed in detail in the following paragraph.

First, systems need to communicate “information about the agent’s current state and goals, intentions, and proposed actions” (Chen et al., 2014, p. 2). This also includes information about the system’s purpose, processes, and current performance. Second, systems need to provide information “about the agent’s reasoning process behind those actions and the constraints/affordances that the agent considers when planning those actions” (Chen et al., 2014, p. 3). Giboney et al. (2015a, p. 2) refer to “descriptions of reasoning processes” as explanations. In the last few years, researchers have extensively studied how explanations effect users (Meske et al., 2020). Explanations should “describe the internals of a system in a way that is understandable to humans” (Gilpin et al., 2018, p. 2). Research shows that explanations can help establish trust (Nunes and Jannach, 2017; Gregor & Benbasat, 1999). Third, systems should communicate “information regarding the agent’s projection of the [...] likelihood of success/failure, and any uncertainty associated with the aforementioned projections” (Chen et al., 2014, p. 3). Epistemic uncertainty is caused by a lack of information and, therefore, can be reduced by either gathering more data or by further refining models (Der Kiureghian & Ditlevsen, 2009). In contrast, aleatory uncertainty refers to the intrinsic randomness of a phenomenon. Stowers et al. (2016) note that uncertainties should be made explicit. Communicating epistemic uncertainty to humans allows them to utilize their domain knowledge to compensate for the AI systems’ lack of knowledge.

### 2.3 Trust in Human-AI Collaboration

Transparency is deeply intertwined with the trust of humans in IT artifacts (Christian & Bunde, 2020; Söllner et al., 2012). Wang & Benbasat define the construct as “an individual’s beliefs in an agent’s competence, benevolence, and integrity” (Wang & Benbasat, 2005, p. 76). They specify that trust in the confidence of agents relates back to the user’s belief “that the trustee has the ability, skills, and expertise to perform effectively in specific domains” (Wang & Benbasat, 2005, p. 76). While the value of trust for collaboration among humans is self-evident (Tschannen-Moran, 2001), transferring the concept of interpersonal trust to the collaboration between humans and AI systems is challenging (Christian & Bunde, 2020). Scholars have long debated whether technological artifacts can or should be recipients of trust and whether human characteristics should be ascribed to them (Wang & Benbasat, 2005). However, the importance of trust for technology acceptance has been shown in numerous studies (Christian & Bunde, 2020; Söllner et al., 2016b; Söllner et al., 2016a). Muir (1994) argues that building trust is an essential means to address complexity and uncertainty because humans cannot have complete knowledge of most AI’s decision models. We argue that trust should be attributed to these artifacts. In the context of AI-based information systems, trust is regarded as a central requirement for successful system application (Siau and Wang, 2018) and has been empirically examined in many cases, such as autonomous driving (Wang et al., 2016a), intelligent medical devices (Hengstler et al., 2016), or recommender systems (O’Donovan & Smyth, 2005).

Measuring trust is challenging. Madsen and Gregor (2000) developed a questionnaire to measure trust between humans and computational agents. Their psychometric instrument differentiates between the cognitive and affective components of trust. They define trust as “the extent to which a user is confident in, and willing to act based on, the recommendations, actions, and decisions of an artificially intelligent decision aid” (Madsen & Gregor, 2000, p. 1). Their definition emphasizes that trust encompasses not only confidence in the actions, but also the willingness to act on the provided advice. If humans have sufficient evidence for the system’s capabilities, their confidence likely arises from reason and fact. However, when sufficient evidence is not available, trust is primarily based on faith (Shaw, 1997). Researchers generally agree that trust and transparency are closely linked. However, only a few researchers have investigated the complex nature of their relationship. As outlined by Kizilcec (2016), transparency can “promote or erode users’ trust in a system by changing beliefs about its trustworthiness”. While the positive effects of transparency are widely recognized, some researchers also report negative side effects. For



example, access to too much information can erode users' trust if the provided information does not match the user's expectations. Transparency can also reduce the understanding of users if the provided information is confusing or overwhelming (Pieters, 2011). In conclusion, the evidence regarding how exactly transparency affects trust is inconclusive (Kizilcec, 2016).

### 3 Research Method

Following the Design Science Research (DSR) paradigm (Hevner et al., 2004), we conducted multiple concurrent design activities and evaluation episodes. We conducted four activities according to the design-evaluation pattern outlined by Sonnenberg and Brocke (2012): IDENTIFY PROBLEM, DESIGN, CONSTRUCTION, and USE. We embraced the evaluation guidelines of the *Framework for Evaluation in Design Science Research (FEDS)* as we combined formative and summative evaluation episodes (Venable et al., 2016). We reported our results utilizing the six dimensions for the communication of DSR projects suggested by vom Brocke and Maedche (2019). Throughout our research process, we aimed to derive prescriptive knowledge on how to design agent transparency in order to achieve effective human-AI collaboration. To empirically evaluate two testable propositions, we developed and implemented two design features in close cooperation with an industry partner.

We began our research by extending our understanding of the problem space and application context through expert interviews (EE1). Through the interviews, we derived three practically-grounded *design requirements* (Jarke et al., 2011) that set the foundation for our DESIGN activities. Building on these requirements, we derived two *design principles* (Gregor & Hevner, 2013). We developed multiple non-functional prototypes to narrow down the problem space (CONSTRUCT). The non-functional prototypes—among others, a system that helps restaurant owners estimate the revenue of their restaurants more accurately—were subsequently discussed in a focus group (EE2). Incorporating the feedback from the focus group, we ultimately implemented two *design features* (Sonnenberg & Brocke, 2012) that provide agent transparency. The design features increased the transparency of the AI system used to forecast the restaurant's revenue by either providing explanations of the AI's reasoning (DP1, Section 4.1) or communicating additional information that allows the domain experts to understand the uncertainty associated with the AI's predictions (DP2, Section 4.1). Finally, we concluded our research with a summative evaluation episode (EE3) by demonstrating the artifacts in two case studies (USE).

### 3.1 Constructs and Testable Propositions

In the summative evaluation episode (EE3), we investigated two testable propositions. Building on the 3-Gap framework outlined in Section 2.1, we aimed to understand how agent transparency—which is a means to reduce the information asymmetry between the human's mental model and the AI's decision model (Gap 1)—affects trust and task outcomes. Accordingly, we formulated the following two testable propositions:

*Agent transparency in AI-based information systems increases trust in the context of human-AI collaboration (TP1):* Trust in technological artifacts is recognized as an important factor in the design of information systems (Lee & See, 2004; Mcknight et al., 2014). Trust is particularly important in the context of complex technical systems, which are characterized by steadily increasing levels of automation (Söllner et al., 2016b; Muir, 1994). As discussed in Section 2.3, effective collaboration requires trust in the capabilities of the other collaborators. Muir (1994) notes that systems should be observable and emphasizes the importance of the “transparency of the automation's behaviour” (p. 1920). In this work, we examine agent transparency as a means to increase the domain experts' level of trust by providing them with additional information on the AI's decision model.

*Agent transparency in AI-based information systems improves task outcomes in the context of human-AI collaboration (TP2):* Ultimately, collaboration between humans and AI systems should improve the task outcome. In our work, the task outcome is measured in the accuracy of the revenue forecast. The knowledge of domain experts is valuable because they can provide subjective judgments in cases where data is difficult to measure objectively (Einhorn, 1974). Additionally, experts can use their domain knowledge to interpret information that models would consider to be outliers (Blattberg and Hoch, 2008). Personnel planners rely strongly on their domain knowledge and consider a variety of factors in their decision-making. Accordingly, we hypothesize that agent transparency should improve the ability of the domain experts to use their knowledge to adjust the AI's forecast and, thus, increase the overall accuracy of the forecast.

