Human-XAI Interaction: A Review and Design Principles for Explanation User Interfaces

Michael Chromik, Andreas Butz

Human-XAI Interaction: A Review and Design Principles for Explanation User Interfaces

Michael Chromik and Andreas Butz

LMU Munich, Munich, Germany
michael.chromik@ifi.lmu.de
butz@ifi.lmu.de

Abstract. The interdisciplinary field of explainable artificial intelligence (XAI) aims to foster human understanding of black-box machine learning models through explanation-generating methods. Although the social sciences suggest that explanation is a social and iterative process between an explainer and an explainee, explanation user interfaces and their user interactions have not been systematically explored in XAI research yet. Therefore, we review prior XAI research containing explanation user interfaces for ML-based intelligent systems and describe different concepts of interaction. Further, we present observed design principles for interactive explanation user interfaces. With our work, we inform designers of XAI systems about human-centric ways to tailor their explanation user interfaces to different target audiences and use cases.

Keywords: explainable AI · explanation user interfaces · interaction design · literature review.

1 Introduction

Intelligent systems based on machine learning (ML) are widespread in many contexts of our lives. Often, their accurate predictions come at the expense of interpretability due to their black-box nature. As consequential predictions of these systems may raise questions by those who are affected or held accountable, there is a call for *"explanations that enable people to understand the decisions"* [85]. Hence, much research is conducted within the emerging domain of explainable artificial intelligence (XAI) and interpretable machine learning (IML) on developing methods and interfaces that human users can interpret – often through some sort of explanation. Often there is not a single explanation to be conveyed [1]. Therefore, the DARPA XAI program describes the XAI process as a two-staged approach. It distinguishes between the explainable model and the explanation user interface [37] and, thus, disentangles analyzing the ML model behavior from communicating it to the user. We define an *explanation user interface (XUI)* as the sum of outputs of an XAI system that the user can directly interact with. An XUI may tap into the ML model or may use one or more explanation generating algorithms to provide relevant insights for

a particular audience. The design of interfaces that *"allow users to better understand underlying computational processes"* is considered a grand challenge of HCI research [86]. Shneiderman considers XUIs as a building block towards *human-centered AI* which aims *"to amplify, augment and enhance human performance"* instead of automating it [85].

However, most XAI research focuses on computational aspects of generating explanations while limited research is reported concerning the human-centered design of the XUI [89, 85, 102]. Similarly, resources targeting practitioners, such as UK's Information Commissioner's Office[1] , who aim to provide practitioners with *"guidance [that] is practically applicable in the real world"*, do not touch on explanation user interfaces nor how to present them to users and instead propose *"...to draw on the expertise of user experience and user interface designers"*. A notable exception is Google's *People+AI Guidebook*[2] which presents case studies of explanations integrated into mobile apps. As the human use of computing is the subject of inquiry in HCI [73], our discipline *"should take a leading role by providing explainable and comprehensible AI, and useful and usable AI"* [105]. In particular, our community is well suited to *"provide effective design for explanation UIs"* [105].

To follow this call and to understand the current practices in the field, we took an HCI perspective and conducted a systematic literature review. The overarching research question (ORQ) of our work is to **survey how researchers designed XUIs in prior XAI work.** From there, we analyze the user interactions offered by the XAI systems and describe observed design patterns. Our work is guided by the following more specific research questions:

- RQ1: How can the different concepts of interaction in XAI be characterized?
- RQ2: What design principles for interactive XUIs can be observed?

The increasing demand for interpretable systems also raises the question how to present this interpretability to users. The contribution of this paper is twofold: First, we provide a structured literature overview of how user interaction has been designed in XAI. Second, we outline design principles for human interaction with XUIs. Our work guides researchers and practitioners through the interdisciplinary design space of XAI from an HCI perspective.

2 Background and Related Work

2.1 Interaction in Surveys of Explainable AI

XAI is an umbrella term for algorithms and methods that extend the output of ML-based systems with some sort of explanation. The goal is *"to explain or to present [the ML-based system] in understandable terms to a human"* [27].

[1] ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-and-the-turing-consultation-on-explaining-ai-decisions-guidance/

[2] pair.withgoogle.com/chapter/explainability-trust/

Multiple reviews of the growing field of XAI exist. They formalize and ground the concept of XAI [1, 3], relate it to adjacent concepts and disciplines [1, 62], categorize methods [36, 57], analyze the user perspective [33], review evaluation practices [65], or outline future research directions [1, 3]. Most of these reviews acknowledge the importance of interaction for XAI only as a side note. For instance, Mueller et al. [65] consider an effective explanation to be *"an interaction"* and *"not a property of statements"*. Adadi et al. [3] state that *"explainability can only happen through interaction between human and machine"*. Abdul et al. [1] present research on interactive explanation interfaces as an important trajectory to advance the XAI research field. However, none of these reviews elaborates how this interaction could be described nor designed to inform researchers and practitioners. To our knowledge, none of the review look at XAI from an interaction design perspective.

