European Journal of Information Systems

ISSN: 0960-085X (Print) 1476-9344 (Online) Journal homepage: www.tandfonline.com/journals/tjis20

Tell me more, tell me more: the impact of explanations on learning from feedback provided by Artificial Intelligence

Maximilian Förster, Hanna R. Broder, Marie C. Fahr, Mathias Klier & Lior Fink

To cite this article: Maximilian Förster, Hanna R. Broder, Marie C. Fahr, Mathias Klier & Lior Fink (2025) Tell me more, tell me more: the impact of explanations on learning from feedback provided by Artificial Intelligence, European Journal of Information Systems, 34:2, 323-345, DOI: 10.1080/0960085X.2024.2404028

To link to this article: https://doi.org/10.1080/0960085X.2024.2404028

EUROPEAN JOURNAL OF INFORMATION SYSTEMS
2025, VOL. 34, NO. 2, 323–345
https://doi.org/10.1080/0960085X.2024.2404028

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Tell me more, tell me more: the impact of explanations on learning from feedback provided by Artificial Intelligence

Maximilian Förster[a], Hanna R. Broder[a], Marie C. Fahr[a], Mathias Klier[a] and Lior Fink[b]

[a]Institute of Business Analytics, University of Ulm, Ulm, Germany; [b]Department of Industrial Engineering & Management, Ben-Gurion University of the Negev, Beer Sheva, Israel

ABSTRACT

Whereas learning is one of the primary goals of Explainable Artificial Intelligence (XAI), we know little about whether, how, and when explanations enhance users' learning from feedback provided by Artificial Intelligence (AI). Drawing on *Feedback Theory* as a fundamental theoretical lens, we formulate a research model wherein explanations enhance informativeness and task performance, contingent on users' prior knowledge, ultimately leading to a higher learning outcome. This research model is tested in a randomized between-subjects online experiment with 573 participants whose task is to match Google Street View pictures to their city of origin. We find a positive effect of explanations on learning outcome, which is fully mediated by informativeness, for users with less prior knowledge. Furthermore, we find that explanations positively impact users' task performance, where this effect is direct for more knowledgeable users and fully mediated by informativeness for less knowledgeable users. We seek to elucidate the mechanisms underlying these effects of explanations on learning from AI feedback in focus groups with AI experts and users. By studying the consequences of explanations as part of AI feedback for users in non-routine inference tasks, we advance the understanding of explanations as facilitators of human learning from AI systems.

ARTICLE HISTORY
Received 14 November 2022
Accepted 9 September 2024

KEYWORDS
Explainable Artificial Intelligence; XAI; AI feedback; learning outcome; informativeness; feedback theory

1. Introduction

Explainable Artificial Intelligence (XAI) serves to improve human interaction with Artificial Intelligence (AI) by automatically providing explanations alongside AI decisions (Adadi & Berrada, 2018). Among different goals of XAI, including the evaluation, improvement, and justification of AI systems, an important goal is to help users learn from AI systems (Meske et al., 2022). While learning from AI systems can take place in various contexts and tasks, the potential of AI systems to enhance human learning is particularly significant for humans who perform non-routine inference tasks (e.g., Gavaz et al., 2021). This significant potential for human learning is rooted in the unique characteristics of AI learning (Jia et al., 2023). AI learns in inference tasks (i.e., tasks that have a right or wrong answer) through complex processes of data processing and analysis, which yield highly accurate decisions even for difficult tasks, thereby providing higher learning capacity than any other technology (Baird & Maruping, 2021; Nishant et al., 2023). By contrast, humans rely on the application of critical and creative thinking skills to overcome their lack of knowledge (Chong et al., 2018). Thus, AI can supplement the human creativity-driven learning approach by data-driven insights (Candelon et al., 2023; Jia et al., 2023; Nishant et al., 2023; Van den Broek et al., 2021). AI has the potential to advance human learning by improving users' cognitive skills (Jia et al., 2023), particularly when users engage in non-routine tasks about which they are less knowledgeable, implying that they have much to learn from the AI. However, a common barrier to such learning is the opacity of AI models, which prevents humans from understanding on which insights AI decisions are based (Berente et al., 2021). By revealing how AI models arrive at their decisions, XAI is promising to improve human learning from AI systems, particularly in non-routine inference tasks in which decision accuracy can significantly benefit from reliance on AI.

However, the ability of XAI to facilitate learning has received little research attention. Existing literature on the impact of XAI on human-AI interaction is mostly related to how explanations serve the purpose of *justifying* AI decisions. By contrast, the purpose of improving human *learning* from AI systems is poorly understood. Whereas explanations generally justify AI decisions before users make decisions, learning from AI systems may particularly occur during the assessment of user decisions, implying AI augmentation in the form of feedback (Leyer et al., 2020). AI feedback is the information a user receives from an AI system, which can be employed to alter the user's knowledge gap after they have conducted a task and carried out their own solution (Hattie & Timperley, 2007;

CONTACT Lior Fink finkl@bgu.ac.il
Supplemental data for this article can be accessed online at https://doi.org/10.1080/0960085X.2024.2404028.

Wongvorachan et al., 2022). Feedback in human learning processes allows humans to make sense of information to reduce the gap between their initial level of knowledge and a reference level (Wongvorachan et al., 2022). Whereas AI augmentation in the form of AI feedback has recently gained importance (Hall et al., 2022), the role of explanations as part of AI feedback has yet to be explored. It is an open question whether explanations alongside AI decisions enrich the information contained in AI feedback so that users can better learn from the feedback, particularly in non-routine inference tasks, wherein there is greater potential for such feedback to enhance users' learning.

Endeavouring to address this open question, the purpose of this paper is to examine the impact of providing automatically generated explanations alongside AI decisions in AI feedback on users' learning. We position our research in the field of AI augmentation of human decision-making, which covers situations in which users perform a non-routine inference task and are provided with AI feedback after having made a decision. Given this position, we draw on *Feedback Theory* to investigate the effect of automatically generated explanations on learning outcome in non-routine inference tasks, mediated through informativeness and task performance, which have been highlighted in the literature as important mechanisms for understanding the effect of feedback on learning. We further rely on Feedback Theory to shed light on the role of users' prior knowledge in moderating the effects of explanations. We test the resulting research model by conducting a randomized between-subjects online experiment with 573 participants wherein we manipulate the existence of automatically generated explanations. In the experiment, participants learn from feedback provided by a state-of-the-art AI model and local explanations produced by an XAI method for image classification. We find a positive effect of explanations on learning outcome, fully mediated by informativeness, for users with less prior knowledge, who are positioned to benefit the most from AI feedback given their larger knowledge gap. Furthermore, we find that explanations positively impact users' task performance, although this effect is mediated by informativeness for users with less prior knowledge. We conduct four focus groups with a total of 10 AI experts and 10 AI users to validate the results of the online experiment and to better understand the mediating role of informativeness and the moderating role of users' prior knowledge in the effect of explanations on learning from AI feedback.

Three important theoretical contributions arise from this study. First, we advance the understanding of how and when automatically generated explanations influence users' learning from AI feedback in non-routine inference tasks. Second, we contribute to the body of knowledge on Feedback Theory by applying this theory to further understanding of XAI, thereby demonstrating the insight that can be gained by conceptualizing AI systems and explanations as sources of feedback for users. Third, we view XAI through a behavioural lens, thus complementing the extensive algorithmic research on how to develop XAI with behavioural research on how XAI impacts human-AI interaction. We, consequently, address the call for theoretical and empirical contributions to sharpen the understanding of the importance of XAI for human-AI interaction (Bayer et al., 2021; Brasse et al., 2023). From a practical standpoint, our findings offer implications for organisations and developers that wish to leverage the potential of AI. The findings suggest that supplementary implementation of XAI methods alongside AI models can help users with less prior knowledge, such as novice employees, to learn from AI feedback, as well as help broader populations of employees and clients to improve their task performance.

We start by providing an overview of the theoretical background and related work. Then, we develop the research model, describe the methodological approach for testing our hypotheses, and present our data analysis and results. We next present the focus groups we conducted to validate and better understand our findings. Subsequently, we describe our key findings, implications for theory and practice, and future research directions. We conclude with a summary.

2. Theoretical background and related work

Learning can be modelled as a cycle with different stages in which learners change their cognitive states, and thus their performance, due to the provision of feedback (Bangert-Drowns et al., 1991; Maier & Klotz, 2022). Learners start from an initial level of knowledge. After receiving a certain task, they activate their stored search and retrieval strategies to find an appropriate solution based on their actual level of knowledge. Subsequently, they carry out this solution. This is followed by the opportunity to receive information about whether and why their solution was correct or not, i.e., feedback. This response gives learners the opportunity to evaluate their solution, possibly leading to a new level of knowledge (Lipnevich & Panadero, 2021).

AI augmentation of human decision-making in the assessment of results through feedback (Leyer et al., 2020) can enrich this learning cycle (Bangert-Drowns et al., 1991), as humans make sense of information to reduce their knowledge gap (Wongvorachan et al., 2022). We draw on the conceptualizations of Feedback Theory to define AI feedback as the information a user receives from an AI system, which can be employed to alter the user's knowledge gap after

they have conducted a task and carried out their own solution (Hattie & Timperley, 2007; Wongvorachan et al., 2022). An AI system comprises an AI model, which produces an AI decision, i.e., the AI system's solution for the task. The AI system's solution provides information on whether the user's solution is correct or not. Additionally, an AI system can feature an XAI method, which generates an explanation alongside the AI decision (Gunning & Aha, 2019). The explanation provides information on why the user's solution is correct or not, based on the AI decision. Thus, AI feedback necessarily comprises the AI decision and – if an XAI method is present – a corresponding explanation provided by the AI system.

AI feedback is particularly relevant to users who perform non-routine inference tasks. IS literature distinguishes between routine and non-routine tasks (e.g., Goodhue, 1995). To perform routine tasks, which are also referred to as "questions" or "exercises" (Krulik & Rudnick, 1993), users can apply known rules and procedures without the need to conduct reasoning (Chong et al., 2018). By contrast, non-routine tasks are perceived as more challenging with lower likelihood of successful completion (Elia et al., 2009). Compared to routine tasks, users conducting non-routine tasks are faced with a significant knowledge gap, where feedback can help reduce the knowledge gap and increase task performance (Gavaz et al., 2021). In addition, to perform non-routine tasks, users cannot rely on a structured and "well-rehearsed approach or pathway" (Woodward et al., 2012, p. 11), but require creative and original thinking (Beghetto, 2017). Compared to routine tasks, conducting non-routine tasks is closely connected to learning. AI feedback can augment human decision-making in preference tasks, as a question of personal preference that is evaluated against internal criteria, and inference tasks, defined by the presence of an external criterion for evaluating task performance (Luan et al., 2014). While AI feedback can be utilized to achieve a new level of knowledge for both preference and inference tasks, learning in preference tasks relies on the interpretation of users' subjective intents (Dodeja et al., 2024). By contrast, inference tasks are associated with well-defined, objective metrics with a ground truth, making them most amenable for mathematical and statistical modelling techniques such as AI (Hastie et al., 2009). In sum, feedback is particularly needed to augment human decision-making in non-routine tasks, and AI feedback appears particularly effective in inference tasks.

Feedback Theory explores the role of feedback in learning processes, with feedback being generally regarded as one of the most powerful influences on learning (Lipnevich & Panadero, 2021). Three dimensions influence the effectiveness of feedback: Conditions related to the instructional content, conditions related to the external feedback agent, and conditions related to the learner (Narciss, 2013). First, conditions related to the instructional content include the instructional domain and topic, the requirements of the learning task, and the instructional materials (Narciss, 2013). Often, these conditions are predefined, for example by a company's field and strategy in the context of corporate learning, or by curricula at schools or universities. Second, conditions related to the feedback agent include the agent's knowledge, credibility, and technical attributes (Narciss, 2013). For instance, researchers emphasize the importance of the agent's design, mode of feedback delivery, timeliness of feedback (Lipnevich et al., 2016), and informativeness of provided feedback (Goldin et al., 2017; Rai et al., 2002). One example are explanations, i.e., human-understandable reasoning for a respective input-output-mapping (Abdul et al., 2018). They serve to mitigate the black-box nature of AI systems by revealing to users how underlying AI models arrive at decisions (Abdul et al., 2018; Förster et al., 2020a). This goal is often achieved through post-hoc explainability, i.e., showing why the AI model has arrived at its decision through "after-the-fact rationale [. . .] for system outputs" (Keane & Kenny, 2019, p. 3). Finally, the effectiveness of feedback for learning is strongly influenced by the conditions of the learner, i.e., individual contextual factors. These factors can facilitate or constrain how well learners are able to improve their competencies toward the desired reference level when provided with feedback by an external agent (Lipnevich et al., 2016). Research on feedback and learning puts emphasis on individual conditions of the learner. Unlike conditions related to the instructional content and conditions related to the external feedback agent, conditions of the learner are hard to control, yet they have considerable impact on feedback effectiveness (Narciss, 2013). Various individual contextual factors, summarized in Table 1, have been highlighted in research on the effectiveness of feedback for learning. Prior knowledge has been a common thread in these studies – both in early pioneering papers (e.g., Ramaprasad, 1983) as well as in more recent work (e.g., Panadero et al., 2018).