### 3.2 Application Context and Problem Description

In the gastronomy sector, wages typically account for a third of all costs. Accordingly, optimizing both the amount of personnel and the utilization of employees by creating suitable work schedules is essential for restaurant owners. Typically, shifts for the following week must be finalized in advance. Notably, many commercially available systems solely focus on the scheduling aspect of personnel planning. However,

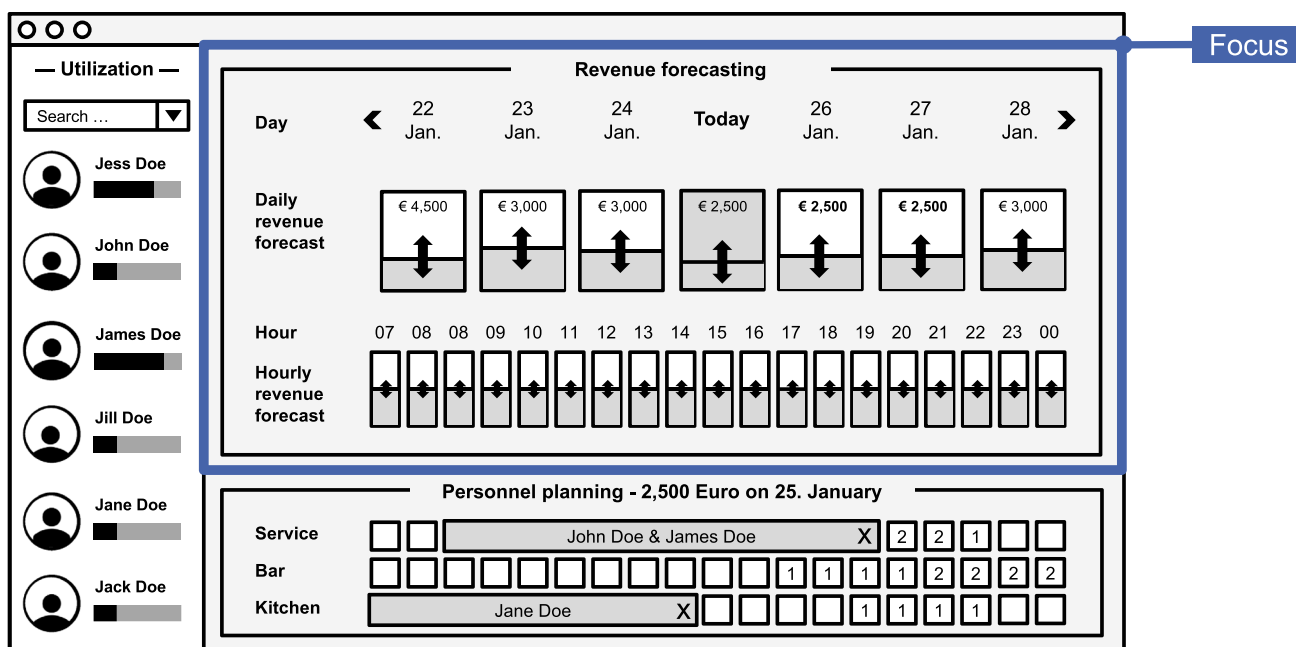


Fig. 2 Schematic visualization of software

the effectiveness of personnel planning depends disproportionately on the ability of restaurant owners to anticipate the restaurant's revenue. Generally, restaurant owners have to predict the revenue of their restaurant(s) for each hour of the next week to assign employees to shifts. Estimating the revenue multiple days in advance is a highly complex task that requires balancing multiple factors, such as historical revenue, weather forecasts, holidays, and upcoming events in the area. Restaurant owners typically rely on their tacit domain knowledge to make these forecasts. Our industry partner has developed a product that surpasses the capabilities of traditional decision support systems—which generally require the restaurant owners to estimate the revenue manually—by utilizing artificial intelligence to predict the restaurant's revenue. However, personnel planners often adjust the provided forecasts. Restaurant owners heavily rely on their tacit knowledge in their decision-making processes—which mirrors decision-making in many high-stakes environments. In the examined application context, restaurant owners can use the AI component (as part of a more comprehensive personnel planning system) to challenge and verify their assumptions.

The system, developed by our industry partner, utilizes an ensemble of multiple machine learning models to forecast the revenue of each restaurant. More specifically, (a) a LightGBM model (i.e., a gradient boosting framework for decision trees) (Ke et al., 2017), (b) a custom neural network, and (c) the model provided by the Prophet library (i.e., an additive model that considers non-linear trends) (Taylor

& Letham, 2018) are used. All models are trained independently for each restaurant, and are then compared against each other to determine the best performing model class for each restaurant. The AI system is dynamically adapting to its environment and is learning new patterns over time.<sup>1</sup> Every night at 11 pm the current model used for each restaurant is retrained to account for new information (i.e., the AI system learns continuously), and every month another round of hyper-parameter tuning and model pre-selection is conducted. The arising collaboration between humans and the AI system offers a unique environment to study human-AI collaboration in a naturalistic environment. Figure 2 provides a schematic visualization of the personnel planning software developed by our industry partner: Its AI-based *revenue forecasting* component, which is the focus of this work, displays the AI's revenue predictions for the upcoming week. The daily revenue forecast is automatically distributed over the operating hours—which is displayed below. The user can modify both the daily and hourly revenue forecast.

### 3.3 Methodology for the Evaluation Episodes

**EE1** We began our research by conducting expert interviews ( $n = 14$ ) to better understand the application context (see Appendix B). In particular, we focused on validating the

<sup>1</sup> Please note that we temporarily suspended this continuous improvement for our experiment, as we aimed to isolate the effects of our transparency treatments.

meaningfulness of the research problem and extending our understanding of the goals and tasks of the users. We conducted narrative interviews—lasting between 60 and 120 minutes—based on a semi-structured interview guideline (Patton, 2002). To ensure a representative portrayal of the application context, we included different restaurant chains utilizing distinct distributions concepts (e.g., holding, franchise). To ensure the reliability of our results, we conducted purposeful sampling until we reached theoretical saturation (Palinkas et al., 2015). We analyzed the interviews according to Patton (2002) through paraphrasing and codification. Based on the refined problem understanding, we derived three requirements (see Section 4.1). Most notably, they informed our decision to examine trust and task outcomes as proxies for human-AI collaboration.

**EE2** We iteratively developed multiple non-functional prototypes. More specifically, we used an exploratory focus group ( $n = 5$ ) to converge the design space and prioritize multiple viable design features. The focus group consisted of domain experts familiar with the restaurant industry as well as method experts experienced in software development, machine learning, and the design of information systems. We conducted the focus group according to the methodology outlined by Tremblay et al. (2010). They highlight that focus groups are an effective way to improve an artifact's design or confirm its utility. Exploratory focus groups study an artifact to propose improvements in the design. Confirmatory focus groups investigate an artifact to establish its utility in the application context.

**EE3** Based on the interview study and the conducted focus group, we implemented two artifacts. Both artifacts increase the transparency of the AI system to help restaurant owners determine when they should use their domain knowledge to revise the provided forecast and when they can trust the system's predictions. We conducted two case studies to demonstrate the utility of the developed artifacts in a naturalistic setting (Hevner et al., 2004). So far, no unified theory has been proposed to measure the effectiveness of human-AI collaboration. Therefore, we relied on established methods from the domain of human-computer interaction and forecasting. Personnel planners generally create work schedules for the following week in advance—usually on Thursdays—based on an hourly forecast of the restaurant's revenue. Accordingly, we structured the case studies as follows: We first recorded three representative planning cycles (i.e., three weeks) at each restaurant to establish a baseline for the current level of trust and task outcome (i.e., the control period). We then introduced one of the implemented design features in each restaurant and recorded three additional planning cycles (i.e., the treatment period). To control for confounding variables, we evaluated the design features

simultaneously in two restaurants that belong to the same restaurant chain and are located less than one kilometer apart. They are of similar size, equally profitable, and have similar opening hours (typically 8 am to 11 pm). The two restaurants are operated by experienced managers with more than five years of experience in the hospitality industry—both actively using the forecasting software developed by our industry partner. For each period, we recorded all hourly revenue forecasts generated by the AI system, as well as the adjustments made by the domain experts. The total number of collected observations varies slightly because observations, when the restaurants were closed, were removed. Based on the collected data, we evaluated the testable proposition as follows:

### 3.4 Measurement of Constructs

To examine the first testable proposition (TP1) (see Section 3.1), we needed to measure the domain experts' trust in the system. As trust is challenging to measure, we combined subjective and objective techniques. We collected empirical data by using a psychometric instrument proposed by Madsen and Gregor (2000) for the measurement of trust between humans and cognitive agents (see Appendix A). The proposed instrument deals specifically with information systems—referred to as intelligent systems in the original work. It does not assume prior knowledge of the underlying technology nor of how exactly a system reaches its conclusions. Additionally, we measured how often the domain experts adjust the forecast (Castelfranchi & Falcone, 2000).