On a broader level, there is a line of research on how to design the overall human interaction with AI-infused systems. For instance, Amershi et al. present guidelines for AI-infused systems [5]. While not explicitly addressing interpretability nor explanations, they point out the importance of making clear why the system did what it did in case of errors. However, their guidelines do not outline what this interaction could look like.

2.2 The XAI Pipeline and Explanation User Interfaces

The XAI process can be broken down into different steps. Murdoch et al. distinguish between the predictive accuracy, the descriptive accuracy, and the relevancy of an XAI system. *Predictive accuracy* is the degree to which the learned ML model correctly extracts the underlying data relationships. *Descriptive accuracy* (also referred to as fidelity) is the degree to which an explanation generation method accurately describes the behavior of the learned ML model. Both accuracies can be objectively measured. In contrast, the subjective *relevancy* describes if the outputs are communicated in a way that they provide insights for a particular audience into a chosen domain problem [67].

The DARPA XAI program illustrates the XAI process as a two-staged approach. It distinguishes between the explainable model and the explanation user interface [37]. The former addresses the predictive and descriptive accuracies, while the latter aims for relevancy. Such a two-staged approach disentangles the XAI process into analyzing the ML model behavior and communicating it to the user. Similarly, Danilevsky et al. [21] differentiate between explainability techniques and explainability visualizations. The former generates *"raw explanations"* typically proposed by AI researchers while the latter is concerned with the presentation of these *"raw explanations"* to users typically guided by HCI researchers. Most open-source methods for XAI provide a single explanation generation method. However, there is a growing number of explanation generation toolkits (e.g., AIX 360[3], Alibi[4], DALEX[5]) that combine multiple state-of-the-art

[3] https://aix360.mybluemix.net/
[4] https://docs.seldon.io/projects/alibi/en/latest/
[5] https://uc-r.github.io/dalex

methods in a uniform programming interface and thus enable rapid prototyping of XUI.

In this work, we define an explanation user interface (XUI) as the sum of outputs of an XAI process that the user can directly interact with. Shneiderman [85] outlines two modes of XUI. *Explanatory* XUIs aim to convey a single explanation (e.g., a visualization or a text explanation). In contrast, *exploratory* XUIs let users freely explore the ML model behavior. They are most effective when users have the power to change or influence the inputs. Arya et al. [7] distinguish between static and interactive explanations. A static explanation *"does not change in response to feedback from the consumer"*. In contrast, interactive explanations allow *"to drill down or ask for different types of explanations [...] until [...] satisfied"*.

3 Methodology

In line with our ORQ, our method for characterizing interaction in XAI was to collect a corpus of publications using the structured search approaches by Kitchenham and Charters [47]. We then analyzed the corpus regarding the interaction concepts followed by the authors as well as the design and interaction functionalities offered to users.

To collect a corpus of candidate publications, we conducted a systematic search in the *ACM Digital Library*. We limited our search to work that has been published at venues relevant to HCI (*Sponsor SIGCHI*). Through initial exploratory search, we obtained an initial understanding of relevant keywords, synonyms, and related concepts that helped us to construct the search query. Different terms are used to describe the field of XAI and XUI [1]. We focused on publications that include user-centered artefacts with explicit forms of explanation for the underlying intelligent behavior. Our primary focus was on research that builds on the potentials of current algorithmic explanation-generating XAI methods and thus often self-identifies as *"XAI"* or *"explainable AI"*. To account for the historic perspectives, we included *"explanation interface"* and *"explanation facility"*. These terms emerged in the 2000s from the recommender systems community and have often been used as a umbrella term for user interfaces covering different explanatory goals [92]. Further, we were interested in research that has a user focus and mentions some form of *"user interaction"*, *"user interface"*, or aspects of *"usability"* or *"interactive"*. We prepended the terms interaction and interface with *"user"* to distinguish them from feature interactions and system interfaces. While not covering the entire dynamic of this interdisciplinary field, this scoping resulted in a diverse set of works from multiple decades that put a focus on the user interface artefact. This resulted in the following search query:

> [[All: "xai"] OR [All: "explainable ai"] OR [All: "explanation facility"] OR [All: "explanation interface"]] AND [[All: "user interaction"] OR [All: "user interface"] OR [All: usability] OR [All: interactive]]

We conducted the search procedure in December 2020, which returned a total of 146 results. We then analyzed the full-text of all results. We excluded 13 results without a contribution (i.e., proceedings, keynotes, workshop summaries). Publications included in our analysis had to present results from *constructive* [73] research that involved an XUI artefact (n=57) or *conceptual* [73] research that addresses interaction in XAI (n=34). Consequently, we excluded 28 results that were not related to XAI and 14 results that were related to XAI but did not present an XUI nor describe interaction. The review was conducted by the first author. The second author was consulted for feedback. Our final set for analysis consisted of 91 publications. We analyzed the selected publications and coded information about the reported XUI and user interactions in a database.

4 Concepts of Interaction in XAI

Following Hornbæk and Oulasvirta [42], *interaction* describes the interplay between two or more constructs. They analyzed the interplay between the constructs human and computer that were discussed in HCI research. From this, they derived seven concepts of interaction: interaction as information transmission, as dialogue, as control, as experience, as optimal behavior, as tool use, and interaction as embodied action. More narrowly, Miller frames XAI as one kind of a human-agent interaction problem where an *"explanatory agent [is] revealing underlying causes to its or another agent's decision making"* [62]. As such, it is about the interplay between a human user and an AI agent that is mediated through an XUI. Tintarev and Masthoff [92] distinguish seven explanatory goals: transparency (answer how the system works), scrutability (allow to question and correct the system), trustworthiness (increase user confidence), persuasiveness (convince user), effectiveness (help user making good decisions), efficiency (help user making decisions faster), and satisfaction (increase usability). As these may be conflicting with one another, designers of XUI *"need to make trade-offs while choosing or designing the form of interface"* [93].

We build on the interaction concepts of Dubin and Hornbaek [42] and apply them to human-XAI interaction. To answer RQ1 (How can the different concepts of interaction in XAI be characterized?), we analyzed the primary interaction concept that authors (implicitly) applied as part of their work. In particular, we focus on the interplay between a user and an AI system that is facilitated through a UI that leverages some kind of explanation to reach an explanatory goal. We abstracted from the purpose that the researchers used the XUI for and instead looked at how a user could interact with it. As such, we approached the concepts of interaction with an *artefactist approach* [90]. Below, we introduce each concept and relate them to surveyed publications. Table 1 summarizes our analysis.

4.1 Interaction as (Information) Transmission

This concept centers around maximizing the throughput of information via a noisy channel. The interaction is about selecting the best message for transmis-

sion from a set of possible messages [42]. It follows the *Shannon-Weaver* [84] model of communication according to which the sender transmits information to the receiver but in between noise is added to the original message.

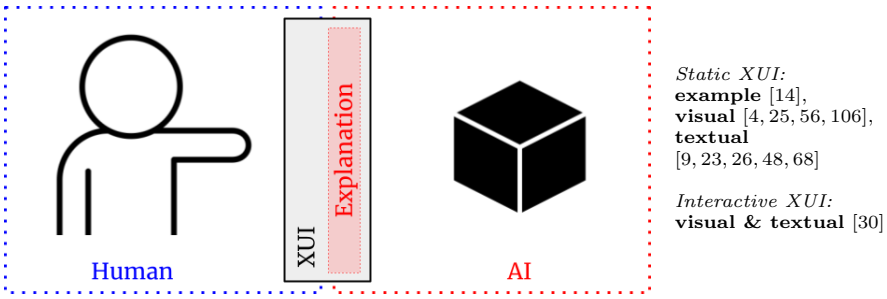

Fig. 1. XAI-interaction as (information) transmission is about presenting an accurate and complete explanation about the AI behavior.

Transfer to XAI: The goal of this interaction centers around presenting users with one complete explanation. Surveyed publications following this concept are mostly driven by the explanatory goal of transparency and acknowledge that *"algorithms should not be studied in isolation, but rather in conjunction with interfaces, since both play a significant role in the perception of explainability"* [25]. They emphasize either (i) the descriptive accuracy of an explanation to describe the underlying AI behavior [26, 30, 48, 56, 68] or (ii) the capacity of a single explanation style [4] or differences between explanation styles [9, 14, 23, 25, 106, 83] to convey information about the behavior to the human. The message is noisy because it may be difficult or even impossible to fully describe the complexity of the AI in a human understandable way, such as with deep neural networks. Unlike interaction as a dialogue, this interaction is mainly about unidirectional communication by presenting a single and static explanation. The XUI is mainly used as a medium for transmitting this explanation.