In IS literature, feedback provided by AI has been investigated in the context of decision support systems (DSS), defined as "interactive computer-based systems, which help decision makers utilize data and models to solve unstructured problems" (Sprague, 1980, p. 1). The usage of AI in DSS serves to enhance their predictive capabilities (Malak et al., 2019; Tyler & Jacobs, 2020). Besides the role of DSS to assist users prior to making a decision, research has recognized the potential of DSS to offer feedback after a decision has been made (e.g., Te'eni, 1991). So far, a limited number of studies investigated feedback mechanisms in DSS. The pioneering papers by Stone (1995) and

Table 1. Individual contextual factors influencing the effectiveness of feedback.

Individual contextual factor(s)	Brief description of impact on effectiveness of feedback	Example
Knowledge gap	The knowledge gap influences how feedback affects learning. The purpose of feedback is to reduce the knowledge gap between the learner's current knowledge level and a reference level. For feedback to be effective, the knowledge gap needs to be considered.	Ramaprasad (1983)
Response certitude	Response certitude – the degree to which the learner expects their response to be a correct one – depends on prior knowledge and influences the effectiveness of feedback.	Kulhavy and Stock (1989)
Learner's initial state	Feedback is more effective when learners demonstrate mindfulness of feedback, which is influenced by the learners' initial state, i.e., factors, such as prior knowledge and interests.	Bangert-Drowns et al. (1991)
Individual differences	Individual differences, such as domain knowledge, motivational beliefs, and strategy knowledge, influence the effectiveness of feedback by shaping the interpretation of feedback.	Panadero et al. (2018)
Individual factors	Individual factors, such as task novelty, creativity, and self-esteem, moderate the effectiveness of feedback interventions.	Kluger and DeNisi (1996)
Learner factors	Learner factors, such as ability, prior success, and gender, are relevant for the effectiveness of feedback on users' learning.	Lipnevich et al. (2016)
Prior knowledge	Effectiveness of feedback depends on whether an appropriate type of feedback is selected in accordance with prior knowledge.	Mason and Bruning (2001)
Individual cognitive factors	Feedback should consider learners' objectives, prior level of knowledge and competencies, and prior level of motivation and vocational skills to be effective.	Narciss (2013)

Te'eni (1991) suggest that DSS feedback can positively influence users' learning and underpin the importance of the design of DSS feedback. Later studies pursued this fundamental idea and investigated what constitutes effective DSS feedback. Montzemi et al. (1996) conducted two experiments with inference tasks where DSS feedback served to help users correct their initial decisions. Their findings indicate that informative guidance is more helpful than suggestive guidance, implying that users need to not only know what they should do next but also understand why they are doing it. Chenoweth et al. (2004) investigated whether and how DSS feedback influences users' willingness to expend effort in decision-making. Their study demonstrates that users are willing to expend more effort if potential accuracy gains are made salient. Kayande et al. (2009) supplement these insights by showing that users will be more likely to accept DSS feedback when their own mental model, i.e., their understandings, reasonings, and predictions, are aligned with the model embedded in the DSS. Taken together, these studies suggest that DSS feedback should be interpretable for users and provide information beyond mere corrective suggestions. However, literature has not yet investigated the impact of explanations in DSS feedback. Furthermore, there is a lack of understanding on how to operationalize the value of DSS feedback from the users' perspectives.

Explanations in DSS feedback are even more relevant for DSS that build on AI models, because the black-box nature of AI models makes respective feedback less interpretable to users (Antoniadi et al., 2021; Schoonderwoerd et al., 2021). XAI methods can help render DSS feedback interpretable to users by automatically generating explanations along AI decisions (Meske et al., 2022). Recent XAI research emphasizes the need to explicitly study the isolated effect of explanations in DSS, without interference from other design characteristics (Nguyen et al., 2024). While empirical evidence on the impact of explanations on DSS users is generally rare (Brasse et al., 2023), existing studies focus on explanations in DSS that assist users prior to making a decision. In such settings, empirical insights shed light on the role of explanations for concepts like trust (e.g., Aechtner et al., 2022; Hamm et al., 2023), understanding of the AI system (e.g., Sieger et al., 2022; Van der Waa et al., 2021), perception of the AI system (e.g., Klein et al., 2023; Y.-F. Wang et al., 2023), and task performance (Van der Waa et al., 2021; Walter et al., 2023). There is a lack of understanding of how explanations affect learning, which is relevant when DSS provide feedback after users have made a decision.

3. Hypothesis development

While the IS literature acknowledges that algorithmic feedback should be interpretable for users to enable learning, the impact of explanations on learning from such feedback has yet to be investigated, despite the recognized importance of learning in IS use. Furthermore, the IS literature is mute on how to conceptualize and operationalize the value of such feedback for learning from the users' perspectives. In this study, we develop a research model to describe the effects of automatically generated explanations as part of AI feedback on users' learning outcome in non-routine inference tasks. Based on Feedback Theory as our theoretical lens, we explore how AI feedback can assist humans in reducing their knowledge gap (Wongvorachan et al., 2022), suggesting that prior knowledge is a key individual contextual factor that *moderates* the effects of explanations. Through the same theoretical lens, we consider informativeness as reflecting the value of feedback from the users' perspectives, suggesting that informativeness is a key condition of the feedback agent that *mediates* the effect of explanations on learning outcome. Table 2 summarises the key concepts we use in hypothesis development.

Table 2. Definition of concepts.

Concept	Definition
Explanations	The provision of automatically generated explanations (Abdul et al., 2018) as part of AI feedback
Informativeness	The user's perception of the value of information in AI feedback, i.e., relevance, usefulness, accuracy of as well as satisfaction with information provided by the AI system (Goldin et al., 2017; Rai et al., 2002)
Learning Outcome	The user's subjective evaluation of their learning with respect to their task derived from the usage of AI feedback (Li et al., 2019)
Task Performance	An objective measure of the user's task performance (Hall et al., 2022)
Prior Knowledge	The user's prior knowledge regarding the task (Bangert-Drowns et al., 1991; Bayer et al., 2021)

3.1. Explanations

For users receiving AI feedback in non-routine inference tasks, explanations are hypothesised to have a direct impact on the informativeness of AI feedback and on the user's task performance, which are considered as important antecedents of learning outcome according to Feedback Theory.

As elaborated above, the effectiveness of feedback is influenced by the value of feedback for the users. Informativeness describes this value – more precisely, the user's perception of the value of information in AI feedback, i.e., the relevance, usefulness, accuracy, and satisfaction with the information provided by the AI system (Goldin et al., 2017; Rai et al., 2002). According to Tempelaar et al. (2015), feedback is informative when two conditions are met: It must be predictive and must enable intervention. Because of its predictive capabilities (e.g., Adadi & Berrada, 2018), an AI model generating decisions can satisfy the first condition. Designing AI feedback in a way that enables intervention is more complex. Due to the black-box nature of AI models, AI decisions only reveal whether a solution is correct or incorrect – but not why (Meske et al., 2022). Automatically generated explanations alongside AI decisions uncover the reasoning of the underlying AI model (e.g., Adadi & Berrada, 2018), thereby revealing to a user why a particular solution is correct or incorrect. In addition, explanations can help to clarify ambiguous issues that may be raised by AI decisions and provide additional information of value. Therefore, if the AI system features an XAI method that generates explanations, AI feedback also allows for intervention, thereby boosting the perceived relevance of (e.g., Conati et al., 2021), usefulness of (e.g., Hamm et al., 2021), and satisfaction with (e.g., Conati et al., 2021) the information provided by the AI system. Consequently, explanations enrich AI feedback, as demonstrated by its positive impact on various characteristics of informativeness. We thus formulate the following hypothesis:

H1: Users receiving explanations in addition to AI decisions perceive the informativeness of AI feedback as higher than users receiving AI decisions only.

While informativeness represents the subjective direct effect of explanations, learning outcome is also likely to be contingent on the ability of explanations to improve objective performance. In many cases, the desired outcome of human-AI interaction is an increase in users' objective task performance (Brasse et al., 2023). AI is often used for decision support, where subsequent task performance depends on whether users trust and follow the AI system's solutions (Brasse et al., 2023). When investigating how automatically generated explanations impact the user's learning from AI feedback, task performance refers to the user's objective learning progress in subsequent tasks. In previous studies, interaction with AI systems has been found to enhance the user's task performance in e-learning contexts through both decision support and feedback mechanisms (Kabudi et al., 2021). It has been suggested that the provision of explanations alongside AI decisions as a decision support tool may even enhance the positive impact of AI systems on task performance by providing additional information on how a decision was derived, enabling users to better challenge AI decisions before carrying out their solution as part of the learning cycle (Schemmer et al., 2022). Also, in the context of AI systems as upfront advice, providing users not only with the AI decision but also with explanations has been shown to improve the user's performance (Van der Waa et al., 2021; Walter et al., 2023). This effect is attributed to the ability of explanations to reduce the cognitive load on users (Lai & Tan, 2019), leaving more capacity for cognitive activities. These positive implications of explanations in AI-based decision support are likely to similarly enhance users' task performance in AI feedback. As explanations support users throughout their learning cycle with more insights as a basis for intervention, we expect them to increase users' knowledge level, resulting in a higher probability of correct solutions in their next task. Thus, we expect AI feedback to increase task performance even more when it includes explanations in addition to AI decisions. Therefore, we postulate the following hypothesis:

H2: Users receiving explanations in addition to AI decisions have higher task performance than users receiving AI decisions only.

3.2. Informativeness

We hypothesise that informativeness, i.e., the user's perception of the value of the information provided by AI feedback, is positively related to two constructs in our research model, namely task performance and learning outcome. First, we expect higher informativeness to enhance users' task performance. Feedback Theory suggests that providing feedback to users can challenge their initial solutions, possibly leading to a new level of knowledge (Bangert-Drowns et al., 1991). Feedback can be characterised as informative if users perceive the value of the information in the feedback as relevant, useful, accurate, and satisfactory (Goldin et al., 2017; Rai et al., 2002). Therefore, we expect feedback that is positively valued by its recipients to also be more likely to be used to challenge initial solutions and to be considered for future solutions. Current literature already emphasises the impact of higher usefulness and satisfaction associated with a system on use intentions (e.g., Conati et al., 2021; Hamm et al., 2021). Therefore, we expect informativeness to stimulate users' processing of AI feedback, leading to higher task performance on future tasks. If AI feedback is less informative for users, they are less likely to rely on it to improve their future performance. We therefore put forward the following hypothesis:

H3: Informativeness of AI feedback for users is positively associated with their task performance.

Second, we suggest that informativeness positively affects learning outcome, which represents users' subjective evaluation of their learning with respect to their task derived from using the AI system (Li et al., 2019). While task performance can be an indicator of whether the users' performance has indeed improved during the learning process, learning outcome describes the learning success from the users' perspectives. Highly informative feedback, i.e., feedback that is predictive and enables intervention from the users' perspectives (Tempelaar et al., 2015), is hypothesised to increase users' perceived learning success, as it helps users to understand not only what mistakes they have made, but also how to avoid them in the future (Wisniewski et al., 2020). As a result, researchers draw a direct link between informativeness and learning outcome (e.g., Panigrahi et al., 2021). If AI feedback is less informative for users, they are less likely to perceive its usage as contributing to their learning. We therefore expect the informativeness of AI feedback to be positively associated with users' learning outcome, as described in the following hypothesis:

H4: Informativeness of AI feedback for users is positively associated with their learning outcome.

3.3. Task performance

We hypothesise that task performance is positively associated with learning outcome. Users who increase their task performance during a learning process are expected to also perceive their learning outcome associated with this process as more positive. Previous research has already described the association between objective learning outcome and perceptions of learning outcome (e.g., López-Pérez et al., 2011; Quadir et al., 2022). For example, Quadir et al. (2022) find a positive relationship between objective learning results and subjective learning perceptions when users are engaged in a blog-based interactive learning environment. López-Pérez et al. (2011) observe that for students using blended learning in higher education, objective final grades are interrelated with learning perceptions. Similarly, in our context of learning from AI feedback, we expect a positive correlation between objective and subjective learning outcomes. When users increase their task performance due to the provision of explanations as part of AI feedback, they are likely to increase their understanding and knowledge of the task. Consequently, they are expected to perceive their learning outcome as more positive. We therefore propose the following hypothesis:

H5: Task performance of users of AI feedback is positively associated with their learning outcome.