To evaluate the second testable proposition (TP2), we needed to measure the task outcome. In the studied application context, the accuracy of the forecast and the quality of the judgemental forecast adjustments are viable means to evaluate the task outcome. The accuracy of an individual forecast can be determined by calculating an error metric between the predicted and actual values—as recorded in each restaurant's point of sale system. Hyndman (2006) suggest scale-free error metrics. They define the accuracy of an individual forecast as a ratio to an average error achieved by a suitable baseline method. The resulting values can be used to compare forecast accuracy between different time series (Hyndman, 2006). Hyndman and Koehler propose the mean absolute scaled error (MASE), which is defined as the mean of the errors scaled “by the in-sample mean absolute error obtained using the naive forecasting method” (Hyndman & Koehler, 2005, p. 694). As the revenue of a restaurant exhibits relatively consistent daily and hourly patterns, we selected the in-sample revenue of the previous week as a suitable naive forecasting method. To evaluate the task outcome, we examined the effect size in terms of *percent error reduction* (Armstrong, 2007). We specifically measured the accuracy of the AI's forecast (Gap 2), as well

as the accuracy of the forecast after the adjustment (Gap 3). We further examined different types (e.g., wrong direction, undershoot, and overshoot) of forecast adjustments (Petrooulos et al., 2016).

## 4 Artifact Design and Construction

### 4.1 Design Requirements and Principles

As outlined in Section 3, we began our research by conducting multiple expert interviews (EE1). The interviewees almost unanimously noted that estimating the revenue of a restaurant is an essential prerequisite for personnel planning. The interviewees reported that they perform revenue forecasting as a dedicated step during the weekly creation of the work schedules. *Alpha* notes that for accurate forecasts, knowledge about the different events happening in the city is essential. *Zeta* said that he<sup>2</sup> continuously tries to save events in his schedule for this purpose. Additionally, most interviewees indicated that building and applying their tacit knowledge is an essential aspect of their work. *Beta* and *Kappa* pondered if and how tacit knowledge could be integrated into the software they use to create the work schedules. *Delta* expressed another unusual sentiment, saying that he will not use the software if he cannot overwrite the AI's decisions if needed. Based on these interviews, we derived three design requirements: “*In AI-based information systems the AI system should support but not replace humans (DR1).*”, “*In AI-based information systems the AI system should explain recommendations (DR2).*”, “*In AI-based information systems the AI system should improve task outcomes (DR3).*” Subsequently, we conducted a focus group with method and domain experts to derive design principles from the design requirements (EE2). The participants discussed the viability of multiple non-functional prototypes. Based on the derived design requirements and the presented literature, we formulated the following two design principles (DF1 & DF2) to facilitate human-AI collaboration:

*Provide AI-based information systems with the ability to communicate information regarding the AI system's reasoning (DP1):* Research shows that users are often skeptical and reluctant to use AI systems, even if they are known to improve task outcomes (Martens & Provost, 2017). In our application context, the domain experts

expressed a general skepticism of automated forecasts during the first evaluation episode (EE1) and expressed the need to understand the reasoning of an AI system in order to comply with it. However, the complexity of models used by the AI system makes communicating its entire reasoning process infeasible. Through the interviews (EE1), we also know that domain experts sometimes look up the revenue of the same day in the previous year and consult the available weather forecasts for the following week to evaluate the AI's suggestions. As the AI system relies on these data sources as well, communicating them to the domain experts should increase their understanding of the decision model. Accordingly, the first design feature allows the AI to provide an explanation for its reasoning to the user—which aligns with the second transparency requirement.

*Provide AI-based information systems with the ability to communicate information regarding the AI system's uncertainty (DP2):* As Stowers et al. (2016, p. 1706) outlined, “knowledge of the state of the world is likely never fully known by any sensing or perceiving” system. Accordingly, the second design principle states that the AI system can only consider a limited subset of information. For example, an experienced personnel planner might consider information that is not available in a digitized format. Similarly, specific information might be available to the AI system but could be challenging to evaluate (e.g., rare events). We embrace this limitation by allowing the AI to communicate information that allows the domain experts to assess the forecast's uncertainty (Stowers et al., 2016). The domain experts can evaluate this information by utilizing their tacit knowledge and adjusting the AI system's forecast if needed. Based on the interviews (EE1), we know that the domain experts attribute high importance to local events. These events are currently not incorporated into the AI system's forecast but are digitally available through several sources. The second design principle, therefore, satisfies the third transparency requirement.

### 4.2 Solution Description and Design Features

We implemented two design features to instantiate the outlined design principles. Figures 3 and 4 provide schematic visualizations of them. We incorporated the features directly into the software offered by our partner company.

Figure 3 visualizes the first design feature (DF1)—which we refer to as the *explanation* component. It consists of a toolbar made up of three components. Each component allows the domain expert to explore a different subset of the data taken into account by the AI system to create the forecast. The decision model currently considers the weather forecast, the restaurant's historical revenue, as well as an indicator of whether a specific day is a holiday in the forecast. Accordingly, the components allow the domain experts to explore the same data utilized by the AI's decision model.

<sup>2</sup> To ensure a steady reading flow, we use only male pronouns (he, his, him) when necessary. This always includes the female gender and/or does not allow any conclusions regarding the gender of the interviewees.



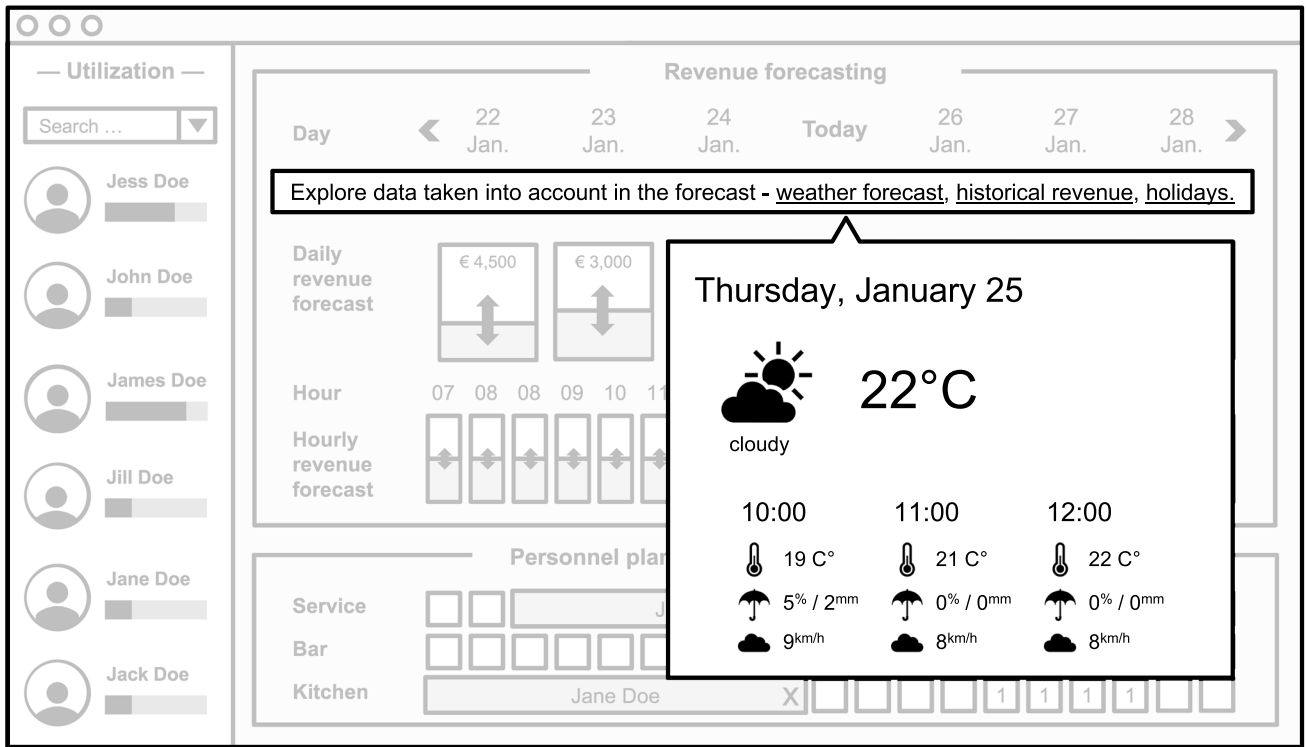


Fig. 3 Schematic visualization of the explanation component (DF1)