Examples: Ehsan et al. [30] present real-time explanations about the actions taken by an autonomous gaming agent in the form of natural language rationales. Alqaraawi et al. [4] study whether saliency maps convey enough information to enable users to anticipate the behavior of an image classifier. Cai et al. [14] compared how well two example-based explanation styles could promote user understanding of a sketch recognition AI. Dodge et al. [23] and Binns et al. [9] study how much different textual explanation styles convey about underlying fairness issues of an ML system. Yang et al. [106] study the differences in spatial layout and visual representation of example-based explanations.

4.2 Interaction as Dialogue

This concept describes a cycle of communication of inputs/outputs by the computer and perception/action by a human. The interaction happens in stages or

turns [42]. It tries to ensure a correct mapping between UI functions and the user's intentions and feedback by the UI to bridge the *gulf of execution* [69].

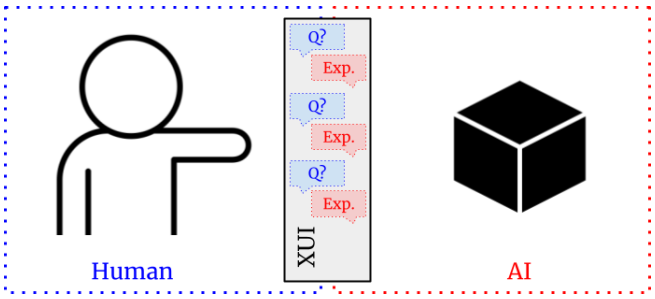

Fig. 2. XAI-interaction as dialogue is about facilitating an iterative communication cycle about the AI behavior.

Transfer to XAI: This concept acknowledges that a single explanation rarely results in a desired level of understanding [1]. Instead, it emphasizes the naturalness and accessibility of (often implicit or simplified) explanations. In contrast to interaction as embodied action, this concept is driven by the user, with the AI responding. Unlike interaction as control, this concept does not change the AI behavior. The goal of the interaction is to provide users with functionalities to gradually build a mental model of the AI behavior. We distinguish between inspection dialogues [10, 18, 28] and natural dialogues [46].

Inspection Examples: Exploratory dialogues allow the user to explore how (possibly hypothetical) changes in inputs lead to changes in the AI prediction or let the user inspect internals of the AI. The XUI is mostly about offering functionalities to iteratively request explanations of the same kind. Explanations have a high fidelity but are implicit. For instance, Cheng et al. [18] present an XUI that allows users to observe how the predictions of a university admission classifier change by freely adjusting the values of input features of applicants. Their exploratory approach was shown to improve users' comprehension although it required more of their time. Bock and Schreiber [10] present an XUI to inspect layers and parameters of deep neural networks in virtual reality. Similarly, Douglas et al. [28] visualize an AI agent's behavior in form of interactive saliency maps in virtual reality.

Natural Examples: Natural dialogues aim to *"lower the threshold of ability required to analyze data"* and thus make XUIs more accessible to end users of XAI. The XUI is about presenting functionalities to request different natural language explanations. The interaction is mostly driven by the human through questions. Explanations are explicit but simplified in the form of textual answers. Kim et al. [46] present an XUI that enables users to ask factoid questions about charts in natural language (e.g., *"What age had the lowest population of males?"*). The XUI provides the answer and an explanation how it was derived from the chart (e.g., *"I looked up 'age' of the shortest blue bar*).

4.3 Interaction as Control

This concept supports a rapid and stable convergence of the human-computer system towards a target state. Building on *control theory*, the interaction is aiming *"to change a control signal to a desired level and updating its behavior according to feedback"* [42].

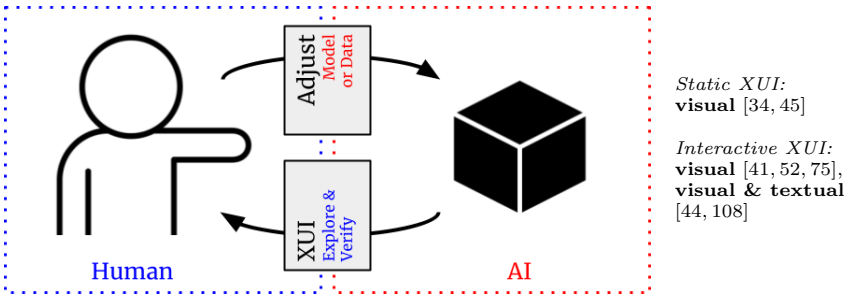

Fig. 3. XAI-interaction as control is about supporting a rapid convergence towards the desired AI behavior.