3.4. Moderating effects of prior knowledge

Both early pioneering work and more recent work on learning from feedback emphasise the role of prior knowledge as a core user-related condition that determines the extent to which users can improve their competences towards the desired reference level when provided with feedback from an external agent (Narciss, 2013). Consequently, we hypothesise that prior knowledge moderates both direct effects of explanations, specifically, their positive effects on users' perceptions of the informativeness of AI feedback and on users' task performance.

First, we expect prior knowledge to moderate the effect of explanations on informativeness. Prior knowledge determines the knowledge level of users upon starting the task. This knowledge level is consequential because users' response to AI feedback is

largely contingent on how the task is related to their prior knowledge (Kulhavy & Stock, 1989). Knowledgeable users may perceive feedback as redundant or too simple and, consequently, become disinterested in feedback (Bangert-Drowns et al., 1991). Therefore, the inclusion of explanations in AI feedback is less likely to have the hypothesised positive effect on informativeness for these users, who have less to benefit from AI feedback to begin with. By contrast, less knowledgeable users are expected to take greater account of feedback because it helps them overcome the significant gap between their initial knowledge and the knowledge required to be successful in the task. As a result, the positive effect of explanations on the informativeness of AI feedback is expected to be larger for less knowledgeable users. In other words, the same explanations will be more informative to less knowledgeable users and less informative to more knowledgeable users. Thus, we hypothesise that the effect of explanations on the informativeness of AI feedback is negatively moderated by users' prior knowledge and propose the following moderation hypothesis:

H6: Prior knowledge of users of AI feedback negatively moderates the effect of explanations on informativeness, so that the effect is higher when prior knowledge is lower.

Second, we expect prior knowledge to moderate the effect of explanations on task performance. According to Feedback Theory, feedback helps to close the gap between an initial knowledge level and a reference level (Ramaprasad, 1983), and this gap is largely determined by users' prior knowledge (Mason & Bruning, 2001). Differences in levels of knowledge are manifested as differences in levels of performance, where higher knowledge should lead to higher performance. For less knowledgeable users, the gap between their preexisting level of performance and the desired level is larger than for more knowledgeable users. Therefore, the less knowledge users possess when they begin the task, the greater the potential of feedback to improve their performance. Explanations provide insights into why a user's solution may be correct or incorrect and form the basis for user interventions, as long as the explanations are novel from the user's perspective. For users with less prior knowledge compared to those with more prior knowledge, explanations are more likely to provide novel insights, stimulate more interventions, and thus lead to greater performance improvements. Therefore, we hypothesise that the effect of explanations on users' task performance is negatively moderated by users' prior knowledge. This hypothesis can further be substantiated by building on the feedback model formulated by Panadero et al. (2018), who postulate that users lacking prior knowledge are more receptive to assistance, such as feedback. More intensive incorporation of feedback in developing evaluative judgements ultimately leads to improved performance. We thus formulate a second moderation hypothesis:

H7: Prior knowledge of users of AI feedback negatively moderates the effect of explanations on task performance, so that the effect is higher when prior knowledge is lower.

Our research model is graphically depicted in Figure 1.

4. Experimental method

To empirically test our hypotheses, we conducted a randomized between-subjects online experiment with a treatment and a control group. We aimed at designing an experimental setting in which users can learn from AI feedback over multiple rounds of the same task and which resembles real-world AI feedback situations. The participants' task was to match Google Street View pictures to their city of origin. As part of the experiment, an AI system was providing participants with AI feedback for their task. For the control group, AI feedback consisted only of the AI decision for the same task. This feedback could be used by participants to assess whether their own

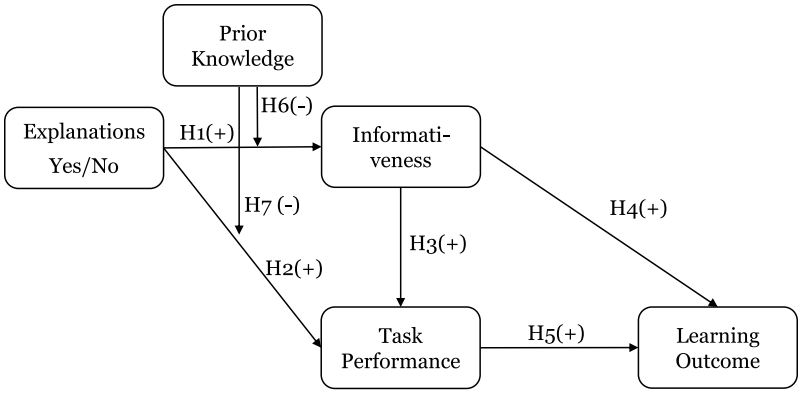


Figure 1. Research model.

decision was correct or not. For the treatment group, the feedback consisted of the AI decision paired with a visual explanation. We implemented a state-of-the-art AI model and XAI method for image classification, which is one of the major tasks of AI systems with increasing attention in XAI research (e.g., Pierrard et al., 2021).

4.1. Experimental setting and AI system

We conducted an online experiment in which participants were supported with AI feedback in a non-routine inference task, specifically matching Google Street View pictures with their originating city over 11 rounds. As defined above, an AI system can consist of an AI model (producing AI decisions) only or an AI model plus an XAI method (automatically generating explanations alongside AI decisions). After each match, the participants received feedback in the form of an AI decision (control group) or an AI decision paired with a visual explanation (treatment group). Google Street View pictures originated from four cities – Berlin and Hamburg in Germany and Tel Aviv and Jerusalem in Israel. Four cities in two different countries were chosen to include both pictures featuring more obvious differences between cities in Western Europe and in the Middle East, as well as pictures featuring more subtle differences between cities within each country. Cities were selected due to popularity (e.g., size and touristic importance) and distinctive differences with regards to building style. Figure 2 illustrates an exemplary round for the treatment group. Here, the automatically generated explanation highlights, for instance, the red brick buildings typical of Hamburg, representing the pixels in the picture that were particularly relevant for the AI decision. Participants in the control group did not receive this visual explanation for why the AI decision was "Hamburg".

The dataset was extracted from Google Street View. To provide all participants with AI decisions, we trained a neural network representing a typical black-box AI model. To create visual post-hoc explanations alongside AI decisions for the treatment group, we used the model-agnostic XAI method LIME. In the following, the dataset, AI model, and XAI method are described in more detail.

4.1.1. Dataset

We extracted 15,000 pictures originating from Berlin and Hamburg in Germany and Tel Aviv and Jerusalem in Israel from Google Street View on January 4 2022. These pictures were divided into 80% training and 20% test data, as common in the literature (Gholamy et al., 2018).

4.1.2. AI model

As AI model, we implemented a neural network with a MobileNetV2 architecture, which is state-of-the-art for image classification (Sandler et al., 2018). We performed hypermeter tuning with the 12,000 training data pictures to optimize the parameters of the network, such as alpha value, optimizer, and number of frozen layers. The optimized network was subsequently used to classify a picture's city of origin (i.e., Berlin, Hamburg, Tel Aviv, or Jerusalem). We achieved an accuracy of 0.88 on the test data, which is a common value in practice (e.g., Ahsan et al., 2020). The neural network served to represent a typical black-box AI model.

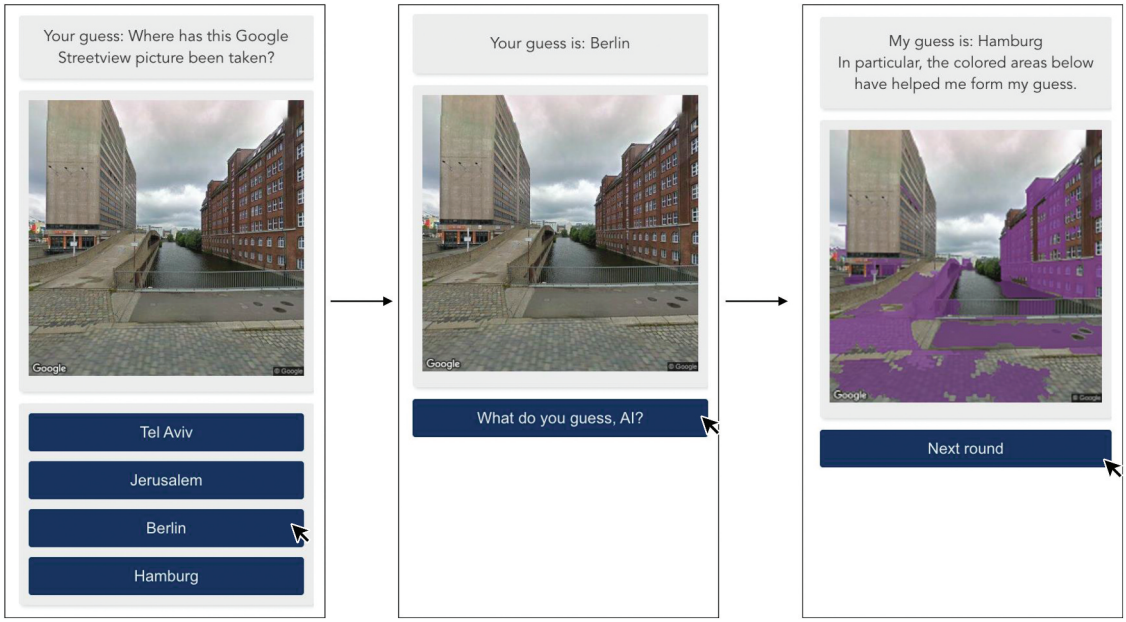


Figure 2. Screenshots from the experiment illustrating one round for the treatment group.

4.1.3. XAI method

For the treatment group, we transformed the AI system into an explainable one by adding an XAI method to the AI model, which automatically generates visual explanations alongside AI decisions. We implemented LIME, which has already been used for various AI applications in the context of image classification (Schallner et al., 2020). In our study, we used LIME to mark areas of the picture that were particularly relevant for the AI decision.

We built upon an instantiation of LIME for picture data (https://github.com/XAI-Demonstrator/visua lime). To generate an explanation, LIME divides a picture into segments using a segmentation method. Then, randomly selected segments are perturbated for a given number of samples. On this basis, a weight is calculated for each segment, which represents its relevance for the AI decision. The resulting explanation takes the form of a picture overlay containing segments in descending order of weight with a desired opacity until the desired coverage is achieved. We selected LIME parameters to optimize robustness and interpretability of explanations (Tan et al., 2023; Zhou et al., 2021) by means of functionally-grounded and human-grounded evaluation (Doshi-Velez & Kim, 2018), as proposed in the literature (Förster et al., 2020b). Following functionally-grounded evaluation, we conducted a grid search for LIME parameters that particularly affected robustness in our setting: segmentation method and the number of samples. For the grid search, we randomly selected 16 pictures from the test data. For each parameter configuration, we generated 20 explanations per picture. We operationalized robustness by two proxy measures: occurrence of segments in the overlay compared to random occurrence and spread of weights of segments over the 20 explanations per picture (Visani et al., 2022). Subsequently, following human-grounded evaluation, we collected user feedback on the interpretability of the explanations in two iterative preliminary studies (with 25 and 52 participants, respectively). In these studies, we varied the parameters segmentation method, coverage, and opacity, which particularly affected interpretability of explanations from the users' perspectives. Only Felzenszwalb and Slic were allowed as segmentation methods in the human-grounded evaluation, as they yielded robust explanations in the functionally-grounded evaluation. Participants in our preliminary studies were asked to provide open feedback on the interpretability of explanations resulting from different parameter configurations for a set of 100 randomly selected pictures from the test data. Based on the results of functionally-grounded and human-grounded evaluation, we set LIME parameters as follows: segmentation method = Felzenszwalb, number of samples = 500, coverage = 0.15, and opacity = 0.5.

4.2. Experimental procedure and measures

The two preliminary studies were not only used to optimize the parametrization of the XAI method, but also to iteratively improve the design of the web interface and to determine appropriate task difficulty regarding picture selection. Appropriate task difficulty was supposed to enable users to learn from the AI system without feeling that the task is excessively difficult or easy. To ensure that, for each picture we set a minimum threshold for users' average task performance of 25% (i.e., users perform better than random guessing) and an upper boundary of average task performance at 88% (i.e., users perform worse than the AI system). We adjusted the picture selection and excluded pictures that did not meet both requirements based on our preliminary studies. For our actual experiment, participants were randomly assigned to the treatment and control groups. The study was conducted as a fully computerized experiment presented via a web interface, which could be accessed via browsers on mobile devices or on computers. Sessions were implemented using the open-source software oTree (Chen et al., 2016).