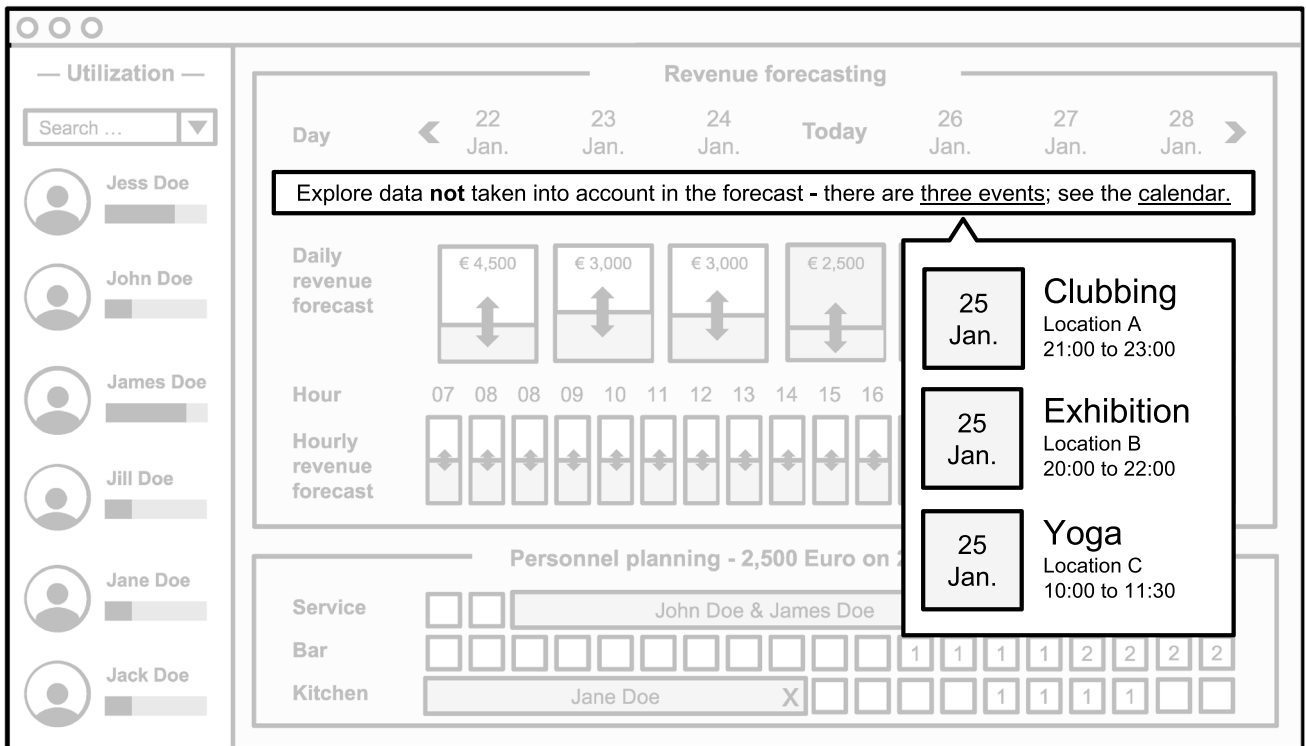


Fig. 4 Schematic visualization of the uncertainty component (DF2)

**Table 1** Summary of the domain experts access to the design features

	Explanation (DF1)		Uncertainty (DF2)	
	Control	Treatment	Control	Treatment
Weather forecast	0	46	N/A	N/A
Historic revenue	0	18	N/A	N/A
Holiday overview	0	7	N/A	N/A
Event overview	N/A	N/A	0	17
Event calendar	N/A	N/A	0	8

Figure 4 visualizes the second design feature (DF2)—which we refer to as the *uncertainty* component. Based on the interviews, we know that domain experts typically examine local events before adjusting the forecast. The AI system does not currently take events into account. To address this limitation, the uncertainty component shows events in the proximity of the restaurants to allow the uncertainty of the AI system's forecast to be assessed. The events are obtained from publicly accessible sources (e.g., event calendar). It is difficult for the AI system to assess these non-periodical events. Therefore, the design feature communicates the events to the restaurant owner to allow for a manual assessment of their effect on the restaurant's revenue. The second design feature embraces the limits of the AI system's knowledge by highlighting that certain information has not yet been incorporated into the forecast. Domain experts can either view events on the selected day or look at all upcoming events on a calendar.

## 5 Results

In Section 3.1 we formulated two testable propositions and argued that agent transparency in an information system affects trust (TP1) as well as task outcomes (TP2) in the context of human-AI collaboration. The following section reports the results of the conducted case studies (EE3). We included capabilities to measure how often the domain experts interacted with the design features during the treatment period to assess their usage. As outlined in Table 1, the domain experts interacted 71 times with the explanation component, particularly to examine the weather and review historical revenue. The event component was used 25 times to investigate 33 events that happened during the treatment period in the vicinity of the restaurants. In the following sections, we utilize the 3-Gap framework (see Section 2.1) to outline how the design features affected trust and task outcomes (TP1 & TP2).

**Table 2** Analysis of the collaborative behaviour (Gap 1)

	Explanation (DF1)		Uncertainty (DF2)	
	Control	Treatment	Control	Treatment
Sample				
– Duration [in days]	21	21	21	21
– Observations [#]	310	293	314	321
Behaviour				
– Reliance	64.84%	36.86%	46.82%	49.22%
– Adjustment	35.16%	63.14%	53.18%	50.78%

### 5.1 Effect of Transparency on Trust

The 3-Gap framework postulates that collaboration between humans and AI systems requires aligning their models. In particular, humans need to understand and internalize the AI's decision model (Gap 1). The implemented design features increase the transparency of the AI's decision model. We, therefore, examine how this transparency affects the first gap, specifically the number of forecasts the domain experts adjust and their trust in the AI system.

In the studied application context, domain experts can exercise control by adjusting the forecast. Table 2 summarizes their adjustment behavior during the control and treatment period. First, we look at the share of forecasts the domain experts adjusted. The results show that the explanation component (DF1) led to a substantial increase in the share of adjusted forecasts (35.16% to 63.14%). In contrast, the uncertainty components (DF2) had a negligible effect on the share of adjusted forecasts (53.18% to 50.78%). Traditionally, researchers would expect trust and control to have a negative relationship (Castelfranchi & Falcone, 2000). Hence, the domain experts' behavior would indicate that DF1 decreases trust and that DF2 does not have a noticeable effect on trust. However, this reasoning does not align with the additional measurement of trust collected through the administered questionnaires. We discuss these contradictory results in Section 6.2.

As outlined in Section 3.3, we utilized the psychometric instrument developed by Madsen and Gregor (2000) to measure the domain experts' level of trust before and after the treatment period. The collected data is provided in the appendices (see Appendix C). Before the treatment period, the domain experts expressed a comparable level of trust. However, more important for our research is the effect of the design features on the level of trust and its different components. Both design features had no noticeable influence on the domain experts' affect-based trust. However, they did affect the domain experts' cognition-based trust in the following ways: The usage of the explanation component (DF1) had a positive effect on cognition-based trust, even though the number of adjusted forecasts increased simultaneously. In particular, increased confidence in the AI's technical competence can be observed ( $M = 3.8$  to  $M = 5.6$ ). The

**Table 3** Analysis of the judgemental forecast adjustments (Gap 1)

	Explanation (DF1)		Uncertainty (DF2)	
	Control	Treatment	Control	Treatment
Tendency				
– Increase	77.06%	83.24%	100%	100%
– Decrease	22.94%	16.76%	0	0
Direction				
– Correct direction: Undershoot	23.85%	42.17%	45.51%	42.33%
– Correct direction: Overshoot	26.61%	14.05%	8.38%	16.56%
– Wrong direction	49.54%	43.78%	46.11%	41.11%

belief that the system uses appropriate methods to make decisions and that it analyzes problems consistently was emphasized. In contrast, the uncertainty component (DF2) led to an overall reduction in trust. Even though trust in the system’s technical capabilities did not change notably, trust in the reliability of the system was reduced ( $M = 3.8$  to  $M = 3.0$ ). Its usage caused doubt in the ability of the AI system to analyze problems consistently, and therefore in its general reliability. Additionally, diminished understandability was reported ( $M = 5.8$  to  $M = 4.6$ ). It was revealed that it became unclear how exactly the AI was supposed to assist domain experts. This aligns with the fact that the domain expert declared a reduced understanding of how to use the system without understanding exactly how it works. These observations are counter-intuitive to the general understanding that control and trust exhibit a negative relationship. We discuss this discrepancy between conventional wisdom and the collected data in Section 6.

**5.2 Effect of Transparency on Task Outcomes**

Ultimately, collaboration between humans and AI systems should improve the task outcome. In the following, we analyze how agent transparency affected the accuracy of the forecast. The accuracy of the forecast depends on the domain experts’ ability to adjust the forecast provided by the AI (Gap 3). As hypothesized in TP1 and TP2, a better understanding of the AI’s decision model (Gap 1) should improve the domain experts’ ability to make judgemental adjustments (Gap 3). As outlined in Section 3.1, we measure the task outcome by examining the accuracy of the judgmental forecast adjustments (Petropoulos et al., 2016) and their effect on the accuracy of the forecast.