Transfer to XAI: This concept aligns with the ideas of interactive ML [29] and ML model tweaking. The XUI feeds control signals from the ML model to the human controller (feedback). These inform the controller how to change parameters of the ML model or its data so that the model adjusts its behavior (feedforward). The goal of the interaction is to reach the AI behavior desired by the controller. We found two streams of research that follow this paradigm. They can be distinguished by their targeted users: *AI experts* [41, 45, 52, 75, 78] or *AI novices* [44, 34, 108].

AI Expert Examples: Explanations are provided mainly on an abstract level as numbers and visualizations. The cycle of exploration and verification drives the process of understanding. The XUI is a standalone application facilitating this interaction while the actual model adjustments are performed in a separate UI (e.g., the development environment). For instance, [78] present an early XUI to debug rule-based expert systems by explaining why a rule was fired. Krause et al. [52] present the interactive visual analytics systems *Prospector*, that supports data scientists in understanding local predictions and deriving actionable insights on how to improve the ML model. They can (i) explore local predictions and simulate counterfactual changes by different ML models to support the formulation of tweaking hypotheses or (ii) verify how their implemented tweaking hypotheses change the prediction behaviour of the ML model. Hohman et al. [41] present *Gamut*, an XUI were *"interactivity was the primary mechanism for exploring, comparing, and explaining"*. User can link local and global explanations, ask counterfactual and compute similar instances. In contrast, Kaur et al. [45] show that the non-interactive XUIs of widely used explainability tools, such as InterpretML or SHAP, hinder experts to effectively control ML models.

AI Novice Examples: These XUI strive *"to effectively communicate relevant technical features of the [ML] model to a non-technical audience"* [108]. These XUIs provide explicit explanations to support the exploration. They also integrate controls for adjusting underlying the ML models without the need of a separate UI. Yu et al. [108] present an XUI for ML classification in the sensitive context of criminal justice. Their XUI enables designers and end-users to explore and understand algorithmic trade-offs based on an interactive confusion matrix and textual explanations. Further, it allows them to adjust model thresholds in a way that reflects their fairness beliefs (feedforward). Ishibashi et al. [44] present an XUI that synergetically combines low-level spectrograms with semantic thumbnails to interactively train a sound recognition AI. Fulton et al. [34] showcase how an XUI can be integrated into games for AI novices to generate usable data for AI experts.

4.4 Interaction as Experience

This concept considers human expectations towards a computer. It is closely related to *user experience* (UX) encompassing a person's emotions, feelings, and thoughts that may be formed before, during, or after interaction [53].

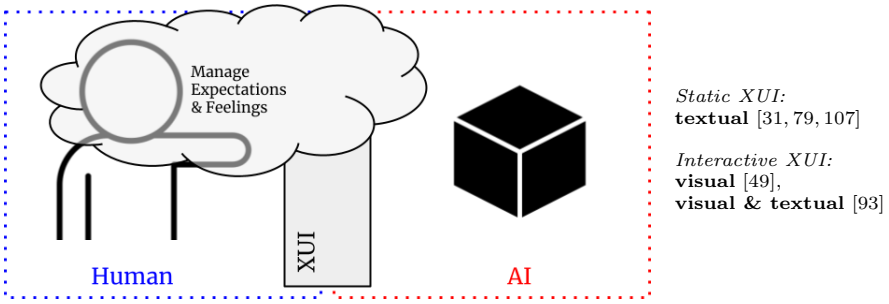

Fig. 4. XAI-interaction as experience is about managing expectations about the AI behavior.

Transfer to XAI: Applied to XAI, this interaction concept emphasizes managing the expectations and preferences of users about the AI. It centers around the explanatory goals of trust [49, 77, 79, 107], satisfaction [93], and persuasiveness [31].

Examples: Knijnenburg et al. show that letting users inspect a recommendation process through an interactive XUI increased their perceived understanding and satisfaction. Tsai et al. [93] investigate the relation of user preferences about explanation styles and user performance. Their results suggest that XUIs preferred by users *"may not guarantee the same level of performance"*. Yin et al. [107] show that a user's trust is impacted by upfront information on the AI's predictive accuracy even after repeated interactions. Pushing this interaction concept, Eiband et al. [31] show with their XUI that even empty (so-called placebic) explanations

can result in a soothing perceived understanding of users. As an intervention, Pilling et al. [77] outline a design fiction of an AI certification body that provides users with standardized AI quality marks (e.g., *"level 4: product is able to explain itself to users on request."*).