The study comprised three parts: The first part consisted of an introduction to the study and of demographic questions. Every participant received introductory information on their task in the experiment and the role of the AI system, i.e., supporting participants in learning by provision of feedback. To motivate participants to perform well on their task, they were informed prior to starting the study that after answering the questionnaire they would receive their performance score, and that the top 50% of the participants would receive a bonus payment. All participants were asked to respond to multiple items about their demographic background (gender, age, educational level, and nationality). We also included two questions on prior visits to Germany and Israel.

The second part comprised the main experiment, where participants completed 11 rounds according to Figure 2, with one round serving as attention check. We designed the experiment to consist of 10 rounds based on the results of the preliminary studies, which suggested that users assessed each picture for 30 seconds, on average. Hence, 10 rounds could be completed within five minutes, which corresponds to the common attention span for micro-learning applied in practice (e.g., Jahnke et al., 2019). Each round started with a Google Street View picture, which was randomly selected from a pool of 86 pictures from the test dataset. The participants were asked to match the picture to its city of origin. To this end, four possible choices – Berlin, Hamburg, Tel Aviv, and Jerusalem – were presented along with the picture. Subsequently, participants were shown their own response along with the picture and a mandatory button to ask the

AI system for feedback. Finally, all participants were presented with the feedback of the AI system. Control group participants received the AI decision along with the picture. Treatment group participants received the AI decision and a visual explanation, i.e., highlighted areas in the picture that were particularly relevant for the AI decision. To imitate real-life feedback scenarios that often lack a "ground truth" but rely on the feedback agent's knowledge, the actual origin of the pictures was not revealed – even if the AI decision was incorrect. One of the 11 rounds served as an attention check, because the picture included a street sign with the name of the city, making the answer straightforward to attentive participants. During the experiment, we tracked the scores of the AI system and the user, which were revealed to participants after they completed all tasks and answered the final questionnaire, described next.

The third part consisted of a questionnaire to measure the subjective constructs of the research model together with two additional attention check questions. Our approach to construct measurement was to rely on existing measures. To measure *informativeness*, four items were adapted from Hsieh and Cho (2011), who integrated one item from Rai et al. (2002) and three items from Y. S. Wang (2003). *Learning outcome* was measured by four items adapted from Li et al. (2019), who used established items from the literature (Barzilai & Blau, 2014). Both constructs were measured on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree". The measurement items for the two constructs can be found in the Appendix (Table A1). Besides subjective measures, we objectively measured prior knowledge and task performance. *Prior knowledge* was defined as at least one prior visit to at least one of the countries from which the pictures originated (Bayer et al., 2021). *Task performance* was measured by tracking the scores of the participants (i.e., number of correct answers in the 10 rounds) as an objective measure of performance (Hall et al., 2022).

4.3. Participants

In April 2022, we recruited 744 participants via Clickworker, an established crowdsourcing platform for academic research (Berg & De Stefano, 2018). Participants were meant to represent the entire population of lay users that can potentially use AI feedback. Therefore, the only restrictions placed on participation were adequate English skills (to allow participants to understand task instructions), completion of all task rounds (to enable potential learning), and passing the one-picture and two-question attention checks (to ensure participants were seriously engaged during the entire experiment). Overall, 171 responses had to be removed because of failing at least one of the attention checks, resulting in a sample size of 573 participants (281 in the treatment group and 292 in the control group). The complete experiment took about 8 minutes per participant. Participants were monetarily compensated for their participation according to their task performance. Every participant received 1.30 EUR for participation regardless of their task performance, and the top 50% also received a performance bonus of 1.00 EUR. Among the 573 participants, there were 337 male participants, 233 female participants, and three participants that identified themselves as gender neutral. The average age of the participants was 35.6 (std = 10.8). Participants originated from 54 countries, with the majority of participants being from Germany (249 participants), followed by Brazil (55 participants), India (45 participants), and Nigeria (26 participants). Overall, the treatment and control groups did not differ significantly in demographic variables (gender, education, nationality) and in prior knowledge, generally confirming the effectiveness of random assignment. Only age differed between the groups as participants in the treatment group were slightly older than participants in the control group.

5. Data analysis and results

The proposed research model portrays the hypothesised relationships among *explanations*, *prior knowledge*, *informativeness*, *task performance*, and *learning outcome*. We apply a two-step approach to analyse the research model and test the hypotheses (Hair et al., 2019). While the first step examines the composition of the first-order constructs (measurement model), the second step tests the structural relationships among these constructs (structural model). This process enables to ensure reliability and validity of the measures before examination of the structural model parameters (Hair et al., 2019). For our analysis, we chose Partial Least Squares (PLS) and the software package Smart PLS 4 because PLS applies a component-based approach to estimation, placing minimal restrictions on sample size, measurement scales, and residual distribution (Chin, 1998).

5.1. Measurement model

The subjective constructs defined above represent reflective constructs, as their indicators do not cause change in the construct but rather reflect change of the construct, are interchangeable, and are expected to covary with one another. Thus, they meet all requirements for reflective constructs (Petter et al., 2007). As criteria for measurement models differ for reflective and formative constructs, we first introduced a robustness check based on confirmatory tetrad analysis (CTA-PLS) to empirically substantiate the measurement model's specifications for

Table 3. Descriptive statistics of constructs.

Construct	Items	CR	AVE	Mean	STD	MIN	MAX
Prior Knowledge (PK)	1	1.00	1.00	0.59	0.49	0.00	1.00
Explanations (EX)	1	1.00	1.00	0.49	0.50	0.00	1.00
Informativeness (I)	4	0.87	0.71	3.13	1.05	1.00	5.00
Task Performance (TP)	1	1.00	1.00	5.32	1.87	0.00	10.00
Learning Outcome (LO)	4	0.91	0.78	3.51	1.07	1.00	5.00

Table 4. Loadings and cross-loadings of items (rows) on constructs (columns).

	EX	I	TP	LO
EX1	**1.000**	0.073	0.068	0.052
I1	0.055	**0.801**	0.168	0.515
I2	0.059	**0.863**	0.125	0.549
I3	0.109	**0.851**	0.092	0.562
I4	0.024	**0.858**	0.110	0.590
TP1	0.068	0.146	**1.000**	0.055
LO1	0.076	0.604	0.085	**0.890**
LO2	0.043	0.586	0.076	**0.909**
LO3	0.047	0.584	0.021	**0.880**
LO4	0.015	0.544	0.009	**0.848**

EX = Explanations, I = Informativeness, TP = Task Performance, LO = Learning Outcome

all constructs with at least four indicators (Gudergan et al., 2008). The results confirmed the reflective measurement model with none of the model-implied vanishing tetrads differing significantly from zero (Gudergan et al., 2008).

To test the adequacy of the reflective measurement model, we examined reliability and internal consistency of the measures as well as convergent and discriminant validity (Hair et al., 2019). First, item reliability was assessed based on indicator loadings, which, as can be seen in Table 4, exceeded the 0.708 recommended threshold for all items (Hair et al., 2019). Second, internal consistency reliability was assessed based on composite reliability (CR) values, which, as can be seen in Table 3, were well above the commonly acceptable threshold of 0.7 for all constructs (Fornell & Larcker, 1981; Hair et al., 2010). Third, convergent validity was evaluated based on the average variance extracted (AVE) of each construct, which should be higher than the variance due to measurement error for that construct. Table 3 confirms that all AVE values met this requirement by being higher than 0.5, satisfying the condition for convergent validity (Hair et al., 2019). Fourth, discriminant validity was assessed in two different ways. In terms of factor loadings, the loading in absolute terms of each item on its assigned construct should exceed its cross-loadings on all other constructs (Chin, 1998). Table 4 shows that each item satisfied this condition. While cross-loadings of items between informativeness and learning outcome were relatively high (above 0.5), the loadings of these items were also relatively high (above 0.8), implying that the conditions for a valid analysis were not violated. In terms of construct variance, the square root of the AVE of each construct should exceed the correlations of the construct with other constructs (Fornell & Larcker, 1981). As can be seen in Table 5, these requirements were well met, implying that the measures demonstrated satisfactory discriminant validity. Given the strong evidence of convergent and discriminant validity, as well as the demonstration of high internal consistency and reliability, the measurement model was deemed acceptable.

Table 5. Correlations among constructs and square roots of AVE.

	EX	I	TP	LO
EX	**1.00**			
I	0.07	**0.84**		
TP	0.07	0.15	**1.00**	
LO	0.05	0.66	0.06	**0.88**

EX = Explanations, I = Informativeness, TP = Task Performance, LO = Learning Outcome

5.2. Structural model

The structural model, reflecting our research model, was estimated first with the full sample of 573 participants, before applying a split-sample approach to estimate how *prior knowledge* moderated the direct effects of *explanations*. The results of path analysis with a bootstrap sample number of 5,000 for the full sample are presented in Figure 3. The results showed that *explanations* positively influenced *informativeness*

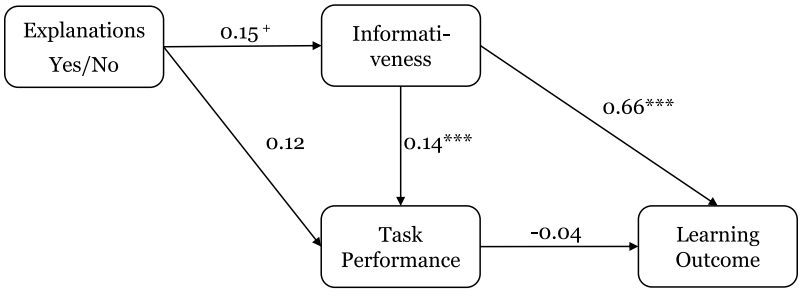


Figure 3. Structural model results for the full sample.

Table 6. Robustness check for nonlinear effects.

	Orig. SPL	Mean	STDEV	|O/STDEV|	*P* value
QE I → TP	0.017	0.017	0.035	0.482	0.630
QE TP → LO	−0.013	−0.013	0.028	0.471	0.638
QE I → LO	−0.020	−0.020	0.033	0.601	0.548

I = Informativeness, TP = Task Performance, LO = Learning Outcome
QE = Quadratic Effects on the endogenous constructs

($\beta = 0.15$, $p < 0.1$), which was positively associated with *task performance* ($\beta = 0.14$, $p < 0.001$) and with *learning outcome* ($\beta = 0.66$, $p < 0.001$). These significant path coefficients provided support for H1, H3, and H4. *Explanations* had no significant effect on *task performance*, implying that H2 was not supported. The results showed no significant effect of *task performance* on *learning outcome*, rejecting H5. The combined paths explained 43.4% of the variance in *learning outcome*, which was our main endogenous construct of interest.

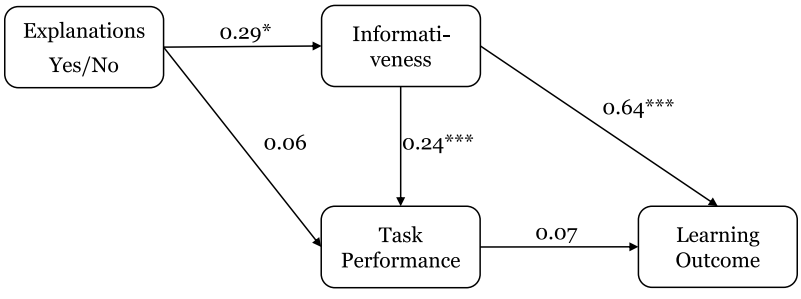
To ensure the robustness of our findings, we performed a series of additional analyses. First, to evaluate whether important structural paths may have been overlooked due to our research model and consequent structural model specification, we tested in a post-hoc analysis the significance of an additional direct (non-mediated) effect of *explanations* on *learning outcome*. The analysis confirmed the non-significance of such a direct effect, implying that this effect was not ignored due to our structural model specification. Second, we tested for common method bias (Podsakoff et al., 2003) by means of a full collinearity test (Kock & Lynn, 2012). Variance inflation factors were generated for all latent variables and were all below 3.3, indicating that the model was free of common method bias (Kock, 2015). Third, as suggested by Hair et al. (2019), we controlled for non-linearity by including three interaction terms to represent the quadratic effects on the endogenous constructs (Svensson et al., 2018). The bootstrapping results, presented in Table 6, indicated that none of the nonlinear effects were significant. Finally, as also suggested by Hair et al. (2019), we checked for unobserved heterogeneity to control for subgroups of data that produce substantially different model estimates. We followed a systematic procedure described by Sarstedt et al. (2017) to identify unobserved heterogeneity. We ran the finite mixture PLS (FIMIX-PLS) procedure on the sample (Sarstedt et al., 2011) and considered between one and 19 segments. This analysis suggested that unobserved heterogeneity did not substantially impact the structural model estimates.