As outlined in Section 4, the design features should decrease the number of adjustments in the wrong direction. The data presented in Table 3 shows that both design features reduced the number of adjustments in the wrong direction—49.54% to 43.78% (DF1) and 46.11% to 41.11% (DF2) respectively. Interestingly, the analysis indicates another effect: On the one hand, even though access to the explanation component (DF1) increased the number of adjustments (35.16% to 63.14%), the overshoots were notably reduced

(26.61% to 14.05%). On the other hand, even though the uncertainty component (DF2) decreased the number of adjustments in the wrong direction (46.11% to 41.11%), it also increased the overshoots (8.38% to 16.56%).

To evaluate how agent transparency affects the task outcome, we analyzed the accuracy of AI forecast and the human forecasts (see Table 4). As outlined in Section 3.1, we used the mean absolute scaled error (MASE) as a measure of accuracy. A value of less than 1 shows that a forecast was better than the naive forecasting method. All forecasts exceeded this baseline. The collected data shows that during the control period—without access to the design features—the domain experts *decreased* the forecasts’ accuracy through their adjustments. More specifically, the error percentage—calculated between the MASE of both forecasts—*increased* by 4.19% (DF1) and 0.73% (DF2) respectively. However, after introducing the design features, the experts *improved* the forecasts’ accuracy and the error percentage was *reduced* by 4.37% (DF1) and 1.68% (DF2) respectively. Agent transparency improved the ability to adjust the forecasts (see Table 4).

**6 Discussion**

The results illustrate how agent transparency affects trust and task outcomes. Agent transparency does not only improves humans’ ability to adjust the forecast but also increases the forecast’s overall accuracy. However, the results indicate that *how* transparency is facilitated affects the domain experts’ trust

**Table 4** Analysis of the forecast accuracy (Gap 2 & 3)

	Explanation (DF1)		Uncertainty (DF2)	
	Control	Treatment	Control	Treatment
Accuracy of the forecasts				
– AI forecast [MASE]	0.5238	0.5688	0.3750	0.4246
– Human forecast [MASE]	0.5458	0.5440	0.3778	0.4174
Error percentage reduction	– 4.19%	+ 4.37%	– 0.73%	+ 1.68%

in the system differently (cf. DF1 and DF2). In this section, we discuss the results, interpret our findings, and relate them to current research.

### 6.1 Effects of the Design Features

Both design features increase agent transparency. The first design feature accomplishes this by making the AI's reasoning more transparent. In contrast, the second design feature creates more transparency as to the uncertainty associated with the AI's forecast. The explanation component (DF1) leads to the hypothesized outcomes. Transparency increased the trust in the system as well as the collaborative task outcome. Wang et al. (2016b) similarly report that, in the context of recommendation agents, explanations increase cognition-based trust. We expected the same results with the second design feature (DF2), as we assumed that providing domain experts with additional information on the uncertainty associated with the AI's forecast would similarly increase their trust in the system and the overall task outcome. However, while the transparency improved the judgmental forecast adjustments, it counter-intuitively reduced the domain expert's level of trust. Even though the trust in the system's technical competence remained mostly unaffected, the domain expert reported that he could not rely on the system to function correctly. However, objectively, the provided information allowed him to better utilize his domain knowledge to adjust the forecast. These results are indicative of a phenomenon called *algorithmic aversion*. As noted by Dietvorst et al. (2018), the "reluctance to use algorithms for forecasting [often] stems from an intolerance of inevitable error" (p. 1156). We expected that the AI's acknowledgment of its limitations would increase the domain expert's trust as it provides the opportunity for suitable forecast adjustments. Instead, the additional information caused the expert to recognize the AI's forecast as imperfect, which had a negative effect on his level of trust. Specifically, the domain expert reported less confidence in the system's reliability after the treatment period.

### 6.2 Trust and Human-AI Collaboration

We show that there is a non-trivial relationship between trust and control. Traditionally, the general understanding is that "when there is control there is no trust, and vice versa" (Castelfranchi and Falcone, 2000, p. 799). However, Castelfranchi & Falcone outline that "trust and control are [not] antagonistic (one eliminates the other) but complementary". They argue that while trust certainly is "needed for relying on the action of another agent" it can also "increase the trust in [an agent], making [the agent] more willing or more effective" (Castelfranchi & Falcone, 2000, p. 821). Their idea that control can build trust is reflected in our results. We argue

that human-AI collaboration requires trust (i.e., granting the AI a certain autonomy) as well as control (i.e., contributing to the joint task outcome).

Trust is a determining factor of effective collaboration (McAlister, 1995). Castelfranchi & Falcone (2000, p. 804) argue that the subjective perception of trust is, among other factors, affected by incomplete information. Given that there will always be a gap between the human's and the AI's model, trust essentially "requires modeling the mind of the other". This assessment aligns with the first gap described in the 3-Gap framework (Kayande et al., 2009) and is also reflected in the collected data (EE3). Our results, therefore, indicate that transparency reduces this gap. However, AI systems are also not necessarily equally competent across all functions (Muir, 1987). Therefore, humans must learn to calibrate their trust to avoid overriding competent functions or failing to override incompetent ones. This is reflected in the behavior of the domain experts (see Table 3 on page 18). Interestingly, the domain experts systematically increased the AI's forecast. Fildes et al. (2007, p. 6) argue that humans exhibit a "bias towards optimism". In the studied application context, this is reflected in the fact that the domain experts can't resist being overly optimistic regarding the revenue of their restaurants. This aligns with Mathews & Diamantopoulous's (1990) comment that bias can be introduced through judgmental forecast adjustments. Given that the algorithm does not characteristically underestimate the revenue, this strategy can only be explained by incomplete information. This behavior also indicates an under-reliance on the AI (Parasuraman & Riley, 1997).

### 6.3 Task Outcome and Human-AI Collaboration

The question remains whether information systems should be made more transparent to enable effective human-AI collaboration. Research shows that allowing humans to adjust forecasts creates a feeling of ownership and credibility (Hyndman & Athanasopoulos, 2018). However, in the control period, the domain experts made the forecast less accurate without additional information. This indicates that human-AI collaboration is not inherently beneficial, as it can negatively affect the task outcome (see Table 4 on page 19). In the control period, the error percentage *increased* by 4.19% (DF1) and 0.73% (DF2) respectively. This effect aligns with the work of several authors who have shown that "people's attempts to adjust algorithmic forecasts often make them worse" (Dietvorst et al., 2018, p. 1156). However, in our study, we observed a notable improvement in the task outcome after the domain experts used the more transparent systems. By utilizing the design features, the domain experts now *reduced* the error percentage by 4.37% (DF1) and 1.68% (DF2) respectively. The magnitude of our results aligns with the work of Petropoulos et al. (2016). They report that judgmental adjustments increased the "overall forecasting performance by 4 percent compared to the current practice" (Petropoulos et al., 2016, p. 850). We, therefore, argue that agent



transparency helped the domain experts improve the task outcome by better utilizing their domain knowledge. Our empirical data confirms this notion. Making the information system more transparent allowed the domain experts to better collaborate with the AI for two reasons: The explanation component allowed the AI's reasoning to be considered before adjustments were made. In contrast, communicating the uncertainty associated with the AI's forecast allowed the AI's lack of knowledge to be compensated where necessary.

#### 6.4 Contributions

On a theoretical level, we derive empirically grounded requirements and principles for the design of AI-based information systems (Buxmann et al., 2021; Maedche et al., 2019) by utilizing expert interviews and a focus group. On that basis, we conduct case studies with two restaurants to gain empirical evidence on the influence of agent transparency on trust and task outcomes. In line with Avital et al. (2019), we can confirm that designing systems for human-AI collaboration requires scholars and practitioners “to expand their considerations beyond performance and to address micro-level issues such as [...] trust” (Avital et al., 2019, p. 2). To support this endeavor, we show that agent transparency is vital for trustful and accurate collaboration between humans and AI systems. In the control phase, collaboration had a negative impact on task outcomes, suggesting that the AI should act autonomously to achieve the best task outcome. However, by facilitating agent transparency, the design features reversed this effect. Hereafter, the collaboration between humans and the AI notably improved the task outcomes. Our findings highlight that further research on the design of human-AI collaboration is needed.