4.5 Interaction as Optimal Behavior

This concept centers around adapting the user behavior to better support their tasks and goals. It acknowledges that the interaction with the system is often constrained, and thus suboptimal. Users are trading off rewards and costs of an interaction. It builds around the idea of *bounded rationality* [87] according to which humans act as "satisficers" who strive for satisfying and sufficient solutions (instead of optimal ones) due to cognitive limitations.

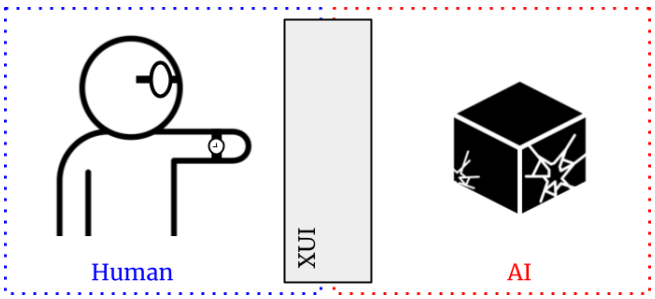

Fig. 5. XAI-interaction as optimal behavior is about adjusting the human behavior despite the cognitive or technical limitations of fully understanding the AI behavior.

Transfer to XAI: Applied to XAI research, the goal of the interaction is to guide users to reach a "satisficing" level of AI understanding for some downstream task. It focuses on providing explanations for *"training humans to have better interactions with AI"*, for example, when they face erroneous AI systems [99] or exhibit misconceptions caused by cognitive biases [97]. We distinguish between research that (i) examines limitations that occur during the interaction with an XAI [12, 13, 24, 60, 61, 70, 97] and (ii) designs interactions to better moderate these limitations [2, 16, 19, 59, 81, 98, 99].

Examples that Examine Limitations: Millecamp et al. [61] studied the impact of personal characteristics on the interaction and perception of XAI in a music recommender setting. They show that the perception and interaction with XUIs is influenced by a user's need for cognition (NFC) (i.e., their tendency to engage in and enjoy effortful cognitive activities). Nourani et al. [70] show that a user's first impression of an AI system influences their overall perception of the system. While a positive first impression may lead to automation bias, a negative first impression may result in a less accurate mental model. They call for XUIs that control a user's first impression and *"continually direct user attention to system strengths and weaknesses throughout user-system interactions"*. Similarly, Bucinca et al. [12] highlight that the effectiveness of XAI is impacted by the

design of the interaction itself. Thus, it is important to take *"into account the cognitive effort and cognitive processes that are employed [by the user]"* during their interpretation of explanations.

Examples that Moderate Limitations: Several of the works designed interactions that *"optimize the performance of the sociotechnical (human+AI) system as a whole"* [12]. For example, Wang et al. [98] provide confidence explanations to help users to gauge when or when not to trust an AI. Similarly, Schaekermann et al. [81] show that highlighting and textually explaining ambiguous predictions helps physicians to *"allocate cognitive resources and reassess their level of trust appropriately for each specific case"*. Abdul et al. [2] propose a visual explanation style that balances cognitive load and descriptive accuracy by limiting the visual chunks to be processed by the user. Further, they present a method to estimate users' cognitive load of explanations. Weisz et al. [99] teach users strategies to effectively interact with a limited capability chatbot in a banking and shopping context. Their interaction aims to explain to users why a chatbot may be unable to provide meaningful responses. For instance, explaining that the chatbot mapped the user's utterance to multiple low confidence intents because the utterance was poorly worded or ambiguous. Mai et al. [59] guide users through a military-inspired structured reflection process, called *after-action review* to understand the behavior of an AI agent. Accompanied by a visual explanation of AI decisions, the reflection process helped users to organize their cognitive process of understanding and kept them engaged.

4.6 Interaction as Tool Use

This concept centers around using computers to augment the user's capabilities beyond the tool itself. Following *activity theory*, the system influences the *"mental functioning of individuals"*. As such, AI can also be used as a tool for learning. For example, the social sciences use word embeddings as a diagnostic tool to quantify changes in society [35].