5.3. Moderation effects

Moderation hypotheses in structural models are commonly tested via a split-sample approach, wherein the same structural model is estimated for the different subsamples of the moderator and the moderated paths are compared across subsamples (Hair et al., 2021). To test the moderation effects of *prior knowledge* (H6 and H7), we applied this approach by splitting the sample according to this binary variable. Doing so resulted in a subsample of 233 participants with less prior knowledge, i.e., no prior visits to one of the countries from which the pictures originated, and a subsample of 340 participants with more prior knowledge, i.e., at least one visit to one of the countries from which the pictures originated. For both subsamples, we estimated all the structural paths. Although the moderation effects were pertinent only to the moderated paths (H1 and H2), for robustness purposes, we aimed at ensuring that the remaining paths (H3, H4, and H5) were not similarly moderated.

5.3.1. Structural model for less knowledgeable participants

The results of path analysis with 5,000 bootstrap samples for the subsample of less knowledgeable participants are presented in Figure 4. The results showed that *explanations* positively influenced *informativeness* ($\beta = 0.29$, $p < 0.05$), which was positively associated with *task performance* ($\beta = 0.24$, $p < 0.001$) and with *learning outcome* ($\beta = 0.64$, $p < 0.001$). These significant path coefficients provided support for H1, H3, and H4 for this subsample. *Explanations* had no



+ p < 0.1 * p < 0.05 ** p < 0.01 *** p < 0.001

Figure 4. Structural model results for the subsample of less knowledgeable participants.

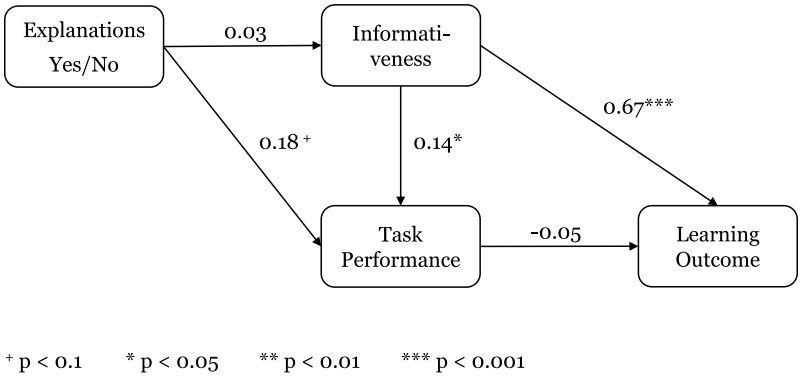


Figure 5. Structural model results for the subsample of more knowledgeable participants.

significant (direct) effect on *task performance*, implying that H2 was not supported. The results indicated no significant effect of *task performance* on *learning outcome*, rejecting H5. The combined paths explained 43.6% of the variance in *learning outcome*. To further understand our structural model results, we also examined the statistical significance of the indirect effects of *explanations* on *task performance* and *learning outcome* in the research model due to the significant direct effect of *explanations* on *informativeness*. These analyses showed that *explanations* had significant indirect effects on *task performance* ($\beta = 0.07$, $p < 0.1$) and *learning outcome* ($\beta = 0.20$, $p < 0.05$). Because *explanations* did not affect these two constructs directly, these indirect effects could be considered as fully mediated by *informativeness*. Finally, we performed the robustness tests, described above, to rule out unobserved direct paths, common method bias, non-linearity, and unobserved heterogeneity and found the results for the subsample of less knowledgeable participants to be robust to these tests.

5.3.2. Structural model for more knowledgeable participants

The results of path analysis with 5,000 bootstrap samples for the subsample of more knowledgeable participants are presented in Figure 5. The results showed that while *explanations* did not influence *informativeness*, providing no support for H1, they did influence *task performance* ($\beta = 0.18$, $p < 0.1$), supporting H2. *Informativeness* was positively associated with *task performance* ($\beta = 0.14$, $p < 0.05$) and with *learning outcome* ($\beta = 0.67$, $p < 0.001$), confirming H3 and H4, respectively. *Task performance* was not associated with *learning outcome*, rejecting H5. The combined paths explained 43.9% of the variance in *learning outcome*. Given these path coefficients, the indirect effects of *explanations* on *task performance* and *learning outcome* were nonsignificant. Finally, we again performed the robustness tests to rule out unobserved direct paths, common method bias, non-linearity, and unobserved heterogeneity and found the results for the subsample of more knowledgeable participants to be robust to these tests.

An integration of the results of path analysis for the two subsamples provided empirical evidence in support of the moderation effects of *prior knowledge*. First, the effect of *explanations* on *informativeness* was significant for less knowledgeable participants but nonsignificant for more knowledgeable participants. These different effects were consistent with H6, confirming that *prior knowledge* negatively moderated the effect of *explanations* on *informativeness*, so that the effect was higher when prior knowledge was lower. Second, the effect of *explanations* on *task performance* was nonsignificant for less knowledgeable participants but significant (although only at the 0.10 level) for more knowledgeable participants. These different effects were opposite to those described in H7, suggesting that *prior knowledge* positively (rather than negatively) moderated the effect of *explanations* on *task performance*. Importantly, the implication of combining the moderation effects of H6 and H7 was that *explanations* positively affected *task performance* for both subsamples, where the effect was mediated by *informativeness* for less knowledgeable participants and direct for more knowledgeable participants. The results for the remaining hypotheses (H3 and H4 were supported while H5 was not supported) were consistent across subsamples, confirming that these effects were not moderated by *prior knowledge* and further strengthening the robustness of our findings. The structural model results are summarized in Table 7.

6. Focus groups

We conducted focus groups to help explain and elaborate on the results of the online experiment. The purpose of the complementary qualitative data was to confirm the quantitative results and expand the understanding of the mechanisms by which explanations facilitate learning from AI feedback. This approach corresponds to an explanatory type of mixed methods design according to Venkatesh et al. (2013).

Table 7. Summary of structural model results.

Hypothesis: Path	Full sample	Subsample of less knowledgeable participants	Subsample of more knowledgeable participants
H1: EX → I	$\beta = 0.15^{+}$	$\beta = 0.29^{*}$	$\beta = 0.03$
H2: EX → TP	$\beta = 0.12$	$\beta = 0.06$	$\beta = 0.18^{+}$
H3: I → TP	$\beta = 0.14^{***}$	$\beta = 0.24^{***}$	$\beta = 0.14^{*}$
H4: I → LO	$\beta = 0.66^{***}$	$\beta = 0.64^{***}$	$\beta = 0.67^{***}$
H5: TP → LO	$\beta = -0.04$	$\beta = 0.07$	$\beta = -0.05$

EX = Explanations, I = Informativeness, TP = Task Performance, LO = Learning Outcome.
$+p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

We recruited a total of 20 participants for four focus groups (five participants each), including two focus groups with 10 AI experts and two focus groups with 10 AI users. The AI experts hold various IT roles, such as CEO of an IT startup, Head of Data Science, Team Lead Analytics & AI, and IT Consultant. They work in different organisations in sectors such as Auditing, Automotive, Healthcare, Manufacturing, and Technology and are deeply involved in AI. They all possess at least three years of relevant AI experience, most of them more than five years. The AI users are students currently enrolled in a bachelor's or master's programme with a subject related to business or management. Details on focus group participants can be found in the Appendix (Tables A2 and A3).

The focus groups were designed according to the guidelines by Stewart and Shamdasani (2017) and conducted via Zoom. To familiarize focus group participants with our research, we gave a presentation of our study (10 slides), which was refined earlier based on a pilot test with five IS Ph.D. students. After this presentation, we initiated a discussion aimed at addressing four open-ended interview questions (Table A4 in the Appendix) about the key findings of the online experiment as well as the practical relevance and boundary conditions of the findings. The participants could discuss and exchange opinions freely during the focus groups. Two researchers accompanied the focus groups, one of them serving as the moderator. The moderator structured the discussions from most to least important topics and from general to specific (Krueger & Casey, 2015). Each focus group lasted about 55 minutes and was recorded and transcribed. As the interviews were conducted in German, the transcriptions were then translated into English. Subsequently, data was coded by two researchers.

Overall, the results of the focus groups confirm the quantitative results. A summary of the responses and sample quotes are presented in the Appendix (Table A5). Furthermore, insights from the focus groups help explain the key findings and boundary conditions. These insights are discussed in the next section, as part of an integrative discussion of our key findings.

7. Discussion

7.1. Key findings

The results of our study offer the fundamental insight that the effect of automatically generated explanations on users' learning from AI feedback differs depending on users' prior knowledge on the task. This insight is associated with three key findings: First, we find in the experiment a positive impact of explanations on learning outcome for less knowledgeable users, which is fully mediated by informativeness. Focus group participants pointed out that informative explanations enhance learning in that they either support users in their decisions (when the explanations align with users' prior knowledge) or provide new insights for subsequent decision-making (when the explanations do not align with users' prior knowledge). Expressed in terms of Feedback Theory, explanations make AI feedback more informative from the users' perspective by revealing the reasoning of the AI model (e.g., Adadi & Berrada, 2018). Thereby, explanations support less knowledgeable users in closing their knowledge gap. Consequently, users' subjective evaluation of their learning success increases. Full mediation of informativeness implies that the entire effect of explanations on learning outcome can be attributed to the ability of explanations to enrich informativeness, suggesting that no additional mechanism is needed to account for this positive effect. As users perceive the informativeness of AI feedback as higher due to the presence of automatically generated explanations, their subjective evaluation of their learning progress improves. Thus, the mere existence of explanations is insufficient for a positive learning experience unless the explanations enhance the users' perception of AI feedback as more informative relative to having AI decisions only. When assessing the impact of XAI on users' learning, future research should focus on the role of informativeness as an important mediator.

Second, our experimental findings show that in contrast to users with less prior knowledge, users with higher prior knowledge do not perceive explanations as enriching the informativeness of AI feedback. Focus group participants attributed this finding – in line with Feedback Theory – to the differences in knowledge gaps depending on users' prior knowledge. Accordingly, explanations in AI feedback are more helpful and provide new insights for users with a larger knowledge gap (i.e., less prior knowledge). By contrast, users with a smaller knowledge gap (i.e., more prior knowledge) find explanations less useful because they more often repeat what users already know. As a result, no effect of explanations on the subjective evaluation of learning success is observed. A possible reason for the absence of such an effect is the type of automatically generated explanations provided in our research setting, specifically, local

explanations that promote the user's understanding of a specific model outcome (Adadi & Berrada, 2018). This type of explanations is also referred to as response-contingent feedback by Mason and Bruning (2001), who suggest that it is appropriate for users with low prior knowledge, whereas users with high prior knowledge are expected to benefit more from topic-contingent feedback (i.e., general elaborative evaluation on the target topic). This latter type of feedback is represented by global explanations in XAI research (Adadi & Berrada, 2018), implying that an intriguing avenue for future research is to investigate the potential positive impact of global explanations on the learning of users with high prior knowledge. The finding that local explanations are more informative and enhance perceptions of learning success when users are less knowledgeable, coupled with the proposition that global explanations may enhance such perceptions when users are more knowledgeable, can extend Feedback Theory to better address the feedback needs of different individuals. In any case, our findings suggest that when exploring the impact of XAI on users' learning, studies should consider the central role of prior knowledge as a critical moderator of the consequences of explanations.

Third, we find a positive impact of including explanations in AI feedback on task performance, which takes a different form contingent on users' prior knowledge. This finding extends the understanding of the role of explanations for the performance of AI users. Whereas existing literature suggests that explanations have a positive impact on task performance when AI is used for decision support (Van der Waa et al., 2021; Walter et al., 2023), our study demonstrates this impact when AI is used for feedback, after a decision has been made. While the question of interest when AI is used for decision support is whether users follow AI decisions, the question of interest in our case is whether users learn from AI feedback. In our study, while both users with less and higher prior knowledge show improvements in task performance in response to the existence of explanations (although the significant effects on task performance are weaker than those on learning outcome), this positive effect of explanations is *direct* for users with higher prior knowledge and *mediated* through informativeness for users with less prior knowledge. Surprisingly, users with higher prior knowledge also show improvements in task performance when explanations are provided as part of AI feedback, although they do not perceive the feedback as more informative, nor do they subjectively evaluate the learning outcome as higher, as observed for users with less prior knowledge. Focus group participants referred to unconscious learning when explaining this finding: Users with more prior knowledge may not immediately appreciate the value of explanations and perceive a learning progress. Nevertheless, they benefit from the additional information provided by explanations in performing their task, suggesting that they possibly learn with less immediate awareness of such learning. Our findings suggest that Feedback Theory should better distinguish between subjective and objective outcomes when considering the individual consequences of feedback. Such a distinction should be particularly salient when the theory is applied to understand how individuals with different levels of knowledge learn from feedback. Our findings demonstrate that perceptions of learning success unnecessarily correlate with task performance, possibly because users may not fully appreciate either the informativeness of feedback for them or the extent to which their performance improved due to feedback. While Feedback Theory acknowledges the distinction between subjective and objective indicators of learning, our findings suggest that this distinction may be critical to understanding how individual contextual factors influence learning from feedback.