The generalizability of our work is based on theoretical inference that draws from the conducted qualitative and quantitative research (Tsang, 2014; Gomm et al., 2000). We predominantly develop the presented explanations based on the relationships between variables observed in the case studies (c.f., Section 6). Lastly, the validity of the studied constructs (c.f., Section 3.3) underlines the rigor of the conducted case studies (Tsang, 2014; Lee and Baskerville, 2003). Consequently, we postulate that our empirically grounded requirement and principles provide a nascent design theory for the role of agent transparency in human-AI collaboration (Jones & Gregor, 2007). This design knowledge might be beneficial for designing AI systems that traditionally require a high level of human monitoring, such as high-stakes decision-making in the medical or judicial domain.

Regarding managerial implications, we show that enhancing business processes through machine learning requires organizations to look beyond task outcomes. The need for collaboration between humans and AI systems will increase

in the coming years. Accordingly, the quality of the designed AI systems will greatly shape the future of work. We have shown that designing transparency requires an intentional approach and that it is vital to provide the right level of agent transparency to ensure a positive effect on trust and task outcomes. It might be worthwhile for companies to provide personalized explanations—as a means to create transparency—based on the characteristics of the individual users. Additionally, we show that trust depends not only on the achieved task outcome but also on a variety of softer factors. Our results provide empirical evidence that a human-centered approach is required to design trustworthy AI systems (Harper, 2019). With the increasing adoption of AI systems in many industries, our work provides directions for designing and managing human-AI interaction.

## 7 Conclusion

In this work, we explore how agent transparency affects human-AI collaboration. Specifically, we study its effects on trust and task outcomes. Our results highlight the importance of transparency, yet, more research is required on how information systems should be designed to ensure effective collaboration (Seeber et al., 2019). We argue that collaboration requires information systems to be designed with both collaborators' cognitive skills and limitations in mind. Our work provides multiple contributions. First, through expert interviews and a focus group, we induce design knowledge (i.e., requirements and principles) applicable to AI-based information systems. Second, we explore the importance of transparency in a naturalistic environment through two case studies. More precisely, we demonstrate that agent transparency can have a mixed effect on trust depending on its realization.

### 7.1 Limitations

Despite the outlined contributions, this research is not without limitations. The naturalistic evaluation of the artifacts provides relevant insights “concerning the artefact's effectiveness in real use” (Venable et al., 2016, p. 81) and empirical results with high internal validity. However, even though we tried to control apparent confounding variables, this approach limits the research's external validity (Gummesson, 2000). Future design episodes are required to strengthen the external validity of the observed phenomena. Another limitation of this work is that we only examined the capability of humans to adjust the AI's forecast. Because of this, the domain experts' accuracy is intrinsically linked to the accuracy of the AI's forecast. Lim and O'Connor (1995) explicitly show that the effectiveness of forecast adjustments depends on the reliability of the forecast. We

could, therefore, not evaluate how accurate the domain experts would have been without the provided forecast. Additionally, we do not know how increased trust and maturation effects will affect the domain experts' accuracy in the long run. Academic literature warns of over-reliance on technological artifacts and the consequences thereof (Parasuraman & Riley, 1997). Lastly, we have evaluated the design features individually. Subsequent design cycles should examine how simultaneous access to both design features would influence the trust of the domain experts as well as the task outcome.

## 7.2 Implications for Future Research and Outlook

The empirical results and the limitations presented in this work provide a basis for further research. Agent transparency is an important facilitator of collaboration between humans and AI systems as it provides users with the information needed to intervene before the realization of a task. One promising direction for further research is bidirectional feedback (Kayande et al., 2009). So far, we have only examined agent transparency as a means for the AI to provide information to humans. However, domain experts could equally offer information on their reasoning to the AI. Explanations can also be used to improve the AI's decision model (Martens & Provost, 2017). Integrating additional feedback mechanisms is likely to have a positive influence on the effectiveness of human-AI collaboration (Zschech et al., 2021). Domain experts should also receive information on the correctness of their adjustments. While our results show that agent transparency is important for collaboration between humans and AI systems, more research is needed to understand how different types of transparency (Chen et al., 2014) affect the effectiveness of human-AI collaboration. Researchers need to explore additional means to increase transparency. In the last few years, numerous technological advances, such as, feature attribution explanations (Lundberg & Lee, 2017), counterfactual explanations (Mothilal et al., 2020), and prescriptive trees (Bertsimas et al., 2019) have been developed. While these approaches are impressive from a technical perspective, little research has been conducted to understand how they affect decision-making in the short and long term, particularly the behavior and perceptions of users. Furthermore, additional research is needed to understand how users' cognitive styles affect their perception and utilization of explanations. On a meta-level, our work motivates further research regarding the "future of work" in settings where AI systems and humans collaboratively work on tasks (Jarrahi, 2018). Designing AI systems that interact with humans is a non-trivial task. As this work shows, concepts like trust and transparency need to be investigated to allow for an effective, calibrated, and reliable symbiosis of humans and AI.

## Appendix A: Questionnaire for Trust Measurement

The following psychometric instrument, proposed by Madsen and Gregor (2000), was used to measure the domain experts' level of trust during the third evaluation episode (EE3).

### 1. Perceived reliability

- R1: The system always provides the advice I require to make my decision.
- R2: The system performs reliably.
- R3: The system responds the same way under the same conditions at different times.
- R4: I can rely on the system to function properly.
- R5: The system analyzes problems consistently.

### 2. Perceived technical competence

- T1: The system uses appropriate methods to reach decisions.
- T2: The system has sound knowledge about this type of problem built into it.
- T3: The advice the system produces is as good as that which a highly competent person could produce.
- T4: The system correctly uses the information I enter.
- T5: The system makes use of all the knowledge and information available to it to produce its solution to the problem.

### 3. Perceived understandability

- U1: I know what will happen the next time I use the system because I understand how it behaves.
- U2: I understand how the system will assist me with decisions I have to make.
- U3: Although I may not know exactly how the system works, I know how to use it to make decisions about the problem.
- U4: It is easy to follow what the system does.
- U5: I recognize what I should do to get the advice I need from the system the next time I use it.

### 4. Faith

- F1: I believe advice from the system even when I don't know for certain that it is correct.
- F2: When I am uncertain about a decision I believe the system rather than myself.

- F3: If I am not sure about a decision, I have faith that the system will provide the best solution.
- F4: When the system gives unusual advice I am confident that the advice is correct.
- F5: Even if I have no reason to expect the system will be able to solve a difficult problem, I still feel certain that it will.

**5. Personal attachment**

- P1: I would feel a sense of loss if the system was unavailable and I could no longer use it.
- P2: I feel a sense of attachment to using the system.
- P3: I find the system suitable to my style of decision-making.
- P4: I like using the system for decision-making.
- P5: I have a personal preference for making decisions with the system.

**Appendix B: Overview of Restaurant Chains and Interviewees**

Table 5 summarizes the sixteen interviews conducted with personnel planners from different restaurant chains during the first evaluation episode (EE1).

**Table 5** Overview of restaurant chains and interviewees

	Distribution	Revenue	Locations	Interviewees
I	Holding	< € 50M	42	<i>Alpha, Beta, Gamma</i>
II	Franchise	< € 50M	115	<i>Delta, Epsilon, Zeta, Eta, Theta, Iota</i>
III	Holding	< € 5M	3	<i>Kappa</i>
IV	Holding	< € 120M	40	<i>Lambda, Mu, Nu, Xi</i>
V	Franchise	–	34	<i>Gamma, Delta</i>

**Appendix C: Questionnaire Results**

Table 6 summarizes the answers the domain experts provided after the control period as well as the treatment period (i.e., after using the respective design features). The answers are provided on a seven-point Likert scale.