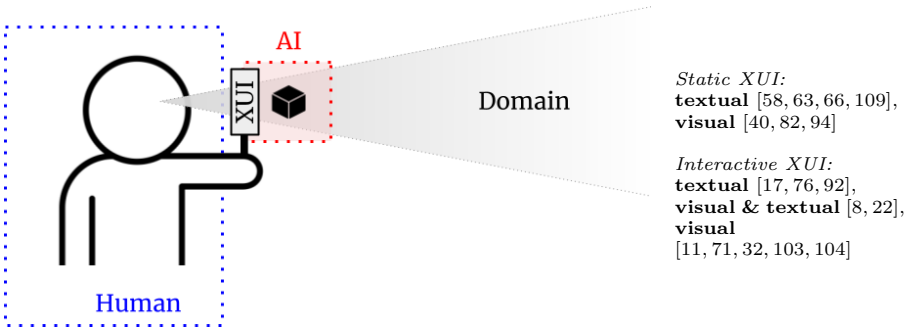

Fig. 6. XAI-interaction as tool use is about facilitating learning from the AI behavior about a given domain.

Transfer to XAI: Applied to XAI, this interaction concept helps humans to find hidden patterns and insights in domain-specific data. To facilitate this learning, some form of explanation is required. The XUI serves as a lens on a domain (beyond the AI behavior) that would otherwise be difficult to understand. In this way, the interaction contributes to augment human thinking.

Examples: Xie et al. [104] assist physicians analyzing chest x-rays of patients through an interactive mixed-modality XUI. Paudyal et al. [76] presents an interactive XUI for a computer-vision based sign language AI. The textual explanations provide learners with feedback on the location, shapes, and movements of their hands. Similarly, Schneeberger et al. [82] use an XUI to let users practice emotionally difficult social situations with a social AI agent. Das et al. [22] present an XUI which provides feedback on a chess player's intended moves. Their visual highlighting and textual explanations significantly improved the performance of chess players in a multi-day user study. They point out the importance of accompanying textual explanations for the AI reasoning. Only showing the visual explanation did not improve performance. Similarly, Feng et al. [32] support players by visually explaining evidences for each uncovered word of a quiz question. Xie et al. [103] use an interactive XUI with visual explanations to give game designers live-feedback on how challenging their created level designs are. Misztal-Radecka and Indurkhya [63] generate textual user stories for personas from large datasets to inform interaction designers about potentially relevant user groups.

Explainable Recommender Systems: In addition, most works on explainable recommender systems follow this interaction concept as their recommendations aim to give users insights about the recommender domain [40]. Some XUIs allow personalization by steering the recommendation behavior and thus, include aspects of the *interaction as control* concept. These user-initiated manipulations dynamically influence the recommendations and serve as a feedforward mechanism. However, users' focus is not about reaching an envisioned end state of AI behavior, but generating useful insights about the domain (or themselves). For example, O'Donovan et al. [71] present *PeerChooser*, an interactive movie recommender that enables users to provide "hints" about their current mood and needs by dragging movie genres closer or further away from their avatar. Bostandjiev et al. [11] use the XUI to explain a music recommendation process and to elicit preferences from users. Users can interactively adjust weights on the input and model level to explore the recommender. Chen et al. [17] present a preference-based recommender to increase users' product knowledge of high-investment products, such as digital cameras and laptops. Their XUI textually explains trade-offs within a set of recommended items.

4.7 Interaction as Embodied Action

This concept centers around collaboration and joint action with a computer. In 1960, Licklider formulated the vision of *man-computer symbiosis* in which

"men and computers [are] to cooperate in making decisions and controlling complex situations" [55]. Humans may be amplified through collaboration with AI. However, effective collaboration goes beyond interaction. In this way, this concept builds on theories from the computer-supported cooperative work (CSCW) community, such as mutual goal understanding, preemptive task co-management and shared progress tracking [96].

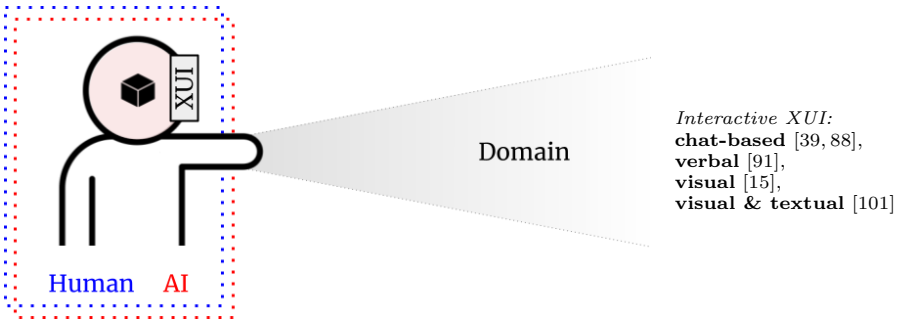

Fig. 7. XAI-interaction as embodied action is about establishing a joint understanding with the AI for an effective collaboration in a given domain.