Finally, based on an assumption that the type of task is an important boundary condition for learning from AI feedback, our findings are specific to non-routine inference tasks. We focus on these tasks because feedback is particularly needed to augment human decision-making in non-routine tasks, and because AI feedback appears particularly effective in inference tasks. Focus group participants confirmed that explanations in AI feedback become superfluous when tasks are overly simplified. Furthermore, focus group participants pointed out that the impact of explanations on learning outcome, mediated by informativeness, may be independent of prior knowledge if users are required to understand the AI, for example, due to legal requirements. Consequently, our findings about the subjective and objective outcomes of XAI should be interpreted within the boundary conditions in which they are observed.

7.2. Implications for theory and practice

Three important theoretical contributions to IS research arise from this study. First, our study is one of the first to contribute to the understanding of the role of explanations for users' learning from AI feedback. Our findings demonstrate that the provision of explanations as part of AI feedback increases users' task performance, as well as perceptions of informativeness and learning outcome among users with less prior knowledge. Our review of the literature suggests that existing studies focus on AI as decision support, attributing the positive effect of explanations on task performance to system trust

(e.g., Lai & Tan, 2019; Schmidt & Biessmann, 2019), persuasiveness (Schmidt & Biessmann, 2019), and advice following (e.g., Bansal et al., 2021). In our research, explanations are part of AI feedback after users have made their decision. Therefore, the primary role of explanations in such a setting is to help humans gain new insights for future decisions (Van den Broek et al., 2021). Our findings shed light on when (i.e., for less knowledgeable users) and how (i.e., through informativeness) explanations facilitate learning from AI feedback. These findings are an important first step towards bridging the gap between the practical importance and existing knowledge about the ability of XAI to facilitate learning.

Second, our study contributes to the integration of Feedback Theory into the context of XAI. Past XAI research has been anchored in different IS theories, including Cognitive Fit Theory, Elaboration Likelihood Model, Technology Acceptance Model, and Activity Theory. Because XAI research has focused on the goal of decision support rather than the goal of learning, a learning theory has yet to be the foundation of studying the consequences of XAI (Brasse et al., 2023). The present study is the first to integrate Feedback Theory into the context of XAI. Our theoretical and empirical analyses suggest that Feedback Theory is a suitable foundation for studying the impact of explanations on learning from AI feedback. By embracing this theoretical framework, we demonstrate the positive effect of explanations on users' task performance and explain variance in the effect of explanations on learning contingent on users' prior knowledge. This study thus takes an important step towards a differentiated, more fine-grained view of learning from explanations when users receive feedback from an AI system. Overall, the study demonstrates the insight that can be gained by conceptualizing AI models and XAI methods as providers of feedback to users.

Third, we contribute to the literature by viewing XAI through a behavioural lens. A recent review of the IS literature on XAI demonstrates the predominant focus on developing novel XAI approaches while devoting much less research attention to investigating user behaviour (Brasse et al., 2023). Among the latter are studies on trust (e.g., Aechtner et al., 2022; Hamm et al., 2023), understanding of the AI system (e.g., Sieger et al., 2022; Van der Waa et al., 2021), perception of the AI system (e.g., Klein et al., 2023; Y.-F. Wang et al., 2023), and task performance (Van der Waa et al., 2021; Walter et al., 2023). To avoid being misled by incorrect assumptions about user behaviour, researchers stress the need to accompany the development of XAI approaches with the creation of knowledge on the impact of explanations on human-AI interaction (Brasse et al., 2023). Our study advances the understanding of XAI as a driver of user learning, thereby strengthening the behavioural foundations of XAI research and allowing XAI development to better respond to the needs of users through a better understanding of their behaviour.

Beyond theoretical contributions, our findings provide the following practical implications for organisations and developers that wish to leverage the potential of AI feedback for their employees and clients. When developing explanations, organisations and developers need to always keep the user's knowledge level in mind. First, automatically generated explanations can help less knowledgeable users to leverage AI feedback for effective learning in new tasks, particularly those that are non-routine and involve making inferences. To accomplish that, AI feedback needs to incorporate explanations that enhance users' perceptions of the feedback as informative. AI feedback that includes informative explanations can help users to narrow the gap between what they know and what they need to know to successfully perform the task. Therefore, such AI feedback can enable organisations to scale up training and education for the masses. Given the important mediating role of informativeness in the effect of explanations on learning, informativeness can serve as a core indicator to measure the effectiveness of explanations introduced as part of AI feedback to identify untapped learning potential. Therefore, organisations and developers should ensure that the explanations provided represent an information gain from the users' perspectives, consistent with the emerging notion of user-centric XAI design (Förster et al., 2020b). Second, explanations as part of AI feedback can help users improve their task performance regardless of their prior knowledge. Our results show that the provision of explanations as part of AI feedback improves task performance for all users either directly or indirectly. Thus, organisations that aim to improve task performance through AI implementation should consider introducing explanations as part of AI feedback, as the performance of all users can potentially benefit from this introduction. While explanations as part of AI feedback contribute to the subjective learning perceptions of less knowledgeable users, their introduction has no downside for the remaining, more knowledgeable users. These users may be less appreciative of the information provided, but their task performance is still expected to improve.

7.3. Limitations and future research directions

Our study has several limitations, which offer opportunities for future research. First, we assess the impact of explanations on users' learning from AI feedback through a research model that mostly includes endogenous constructs (i.e., only the inclusion of explanations is manipulated through random assignment) that

are measured at the same time. The main implication of this research design is that causality can be validly attributed only to the effects of explanations, whereas the direction of other effects in the research model is based on theoretical rather than empirical considerations. Another implication of this research design is variance in effect sizes, which are relatively small for the effects of explanations on informativeness and task performance. However, such effect sizes are more common for the effects of a manipulated between-subjects variable in an experimental design than for the associations between constructs measured by a single instrument in a survey design. Second, our study design was limited to one specific task (matching pictures to their city of origin) in which users were – beyond monetary incentives as part of the experiment – not personally invested. It is to be expected that the effect of explanations may differ in scenarios where users are personally or professionally affected by the outcome of their task (e.g., employees performing professional tasks, such as credit risk assessment, or patients determining the correct dose of medicine for diabetes self-management). Future studies on users' learning from AI feedback should investigate different tasks, including tasks with different levels of personal investment and complexity. Third, we considered prior knowledge as a binary variable, based on whether participants visited at least one of the countries, which could be correlated with whether participants reside in one of the countries (many of the participants reside in Germany), implying that we cannot confirm the accurate source of prior knowledge in our experiment. We also cannot confirm that our sample is representative of the global population of users, as is typical in online experiments. While we considered prior knowledge as endogenous and analytically ruled out the possibility of significant unobserved heterogeneity, future studies should seek to manipulate this variable (e.g., by providing knowledge to some participants) to confirm the meaningful moderating effects we observe. Future studies may also benefit from a more differentiated view of users' prior knowledge. Fourth, in our experiment, we focused on local explanations generated with LIME, a widely used and state-of-the-art local XAI method for image classification. While we instantiated LIME rigorously according to state-of-the-art literature, our findings about the consequences of explanations should be generalised with caution to other XAI methods. Furthermore, while our results demonstrate the impact of local explanations on learning, the learning experience of users with higher prior knowledge may benefit more from global explanations. Therefore, future research should consider how other types of explanations, such as global explanations, influence learning from AI feedback. Fifth, while we implemented a real AI model with an accuracy that is common in practice, our experiment did not allow to distinguish learning outcome depending on whether AI feedback was correct or not. Future research may benefit from studying the negative consequences for performance and learning when the AI provides incorrect feedback to users. Finally, although we tracked task performance as an objective measure, we measured subjective learning outcome through self-reported items. Future research can benefit from including additional objective measures, for instance pre- and post-task measures of performance, as well as from designing longitudinal studies to assess the long-term effects of explanations on learning from AI feedback.

8. Conclusion

While the emerging research field of XAI shows remarkable progress in the development of approaches that automatically generate explanations alongside AI decisions to make AI systems more traceable to humans, the effects of explanations on human-AI interaction are still poorly understood, particularly when AI serves for learning rather than decision support. Focusing on the use of AI for feedback in non-routine inference tasks, this study investigated the impact of automatically generated explanations on users' learning. Drawing on Feedback Theory as our fundamental theoretical lens, we investigated how explanations enhance the informativeness of AI feedback and users' task performance, contingent on users' prior knowledge, ultimately leading to higher learning outcome. We tested our hypotheses in a randomized between-subjects online experiment with 573 participants, where local explanations for image classification were included in AI feedback only for the treatment group. The results show that explanations enhance perceptions of learning only among users with less prior knowledge by increasing their perceptions of the informativeness of AI feedback as a mediating mechanism. Furthermore, explanations improve users' task performance directly for more knowledgeable users and indirectly (mediated by informativeness) for less knowledgeable users. These mediated and moderated effects of explanations in AI feedback on subjective and objective outcomes are confirmed and extended in focus groups with AI experts and users. Altogether, our findings advance the understanding of how and when providing explanations as part of AI feedback facilitates learning, thereby promoting an agenda of studying how XAI improves human learning.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Péter Horváth Foundation.

ORCID

Lior Fink http://orcid.org/0000-0001-5165-5259

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. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3173574.3174156

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

Aechtner, J., Cabrera, L., Katwal, D., Onghena, P., Valenzuela, D., & Wilbik, A. (2022). Comparing user perception of explanations developed with XAI methods. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1–7). https://doi.org/10.1109/FUZZ-IEEE55066.2022.9882743

Ahsan, M. M., Gupta, K. D., Islam, M. M., Sen, S., Rahman, L., & Hossain, M. S. (2020). Study of different deep learning approach with explainable ai for screening patients with COVID-19 symptoms: Using ct scan and chest x-ray image dataset. *arXiv preprint arXiv:2007.12525*.https://doi.org/10.48550/arXiv.2007.12525

Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A., & Mooney, C. (2021). Current challenges and future opportunities for XAI in machine learning-based clinical decision support Systems: A systematic review. *Applied Sciences*, *11*(11), 5088. https://doi.org/10.3390/app11115088

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*(1), 315–341. https://doi.org/10.25300/MISQ/2021/15882

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213–238. https://doi.org/10.3102/00346543061002213

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. (2021). Does the whole exceed its parts? The effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). https://doi.org/10.1145/3411764.3445717

Barzilai, S., & Blau, I. (2014). Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. *Computers & Education*, *70*, 65–79. https://doi.org/10.1016/j.compedu.2013.08.003

Bayer, S., Gimpel, H., & Markgraf, M. (2021). The role of domain expertise in trusting and following explainable AI decision support systems. *Journal of Decision Systems*, *32*(1), 110–138. https://doi.org/10.1080/12460125.2021.1958505

Beghetto, R. A. (2017). Lesson unplanning: Toward transforming routine tasks into non-routine problems. *ZDM Mathematics Education*, *49*(7), 987–993. https://doi.org/10.1007/s11858-017-0885-1

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

Berg, J., & De Stefano, V. (2018). Employment and regulation for clickworkers. In M. Neufeind, J. O'Reilly, & F. Ranft (Eds.), *Work in the digital age: Challenges of the fourth industrial Revolution* (pp. 175–184). Rowman & Littlefield International Ltd.

Brasse, J., Broder, H., Förster, M., Klier, M., & Sigler, I. (2023). Explainable artificial intelligence in information systems: A review of the status quo and future research directions. *Electronic Markets*, *33*(26). https://doi.org/10.1007/s12525-023-00644-5

Candelon, F., Krayer, L., Rajendran, S., & Zuluaga Martínez, D. (2023). *How people Can create – and destroy – value with generative AI*. BCG Henderson Institute. https://www.bcg.com/publications/2023/how-people-create-and-destroy-value-with-gen-ai

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

Chenoweth, T., Dowling, K. L., St, L., & D, R. (2004). Convincing DSS users that complex models are worth the effort. *Decision Support Systems*, *37*(1), 71–82. https://doi.org/10.1016/S0167-9236(03)00005-8

Chin, W. W. (1998). Issues and opinions on structural equation modeling. *MIS Quarterly*, *22*(1), 7–16.