**Table 6** Questionnaire results on a seven-point Likert scale

	Explanation (DF1)		Uncertainty (DF2)	
	Control	Treatment	Control	Treatment
<b>Perceived reliability</b>				
– R1	5	5	3	3
– R2	5	5	5	3
– R3	5	5	3	4
– R4	5	5	4	3
– R5	3	5	4	2
<b>Perceived tech. competence</b>				
– T1	3	6	5	5
– T2	4	6	3	4
– T3	4	5	2	1
– T4	4	6	5	5
– T5	4	5	6	5
<b>Perceived understandability</b>				
– U1	5	6	6	6
– U2	6	6	6	5
– U3	3	2	6	1
– U4	6	6	6	6
– U5	4	5	5	5
<b>Faith</b>				
– F1	5	4	2	2
– F2	3	3	1	1
– F3	3	4	5	4
– F4	3	4	2	1
– F5	4	4	3	3
<b>Personal attachment</b>				
– P1	5	6	4	3
– P2	5	5	2	2
– P3	5	5	5	4
– P4	5	5	3	3
– P5	5	5	2	2

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of Interests** The authors declare that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abdul, A., Vermeulen, J., Wang, D., Lim, B.Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–18).
- Ardila, D., Kiraly, A.P., Bharadwaj, S., Choi, B., Reicher, J.J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D.P., & Shrivaya, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, 25, 954–961.
- Armstrong, J. (2007). Scott Significance tests harm progress in forecasting. *International Journal of Forecasting*, 23(2), 321–327.
- Avital, M., Hevner, A., & Schwartz, D. (2019). It takes two to Tango: Choreographing the interactions between human and artificial intelligence. In *Proceedings of the 27th european conference on information systems (ECIS)*. Stockholm & Uppsala, Sweden.
- Benyon, D. (2014). *Designing Interactive Systems: A comprehensive guide to HCI, UX and interaction design*. United Kingdom: Pearson, 3rd edition.
- Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *Management Information Systems Quarterly*, 45(3).
- Bertsimas, D., Dunn, J., & Mundru, N. (2019). Optimal prescriptive trees. *INFORMS Journal on Optimization*, 1(2), 164–183.
- Blattberg, R.C., & Hoch, S.J. (2008). Database models and managerial intuition: 50% Model + 50% Manager. *Management Science*, 36(8), 887–899.
- Buxmann, P., Hess, T., & Thatcher, J.B. (2021). AI-based information systems. *Business & Information Systems Engineering*, 63(1), 1–4.
- Castelfranchi, C., & Falcone, R. (2000). Trust and control: A dialectic link. *Applied Artificial Intelligence*, 14(8), 799–823.
- Chen, J.Y.C., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M.J. (2014). Situation awareness-based agent transparency. US Army Research Laboratory, pp. 1–29.
- Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, 60.
- Crouser, R.J., & Chang, R. (2012). An affordance-based framework for human computation and human-computer collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2859–2868.
- Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J.M. (2019a). Hybrid intelligence. *Business & Information Systems Engineering*, 61(5), 637–643.
- Dellermann, D., Lipusch, N., Ebel, P., & Leimeister, J.M. (2019b). Design principles for a hybrid intelligence decision support system for business model validation. *Electronic Markets*, 29(3), 423–441.
- Dietvorst, B.J., Simmons, J.P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3), 1155–1170.
- Drury, J.L., Scholtz, J., & Yanco, H.A. (2003). Awareness in human-robot interactions. In *Proceedings of the 2003 IEEE international conference on systems, man and cybernetics* (pp. 912–918). IEEE.
- Einhorn, H.J. (1974). Expert judgment: Some necessary conditions and an example. *Journal of Applied Psychology*, 59(5), 562.
- Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32–64.
- Fildes, R., & Goodwin, P. (2013). Forecasting support systems: What we know, what we need to know. *International Journal of Forecasting*, 29(2), 290–294.
- Fildes, R., Goodwin, P., & et al. (2007). Good and bad judgement in forecasting: Lessons from four companies. *Foresight*, 8, 5–10.
- Fleischmann, K.R., & Wallace, W.A. (2005). A covenant with transparency: Opening the black box of models. *Communications of the ACM*, 48(5), 93–97.
- Giboney, J.S., Brown, S.A., Lowry, P.B., & Nunamaker, J.F. (2015a). User acceptance of knowledge-based system recommendations: Explanations, arguments, and fit. *Decision Support Systems*, 72, 1–10.
- Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In *Proceedings of the 5th international conference on data science and advanced analytics (DSAA)*.
- Gomm, R., Hammersley, M., & Foster, P. (2000). Case study and generalization. In R. Gomm, M. Hammersley, & P. Foster (Eds.) *Case study method* (pp. 98–115).
- Gregor, S., & Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly*, 23(4), 497–530.
- Gregor, S., & Hevner, A.R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2), 337–355.
- Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, 29(2), 101614.
- Gummesson, E. (2000). *Qualitative methods in management research*. Sage.
- Wilson, H.J., Daugherty, P.R., Wilson, J., & Daugherty, P. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, 96(4), 114–123.
- Harper, R.H.R. (2019). The role of HCI in the Age of AI. *International Journal of Human-Computer Interaction*, 35(15), 1331–1344.
- Patrick, H., Schemmer, M., Vössing, M., & Kühl, N. (2021). Human-AI complementarity in hybrid intelligence systems: A structured literature review. In *PACIS 2021 Proceedings*.
- Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, 105, 105–120.



- Hevner, A.R., March, S.T., Park, J., Ram, S., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- Holzinger, A. (2016). Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics*, 3(2), 119–131.
- Hosseini, M., Shahri, A., Phalp, K., & Ali, R. (2017). Engineering transparency requirements: A modelling and analysis framework. *Information Systems*, 74(1), 3–22.
- Hosseini, M., Shahri, A., Phalp, K., & Ali, R. (2018). Four reference models for transparency requirements in information systems. *Requirements Engineering*, 23(2), 251–275.
- Hyndman, R.J. (2006). Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, 4(4), 43–46.
- Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Hyndman, R.J., & Koehler, A.B. (2005). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679–688.
- Peter, J. (1986). Introduction to expert systems.
- Jarke, M., Loucopoulos, P., Lyytinen, K., Mylopoulos, J., & Robinson, W. (2011). The brave new world of design requirements. *Information Systems*, 36(7), 992–1008.
- Jarrahi, M.H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586.
- Jensen, M.L., Lowry, P.B., Burgoon, J.K., & Nunamaker, J.F. (2010). Technology dominance in complex decision making: The case of aided credibility assessment. *Journal of Management Information Systems*, 27(1), 175–202.
- Jones, D., & Gregor, S. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems*, 8(5), 1.
- Kagermann, H. (2015). Change through digitization—value creation in the age of Industry 4.0. In A. Horst, H. Meffert, A. Pinkwart, & R. Reichwald (Eds.) *Management of permanent change* (pp. 23–33). Springer Fachmedien Wiesbaden, Wiesbaden.
- Kamar, E. (2016). Directions in hybrid intelligence: Complementing AI systems with human intelligence. In *IJCAI international joint conference on artificial intelligence* (pp. 4070–4073).
- Kayande, U., De Bruyn, A., Lilien, G.L., Rangaswamy, A., & Van Bruggen, G.H. (2009). How incorporating feedback mechanisms in a DSS affects DSS evaluations. *Information Systems Research*, 20(4), 527–546.
- Ke, G., Qi, M., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
- Kiouris, S. (2002). Interactivity: a concept explication. *New Media & Society*, 4(3), 355–383.
- Der Kiureghian, A., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2), 105–112.
- Kizilcec, R.F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 2390–2395).
- Kolbjornsrud, V., Amico, R., & Thomas, R.J. (2016). The promise of artificial intelligence: Redefining management in the workforce of the future. Technical report, Accenture.
- Kurzweil, R. (1990). *The age of intelligent machines*. MIT Press.
- Lee, A.S., & Baskerville, R.L. (2003). Generalizing generalizability in information systems research. *Information Systems Research*, 14(3), 221–243.
- Lee, J.D., & See, K.A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors*, 46(1), 50–80.
- Leite, J.C.S.D.P., & Cappelli, C. (2010). Software transparency. *Business & Information Systems Engineering*, 2(3), 127–139.
- Lim, J.S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making*, 8(3), 149–168.
- Lundberg, S.M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.) *Advances in neural information processing systems 30* (pp. 4765–4774). Curran Associates, Inc.
- Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. *Proceedings of 11th Australasian Conference on Information Systems*, 53, 6–8.
- Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-based digital assistants. *Business & Information Systems Engineering*, 61(4), 535–544.
- Maedche, A., Morana, S., Schacht, S., Werth, D., & Krumeich, J. (2016). Advanced user assistance systems. *Business & Information Systems Engineering*, 58(5), 367–370.
- Manyika, J. (2017). A future that works: automation, employment and productivity. Technical report, McKinsey Global Institute.
- Martens, D., & Provost, F. (2017). Explaining data-driven document classifications. *MIS Quarterly*, 38(1), 73–99.
- Mathews, B.P., & Diamantopoulos, A. (1990). Judgemental revision of sales forecasts: Effectiveness of forecast selection. *Journal of Forecasting*, 9(4), 407–415.
- Mcallister, D.J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *The Academy of Management Journal*, 38(1), 24–59.
- Mcknight, D., Carter, M., Thatcher, J.B., & Clay, P.F. (2014). Trust in a specific technology. *ACM Transactions on Management Information Systems*, 2(2), 1–25.
- Christian, M., & Bunde, E. (2020). Transparency and trust in Human-AI-interaction: The role of model-agnostic explanations in computer vision-based decision support. In D. Helmut L. Reinerman-Jones (Eds.) *Artificial intelligence in HCI* (pp. 54–69). Cham: Springer International Publishing.
- Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2020). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. *Information Systems Management*, pp 1–11.
- Morana, S., Pfeiffer, J., & Adam, M.T.P. (2018). Call for papers, Issue 3/2020, user assistance for intelligent systems. *Business & Information Systems Engineering*, 60(6), 571–572.
- Mothilal, R.K., Sharma, A., & Tan, C. (2020). Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 607–617).
- Muir, B.M. (1987). Trust between humans and machines. *International Journal of Man-Machine Studies*, 27, 327–339.
- Muir, B.M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11), 1905–1922.
- Ford, N.F. (1985). Decision support systems and expert systems: A comparison. *Information & Management*, 8(1), 21–26.
- Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3-5), 393–444.
- O'Donovan, J., & Smyth, B. (2005). Trust in recommender systems. In *Proceedings of the 10th international conference on Intelligent user interfaces* (pp. 167–174). ACM.
- Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation

- research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, and disuse. *Human Factors*, 39(2), 230–253.
- Patton, M.Q. (2002). *Qualitative research and evaluation methods*, vol. 3rd. Sage Publications.
- Petropoulos, F., Fildes, R., & Goodwin, P. (2016). Do ‘big losses’ in judgmental adjustments to statistical forecasts affect experts’ behaviour? *European Journal of Operational Research*, 249(3), 842–852.
- Phillips-Wren, G. (2013). Intelligent decision support systems. *Multicriteria Decision Aid and Artificial Intelligence: Links, Theory and Applications*, pp. 25–43.
- Pieters, W. (2011). Explanation and trust: What to tell the user in security and AI? *Ethics and Information Technology*, 13(1), 53–64.
- Power, D.J. (2002). *Decision support systems: Concepts and resources for managers*. Quorum Books.
- Power, D.J., Heavin, C., & Keenan, P. (2019). Decision systems redux. *Journal of Decision Systems*, 28(1), 1–18.
- Rai, A., Constantinides, P., & Sarker, S. (2019). Next-generation digital platforms: Toward Human–AI hybrids. *MIS Quarterly*, 43(1), iii–x.
- Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson Education Limited, pp. 1151.
- Seeber, I., Bittner, E., Briggs, R.O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2019). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*.
- Shaw, R.B. (1997). *Trust in the balance: Building successful organizations on results, integrity, and concern*. Jossey-Bass.
- Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal*, 31(2).
- Söllner, M., Benbasat, I., Gefen, D., Leimeister, J.M., & Pavlou, P.A. (2016a). Trust. In A. Bush A. Rai (Eds.) *MIS quarterly research curations*.
- Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., & Leimeister, J.M. (2012). Understanding the formation of trust in IT artifacts. In J.F. George (Ed.) *Proceedings of the international conference on information systems ICIS 2012*.
- Söllner, M., Hoffmann, A., & Leimeister, J.M. (2016b). Why different trust relationships matter for information systems users. *European Journal of Information Systems*, 25(3), 274–287.
- Sonnenberg, C., & Brocke, J.V. (2012). Evaluations in the science of the artificial – reconsidering the build-evaluate pattern in design science research. In K. Pfeffers, M.A. Rothenberger, & B. Kuechler (Eds.) *Design science research in information systems. Advances in theory and practice. DESRIST 2012. Lecture Notes in Computer Science* (pp. 381–397). Springer Berlin Heidelberg.
- Stowers, K., Kasdaglis, N., Newton, O., Lakhmani, S., Wohleber, R., & Chen, J. (2016). Intelligent agent transparency: The design and evaluation of an interface to facilitate human and intelligent agent collaboration. In *Proceedings of the human factors and ergonomics society* (pp. 1704–1708).
- Street, C.T., & Meister, D.B. (2004). Small business growth and internal transparency: The role of information systems. *MIS Quarterly*, pp 473–506.
- Taylor, S.J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45.
- Terveen, L.G. (1995). Overview of human-computer collaboration. *Knowledge-Based Systems*, 8 (2-3), 67–81.
- Tremblay, M.C., Hevner, A.R., & Berndt, D.J. (2010). The use of focus groups in design science research. In H. Alan S. Chatterjee (Eds.) *Design research in information systems* (pp. 121–143). Boston, MA: Springer.
- Tsang, E.W.K. (2014). Generalizing from research findings: The merits of case studies. *International Journal of Management Reviews*, 16(4), 369–383.
- Tschannen-Moran, M. (2001). Collaboration and the need for trust. *Journal of Educational Administration*, 39(4), 308–331.
- Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems*, 25(1), 77–89.
- vom Brocke, J., & Maedche, A. (2019). The DSR grid: six core dimensions for effectively planning and communicating design science research projects. *Electronic Markets*, 29(3), 379–385.
- Vössing, M., Potthoff, F., Kühl, N., & Satzger, G. (2019). Designing useful transparency to improve process performance—evidence from an automated production line. In *Proceedings of the 27th european conference on information systems (ECIS)*.
- Wang, N., Pynadath, D.V., & Hill, S.G. (2016a). Trust calibration within a human-robot team: Comparing automatically generated explanations. In *The Eleventh ACM/IEEE international conference on human robot interaction* (pp. 109–116). IEEE Press.
- Wang, W., & Benbasat, I. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, 6(3), 72–101.
- Wang, W., & Benbasat, I. (2008). Attributions of trust in decision support technologies: A study of recommendation agents for e-commerce. *Journal of Management Information Systems*, 24(4), 249–273.
- Wang, W., Qiu, L., Kim, D., & Benbasat, I. (2016b). Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. *Decision Support Systems*, 86, 48–60.
- Watson, H.J. (2017). Preparing for the cognitive generation of decision support. *MIS Quarterly Executive*, 16(3).
- Zschech, P., Walk, J., Heinrich, K., Vössing, M., & Kühl, N. (2021). A picture is worth a collaboration: Accumulating design knowledge for computer-vision-based hybrid intelligence systems. In *29th european conference on information systems (ECIS 2021), June 14-16, 2021*.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Michael Vössing** is a research group lead at the Karlsruhe Service Research Institute (KSRI) and the Institute of Information Systems and Marketing (IISM) at the Karlsruhe Institute of Technology (KIT). He is studying how information systems should be designed to support effective collaboration between humans and artificial intelligence (AI). Currently, he and his team are developing tools for human-AI collaboration in multiple domains—most recently in the Architectural Engineering and Construction (AEC) industry. He is internationally collaborating with multiple institutions like ETH Zürich, Technion - Israel Institute of Technology, and West Virginia University.

**Niklas Kühl** is head of the Applied AI in Services Lab at the Karlsruhe Service Research Institute (KSRI) and the Institute of Information Systems and Marketing (IISM) at the Karlsruhe Institute of Technology (KIT). He has been working on machine learning (ML) and artificial intelligence (AI) in different domains since 2014. He is working on conceptualizing, designing and implementing AI in Service Systems with a focus on inter-organizational learning as well as the fair and effective interplay between users and intelligent agents. Currently, he and his team are actively working on different ML & AI solutions within industrial services, sales forecasting, production lines or fairness. He is internationally collaborating with multiple institutions like the University of Texas and the MIT-IBM Watson AI Lab.

**Matteo Lind** holds a Master of Science (M.Sc.) in Industrial Engineering and Management from the Karlsruhe Institute of Technology (KIT). Since 2019 he has been working on machine learning (ML) and artificial intelligence (AI) in the area of demand forecasting and order fulfillment. His work includes data analysis, model tuning, proof-of-concept development, and prototyping. He is currently working as a Senior Data Science Consultant at Blue Yonder.

**Gerhard Satzger** is Director of the Karlsruhe Service Research Institute (KSRI) and the Institute of Information Systems and Marketing (IISM) at the Karlsruhe Institute of Technology (KIT). His research

interests are the use of data and artificial intelligence to conceive, develop, and transform service business models. Before joining KSRI, he has gathered multi-year industry experience in various national and international roles within IBM, among others as head of their European Chief Analytics Office and as CFO of IBM's technology service business in Central Europe. He has obtained a formal postdoctoral lecturer qualification from the University of Augsburg, a Ph.D. in Information Systems from the University of Giessen and an MBA from Oregon State University/USA, after graduating from the University of Karlsruhe with a degree in Industrial Engineering and Management.