Chong, M., Shahrill, M., Putri, R. I. I., & Zulkardi, Z. (2018). Teaching problem solving using non-routine tasks. *AIP Conference Proceedings*, *1952*. https://doi.org/10.1063/1.5031982.

Conati, C., Barral, O., Putnam, V., & Rieger, L. (2021). Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence*, *298*. https://doi.org/10.1016/j.artint.2021.103503

Dodeja, L., Tambwekar, P., Hedlund-Botti, E., & Gombolay, M. (2024). Towards the design of user-centric strategy recommendation systems for collaborative human–ai tasks. *International Journal of Human-Computer Studies*, *184*, 103216. https://doi.org/10.1016/j.ijhcs.2023.103216

Doshi-Velez, F., & Kim, B. (2018). Considerations for evaluation and generalization in interpretable machine learning. In Escalante, H. J., Escalera, S., Baró, X., Güçlütürk, Y., Güçlü, U., & van Gerven, M. (Eds.), *Explainable and interpretable models in computer vision and machine learning* (pp. 3–17). The Springer Series on Challenges in Machine Learning.

Elia, I., van den Heuvel-Panhuizen, M., & Kolovou, A. (2009). Exploring strategy use and strategy flexibility in non-routine problem solving by primary school high achievers in mathematics. *ZDM Mathematics Education*, *41*(5), 605–618. https://doi.org/10.1007/s11858-009-0184-6

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, *18*(3), 382. https://doi.org/10.1177/002224378101800313

Förster, M., Klier, M., Kluge, K., & Sigler, I. (2020a). Evaluating explainable artificial intelligence: What users really appreciate. In *Proceedings of the 28th European*

Conference on Information Systems (ECIS). https://aisel.aisnet.org/ecis2020_rp/195

Förster, M., Klier, M., Kluge, K., & Sigler, I. (2020b). Fostering human agency: A process for the design of user-centric XAI systems. In *Proceedings of the Forty-First International Conference on Information Systems (ICIS)*. https://aisel.aisnet.org/icis2020/hci_artintel/hci_artintel/12

Gavaz, H. O., Yazgan, Y., & Arslan, Y. (2021). Non-routine problem solving and strategy flexibility: A quasi-experimental study. *Journal of Pedagogical Research*, *5*(3), 40–54. https://doi.org/10.33902/JPR.2021370581

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). *Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation*. Departmental technical reports (CS), 1209.

Goldin, I., Narciss, S., Foltz, P., & Bauer, M. (2017). New directions in formative feedback in interactive learning environments. *International Journal of Artificial Intelligence in Education*, *27*(3), 385–392. https://doi.org/10.1007/s40593-016-0135-7

Goodhue, D. L. (1995). Understanding user evaluations of information systems. *Management Science*, *41*(12), 1827–1844. https://doi.org/10.1287/mnsc.41.12.1827

Gudergan, S. P., Ringle, C. M., Wende, S., & Will, A. (2008). Confirmatory tetrad analysis in PLS path modeling. *Journal of Business Research*, *61*(12), 1238–1249. https://doi.org/10.1016/j.jbusres.2008.01.012

Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine*, *40*(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Hair, J., Black, W., Babin, B., & Anderson, R. (2010). Multivariate data analysis. *A Global Perspective*, *14*(3), 274–286.

Hair, J., Hult, G. T. M., Ringle, C., Sarstedt, M., Danks, N. P., & Ray, S. (2021). Partial least squares structural equation modeling (PLS-SEM) using R – a workbook. *Springer Cham*. https://doi.org/10.1007/978-3-030-80519-7

Hair, J., Risher, J., Sarstedt, M., & Ringle, C. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hall, K. R., Harrison, D. E., Ajjan, H., & Marshall, G. W. (2022). Understanding salesperson intention to use AI feedback and its influence on business-to-business sales outcomes. *Journal of Business & Industrial Marketing*, *37*(9), 1787–1801. https://doi.org/10.1108/JBIM-04-2021-0218

Hamm, P., Klesel, M., Coberger, P., & Wittmann, H. F. (2023). Explanation matters: An experimental study on explainable AI. *Electronic Markets*, *33*(1). https://doi.org/10.1007/s12525-023-00640-9

Hamm, P., Wittmann, H. F., & Klesel, M. (2021). Explain it to me and I will use it: A proposal on the impact of explainable AI on use behavior. In *Proceedings of the Forty-Second International Conference on Information Systems (ICIS)*. https://aisel.aisnet.org/icis2021/ai_business/ai_business/9

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer New York Inc. https://doi.org/10.1007/978-0-387-84858-7

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hsieh, P. A. J., & Cho, V. (2011). Comparing e-learning tools' success: The case of instructor–student interactive vs. self-paced tools. *Computers & Education*, *57*(3), 2025–2038. https://doi.org/10.1016/j.compedu.2011.05.002

Jahnke, I., Lee, Y.-M., Pham, M., He, H., & Austin, L. (2019). Unpacking the inherent design principles of mobile microlearning. *Technology, Knowledge, and Learning*, *25*(3), 585–619. https://doi.org/10.1007/s10758-019-09413-w

Jia, N., Luo, X., Fang, Z., & Liao, C. (2023). When and how artificial intelligence augments employee creativity. *Academy of Management Journal*. https://doi.org/10.2139/ssrn.4397280

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). Ai-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, *2*, 100017. https://doi.org/10.1016/j.caeai.2021.100017

Kayande, U., De Bruyn, A., Lilien, G., Rangaswamy, A., & Bruggen, G. (2009). How incorporating feedback mechanisms in a DSS affects DSS evaluations. *Information Systems Research*, *20*(4), 527–546. https://doi.org/10.1287/isre.1080.0198

Keane, M. T., & Kenny, E. M. (2019). How case-based reasoning explains neural networks: A theoretical analysis of XAI using post-hoc explanation-by-example from a survey of ANN-CBR twin-systems. In K. Bach & C. Marling (Eds.), *Case-based reasoning research and development, lecture notes in computer science (11680)*. Springer. https://doi.org/10.1007/978-3-030-29249-2_11

Klein, U., Depping, J., Wohlfahrt, L., & Fassbender, P. (2023). Application of artificial intelligence: Risk perception and trust in the work context with different impact levels and task types. *AI & Society*, 1–12. https://doi.org/10.1007/s00146-023-01699-w

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254. https://doi.org/10.1037/0033-2909.119.2.254

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

Kock, N., & Lynn, G. S. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems*, *13*(7), 546–580. https://doi.org/10.17705/1JAIS.00302

Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research*. Sage Publications.

Krulik, S., & Rudnick, J. A. (1993). *Reasoning and problem solving: A handbook for elementary school teachers*. Allyn and Bacon.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*(4), 279–308. https://doi.org/10.1007/BF01320096

Lai, V., & Tan, C. (2019). On human predictions with explanations and predictions of machine learning models: A case study on deception detection. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT)* (pp. 29–38). https://doi.org/10.1145/3287560.3287590

Leyer, M., Oberlaender, A., Dootson, P., & Kowalkiewicz, M. (2020). Decision-making with artificial intelligence: Towards a novel conceptualization of patterns. In *Proceedings of the 24th Pacific Asia*

Conference on Information Systems (PACIS). https://aisel.aisnet.org/pacis2020/224

Li, R., Meng, Z., Tian, M., Zhang, Z., & Xiao, W. (2019). Modelling Chinese EFL learners' flow experiences in digital game-based vocabulary learning: The roles of learner and contextual factors. *Computer Assisted Language Learning*, *34*(4), 483–505. https://doi.org/10.1080/09588221.2019.1619585

Lipnevich, A. A., Berg, D. A., & Smith, J. K. (2016). Toward a Model of Student Response to Feedback. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 69–185). New York: Routledge.

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education*, *6*. https://doi.org/10.3389/feduc.2021.720195

López-Pérez, V., Pérez-López, M. C., & Rodríguez-Ariza, L. (2011). Blended learning in higher education: Students' perceptions and their relation to outcomes. *Computers & Education*, *56*(3), 818–826. https://doi.org/10.1016/j.compedu.2010.10.023

Luan, S., Schooler, L. J., & Gigerenzer, G. (2014). From perception to preference and on to inference: An approach–avoidance analysis of thresholds. *Psychological Review*, *21*(3), 501–525. https://doi.org/10.1037/a0037025

Maier, U., & Klotz, C. (2022). Personalized feedback in digital learning environments: Classification framework and literature review. *Computers and Education: Artificial Intelligence*, *3*, 100080. https://doi.org/10.1016/j.caeai.2022.100080

Malak, J. S., Zeraati, H., Nayeri, F. S., Safdari, R., & Shahraki, A. D. (2019). Neonatal intensive care decision support systems using artificial intelligence techniques: A systematic review. *Artificial Intelligence Review*, *52*(4), 2685–2704. https://doi.org/10.1007/s10462-018-9635-1

Mason, B. J., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us*. Center for Instructional Innovation, University of Nebraska–Lincoln.

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. *Information Systems Management*, *39*(1), 53–63. https://doi.org/10.1080/10580530.2020.1849465

Montzemi, A. R., Wang, F., Khalid Nainar, S. M., & Bart, C. K. (1996). On the effectiveness of decisional guidance. *Decision Support Systems*, *18*(2), 181–198. https://doi.org/10.1016/0167-9236(96)00038-3

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning. *Digital Education Review*, *23*, 7–26.

Nguyen, T., Canossa, A., & Zhu, J. (2024). How human-centered explainable AI interface are designed and evaluated: A systematic survey (arXiv: 2403.14496). *arXiv*.https://doi.org/10.48550/arXiv.2403.14496

Nishant, R., Nguyen, T., Teo, T. S., & Hsu, P. F. (2023). Role of substantive and rhetorical signals in the market reaction to announcements on AI adoption: A configurational study. *European Journal of Information Systems* 23, 1–43. https://doi.org/10.1080/0960085X.2023.2243892

Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2018). Using formative assessment to influence self-and co-regulated learning: The role of evaluative judgement. *European Journal of Psychology of Education*, *34*(3), 535–557. https://doi.org/10.1007/s10212-018-0407-8

Panigrahi, R., Srivastava, P. R., & Panigrahi, P. K. (2021). Effectiveness of e-learning: The mediating role of student engagement on perceived learning effectiveness. *Information Technology & People*, *34*(7), 1840–1862. https://doi.org/10.1108/ITP-07-2019-0380

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, *31*(4), 623–656. https://doi.org/10.2307/25148814

Pierrard, R., Poli, J. P., & Hudelot, C. (2021). Spatial relation learning for explainable image classification and annotation in critical applications. *Artificial Intelligence*, *292*. https://doi.org/10.1016/j.artint.2020.103434

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Quadir, B., Yang, J. C., & Chen, N.-S. (2022). The effects of interaction types on learning outcomes in a blog-based interactive learning environment. *Interactive Learning Environments*, *30*(2), 293–306. https://doi.org/10.1080/10494820.2019.1652835

Rai, A., Lang, S. S., & Welker, R. B. (2002). Assessing the validity of is success models: An empirical test and theoretical analysis. *Information Systems Research*, *13*(1), 50–69. https://doi.org/10.1287/isre.13.1.50.96

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, *28*(1), 4–13. https://doi.org/10.1002/bs.3830280103

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2018.00474

Sarstedt, M., Becker, J.-M., Ringle, C. M., & Schwaiger, M. (2011). Uncovering and treating unobserved heterogeneity with FIMIX-PLS: Which model selection criterion provides an appropriate number of segments? *Schmalenbach Business Review*, *63*(1), 34–62. https://doi.org/10.1007/BF03396886

Sarstedt, M., Ringle, C., & Hair, J. (2017). Partial least squares structural equation modeling. In C. Homburg, M. Klarmann, & A. Vomberg (Eds.), *Handbook of market research*. Springer. https://doi.org/10.1007/978-3-319-05542-8_15-1

Schallner, L., Rabold, J., Scholz, O., & Schmid, U. (2020). Effect of superpixel aggregation on explanations in LIME – a case study with biological data. In P. Cellier & K. Driessens (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp.147–158). https://doi.org/10.1007/978-3-030-43823-4_13

Schemmer, M., Hemmer, P., Nitsche, M., Kühl, N., & Vössing, M. (2022). A meta-analysis of the utility of explainable artificial intelligence in human-ai decision-making. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 617–626). https://doi.org/10.48550/arXiv.2205.05126

Schmidt, P., & Biessmann, F. (2019). Quantifying interpretability and trust in machine learning systems. *arXiv preprint arXiv:1901.08558*. https://doi.org/10.48550/arXiv.1901.08558

Schoonderwoerd, T. A. J., Jorritsma, W., Neerincx, M. A., & van den Bosch, K. (2021). Human-centered XAI: Developing design patterns for explanations of clinical

decision support systems. *International Journal of Human-Computer Studies*, *154*, 102684. https://doi.org/10.1016/j.ijhcs.2021.102684

Sieger, L. N., Hermann, J., Schomäcker, A., Heindorf, S., Meske, C., Hey, C. C., & Doğangün, A. (2022). User involvement in training smart home agents: Increasing perceived control and understanding. In *Proceedings of the 10th International Conference on Human-Agent Interaction* (pp. 76–85). https://doi.org/10.1145/3527188.3561914

Sprague, R. H. (1980). A framework for the development of decision support systems. *MIS Quarterly*, *4*(4), 1–26. https://doi.org/10.2307/248957

Stewart, D. W., & Shamdasani, P. (2017). Online focus groups. *Journal of Advertising*, *46*(1), 48–60. https://doi.org/10.1080/00913367.2016.1252288

Stone, D. N. (1995). The joint effects of DSS feedback and users' expectations on decision processes and performance. *Journal of Information Systems*, *9*(1), 23–41.

Svensson, G., Ferro, C., Høgevold, N., Padin, C., Sosa Varela, J. C., & Sarstedt, M. (2018). Framing the triple bottom line approach: Direct and mediation effects between economic, social and environmental elements. *Journal of Cleaner Production*, *197*(1), 972–991. https://doi.org/10.1016/j.jclepro.2018.06.226

Tan, Z., Tian, Y., & Li, J. (2023). GLIME: General, stable and local LIME explanation. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.), *Advances in Neural Information Processing Systems* (pp. 36250–36277). Curran Associates, Inc.

Te'eni, D. (1991). Feedback in DSS as a source of control: Experiments with the timing of feedback. *Decision Sciences*, *22*(3), 644–655. https://doi.org/10.1111/j.1540-5915.1991.tb01287.x

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, *47*, 157–167. https://doi.org/10.1016/j.chb.2014.05.038

Tyler, N. S., & Jacobs, P. G. (2020). Artificial intelligence in decision support systems for type 1 diabetes. *Sensors*, *20*(11), 3214. https://doi.org/10.3390/s20113214

Van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. *MIS Quarterly*, *45*(3), 1557–1580. https://doi.org/10.25300/MISQ/2021/16559

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

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative Divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, *37*(1), 21–54. https://doi.org/10.25300/MISQ/2013/37.1.02

Visani, G., Bagli, E., Chesani, F., Poluzzi, A., & Capuzzo, D. (2022). Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models. *The Journal of the Operational Research Society*, *73*(1), 91–101. https://doi.org/10.1080/01605682.2020.1865846

Walter, M., Broder, H., & Förster, M. (2023). Boosting benefits, offsetting obstacles – the Impact of explanations on AI users' task performance. In *Wirtschaftsinformatik 2023 Proceedings* (Vol. *29*). https://aisel.aisnet.org/wi2023/29

Wang, Y.-F., Chen, Y.-C., & Chien, S.-Y. (2023). Citizens' intention to follow recommendations from a government-supported ai-enabled system. *Public Policy and Administration*, 095207672311761. https://doi.org/10.1177/09520767231176126

Wang, Y. S. (2003). Assessment of learner satisfaction with asynchronous electronic learning systems. *Information & Management*, *41*(1), 75–86. https://doi.org/10.1016/S0378-7206(03)00028-4

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.03087

Wongvorachan, T., Lai, K. W., Bulut, O., Tsai, Y. S., & Chen, G. (2022). Artificial intelligence: Transforming the future of feedback in education. *Journal of Applied Testing Technology*, *23*(1), 95–116. https://jattjournal.net/index.php/atp/article/view/170387

Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., & Ogbuehi, P. (2012). *Improving mathematical problem solving in grades 4 through 8: A practice guide (NCEE 2012–4055)*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Zhou, Z., Hooker, G., & Wang, F. (2021). S-lime: Stabilized-lime for model explanation. In *Proceedings of the Twenty-Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/3447548.3467274

Appendix

Table A1. Constructs and items in the experiment.

Construct	Item	Source
Informativeness	The AI provides me with accurate information to match Google Street View pictures to their four originating cities.	Hsieh and Cho (2011)
	The AI provides me with useful information to match Google Street View pictures to their four originating cities.	
	The AI provides me with relevant information to match Google Street View pictures to their four originating cities.	
	Overall information provided by the AI on matching Google Street View pictures to their four originating cities is satisfactory.	
Learning Outcome	Interaction with the AI was useful for learning about matching Google Street View pictures to their four originating cities.	Li et al. (2019)
	The AI helped me to learn about matching Google Street View pictures to their four originating cities.	
	The AI facilitated my understanding of matching Google Street View pictures to their four originating cities.	
	My knowledge of matching Google Street View pictures to their four originating cities was increased by interacting with the AI.	

Table A2. Information about the AI experts taking part in the focus groups.

ID	Current IT role	Sector	AI experience
E1	IT Project Manager	Automotive	>5 years
E2	CEO of an IT startup	AI Technology	>5 years
E3	Team Lead Analytics & AI	Mechanical Engineering	>15 years
E4	Data Engineer	Manufacturing	>3 years
E5	Data Scientist	Sensor Technology	>5 years
E6	CEO of an IT startup	AI Technology	>5 years
E7	IT Consultant	Auditing	>5 years
E8	IT Consultant	Corporate Strategy	>5 years
E9	Managing IT Consultant	Digital Transformation	>5 years
E10	Head of Data Science	Healthcare	>5 years

Table A3. Information about the AI users taking part in the focus groups.

ID	Current study program	Years of study
U1	M.Sc. Management and Economics	8 years
U2	M.Sc. Management and Economics	5 years
U3	M.Sc. Management and Economics	6 years
U4	M.Sc. Mathematics and Management	5 years
U5	M.Sc. Mathematics and Management	5 years
U6	B.Sc. Business Informatics	2 years
U7	B.Sc. Management and Economics	3 years
U8	M.Sc. Mathematics and Management	5 years
U9	M.Sc. Management and Economics	5 years
U10	B.Sc. Management and Economics	3 years

Table A4. Interview questions in the focus groups.

ID	Question (Q)
Q1	We find a positive impact of explanations on learning outcome for less knowledgeable users, which is fully mediated by informativeness. Based on your experience, is this finding comprehensible? Please explain and elaborate on this finding.
Q2	We find that the mediated impact of explanations on learning outcome is moderated by users' prior knowledge. Based on your experience, is this finding comprehensible? Please explain and elaborate on this finding.
Q3	We find a positive impact of including explanations in AI feedback on task performance regardless of users' prior knowledge. Based on your experience, is this finding comprehensible? Please explain and elaborate on this finding.
Q4	What is the practical relevance of these findings and what are the boundary conditions?

Table A5. Interview data collected in the focus groups.

Q	Summary of responses	Example quotes (IDs in Tables A2 and A3)
Q1	The participants confirmed that explanations can positively impact users' learning outcome, and that this effect can be mediated by the informativeness of AI feedback. Several participants pointed out that informative explanations enhance learning as they either support users in their decisions (when the explanations align with users' prior knowledge) or provide new insights for subsequent decision-making (when the explanations do not align with users' prior knowledge). Some participants highlighted that informative explanations can trigger knowledge activation and stimulate cognitive engagement, thereby positively impacting learning outcome. A few participants suggested that users might learn to imitate AI decision-making with the help of explanations.	*"Users who find the explanation informative probably learn more from AI feedback". (U7)* *"Informative explanations give users hints about what factors to consider when making further decisions, which leads to the feedback being perceived as helpful". (U5)* *"Informative explanations can sharpen the eye for the peculiarities in the image". (E5)* *"An explanation can support the users in their own decision if the same factors were relevant for the AI decision. But it can also provide additional information if it reveals that the AI considered other factors important. In both cases, the explanation is perceived as informative". (U3)* *"An informative explanation could stimulate a thought process, for example, if the floor is marked, which the users did not consider relevant for their decision". (U6)* *"Explanations could be a trigger to activate implicit knowledge". (E9)* *"An informative explanation stimulates thinking and creates a cognitive incentive to learn". (E7)* *"Explanations provide users with the possibility to adopt and imitate the strategy of the AI". (E3)*
Q2	The participants confirmed the finding that the mediated impact of explanations on learning outcome is moderated by users' prior knowledge. Participants provided two possible interpretations of this finding. The first interpretation refers to the differences in knowledge gaps depending on users' prior knowledge. Accordingly, explanations in AI feedback are more helpful and provide new insights for users with a larger knowledge gap (i.e., less prior knowledge). By contrast, users with a smaller knowledge gap (i.e., more prior knowledge) find them less useful because explanations more often repeat what users already know. The second interpretation refers to how users value additional information depending on their perception of their experience. Users with a perception of high prior knowledge may place less value on the additional information provided by explanations.	*"The results make sense to me. If the cities are unfamiliar, users can learn a lot from the explanation. With more prior knowledge, the learning potential is less pronounced, as the explanation matches one's own knowledge". (U9)* *"Users with more prior knowledge can already explain the marked areas on their own and find the explanation less informative than users with less prior knowledge". (U8)* *"Even if the explanation contains new and relevant information, its helpfulness is perceived as less extreme by users with more prior knowledge than by users with less prior knowledge, because their knowledge gaps are smaller". (U4)* *"Users with prior knowledge do not perceive the additional information provided by explanations as helpful". (E4)* *"Less experienced users may be more influenced by the explanation than more experienced users". (U6)* *"Users with less prior knowledge appreciate the information provided by the explanation more". (E10)* *"A user with prior knowledge may perceive less value in the explanation because the AI strategy revealed may be in contrast to the user's own strategy". (U3)*
Q3	The participants shared the belief that including explanations in AI feedback positively impacts users' task performance, regardless of their prior knowledge. Participants offered different interpretations of this finding, differentiating between users with less and more prior knowledge. On the one hand, users with less prior knowledge find explanations particularly valuable and learn consciously from the additional information provided. On the other hand, users with more prior knowledge may not immediately appreciate the value of explanations and perceive a learning progress. Nevertheless, they benefit from the additional information provided in performing their task, suggesting unconscious learning. Elaborating this further, some participants underpinned that engaging with explanations encourages active thinking and facilitates learning, even if the explanations are not perceived as particularly informative.	*"In summary, this means that explanations always help". (E3)* *"While only users with less prior knowledge explicitly value explanations, users with more knowledge also benefit from them". (E10)* *"Explanations are helpful regardless of prior knowledge. However, users' evaluation of the value of explanations depends on their prior knowledge". (U4)* *"Since AI strategy often differs from human strategy, users can always learn from explanations along AI decisions. However, users with prior knowledge may appreciate this less". (U3)* *"Users with less prior knowledge learn more consciously while users with more prior knowledge learn rather unconsciously. Users with more prior knowledge may also be more convinced of themselves and see the learning effect as their own achievement". (U6)* *"Users with more prior knowledge think they already know what to pay attention to. But subconsciously, they still learn from the explanations". (E5)* *"By being forced to deal with the explanations, users are more likely to decide by active thinking than by intuition". (E7)* *"The time that users take to look at the explanation alone can have a positive effect on their learning success". (E4)*
Q4	Several participants recognized the direct business relevance of the findings, for example in employee training. Some participants highlighted the particular relevance of the findings for tasks in which users tend to be overconfident in their abilities. The participants identified three boundary conditions and constraints for the findings. First, the explanations must be interpretable by the users. Second, the tasks should be complex enough to prevent explanations from becoming superfluous. Third, the impact of explanations on learning outcome, mediated by informativeness, may be independent of prior knowledge if users are required to understand the AI, for example, due to legal requirements.	*"The results are transferable to image classification. For example, in radiology". (U8)* *"To support employees in their daily decisions, it would be useful to back up the AI with an explanation to increase acceptance and promote further training". (E1)* *"A novice programmer might find explanations helpful, while the expert just wants to make a quick decision". (U6)* *"Especially for questions where users would intuitively be wrong, an explanation would be useful to make users aware of their overconfidence". (E3)* *"The effectiveness of explanations in AI feedback strongly depends on the type and quality of the explanations". (U4)* *"The user's ability to interpret explanations must be present". (E9)* *"I can't imagine that the same results occur in simple tasks. There may be no significant effect of the explanations, or they may quickly be perceived as superfluous". (U7)* *"Explanations in AI feedback should be helpful for all users who need to understand the AI, for instance, due to legal reasons". (E2)*