Transfer to XAI: Applied to XAI, explanations are a crucial component for effective cooperation. A lack of explanatory communication resulted in dissatisfaction [38, 72]. In this way, XUIs contribute to the augmentation of human actions. A symbiotic relationship for which this is especially important involves *autonomous systems*. Autonomous systems in high-risk scenarios have a high degree of autonomy and thus *"need to explain what they are doing and why"* [39]. In such a setting, it is crucial for humans and agents alike to communicate each other's capabilities and intended next steps with respect to a common goal, often in real-time. We identified XUIs which are not only about understanding AI agents (*interaction as transmission*), but which enabled them to also influence the agents' actions – and vice versa [15, 39, 80]. Unlike *interaction as control* the interaction is not only driven by the human controller, but by both parties [6, 91, 101].

Examples: Tabrez et al. [91] present an AI agent that analyzes the game decisions of a human collaborator in a collaborative game setting and verbally interrupts the human in case the common goal becomes unattainable because of a wrong move. The AI agent dynamically constructs a *theory of mind* of the human collaborator and provides tailored explanations that aim to correct their understanding of the game situation. Chakraborti et al. [15] present an XUI that coordinates mission plans between a semi-autonomous search and rescue robot and a human commander who has an incomplete and possibly outdated map of the robot's environment. Visual explanations are embedded as changes in the commander map. The commander can either request (i) an optimal plan by the robot and explanations for this plan, or (ii) a potentially suboptimal

plan that is aligned with the commander's expectations. As such, the XUI reconciles potential mismatches about the plans between robot and commander. Hastie et al. [39] and Robb et al. [80] present an XUI that provides operators of autonomous underwater vehicles with why and why not explanations in real-time via a chat interface. Further, users can influence actions of the autonomous system through the XUI (e.g. setting reminders). Their XUI was reported to increase the situation awareness of operators and adjusted their mental model of system capabilities.

Table 1. Surveyed XAI publications categorized according to the different concepts of interaction by Hornbæk and Oulasvirta [42].

Interaction Concept	**Interaction Goal** *applied to XAI*	**References**
Transmission	Present users with accurate or complete explanation about AI behavior. *Explanatory goal: transparency*	[4, 9, 14, 23, 25, 26, 30, 48, 56, 68, 106]
Dialogue	Facilitate natural and iterative conversation about AI behavior. *Explanatory goals: transparency, scrutability*	[10, 18, 28, 46]
Control	Support rapid convergence towards desired AI behavior. *Explanatory goal: effectiveness*	[34, 41, 44, 45, 51, 52, 75, 78, 108]
Experience	Manage expectations about AI behavior. *Explanatory goals: satisfaction, trust, persuasiveness*	[31, 49, 77, 79, 93, 107]
Optimal Behavior	Adjust human behavior despite limitations of fully understanding the AI behavior. *Explanatory goal: efficiency*	[2, 12, 13, 16, 19, 24, 50, 59, 61, 60, 70, 81, 97–99]
Tool Use	Facilitate learning from AI behavior about a given domain. *Explanatory goals: effectiveness*	[8, 11, 17, 22, 71, 32, 40, 58, 63, 66, 76, 82, 92, 94, 103, 104, 109]
Embodied Action	Establish a joint understanding with the AI for an effective collaboration in a given domain. *Explanatory goal: effectiveness*	[15, 39, 80, 88, 91, 101]

5 Design Principles for Interactive XUI

In the last section, we described the general interplay between the XAI system and the user. Below, we will focus on the interactive qualities of the XUI itself. Vilone et al. define interactivity as *"the capacity of an explanation system to reason about previous utterances both to interpret and answer users' follow-up questions"* [95]. We expand this definition by building on the concept of *explanation facilities* that dates to the era of rule-based expert systems. Moore and Paris [64] proposed that a good explanation facility should, among others, fulfill the requirements of *naturalness* (explanations in natural language following a dialogue), *responsiveness* (allow follow-up questions), *flexibility* (make use of multiple explanation methods), and *sensitivity* (provided explanations should be informed by the user's knowledge, goal, context, and previous interaction). We analyzed our sample of XAI publications through the lens of these requirements to answer RQ2 (What design principles for interactive XUIs can be observed?). We found common interaction strategies and design recommendations [17, 45, 80, 104] that address aspects of these requirements. We unify and present them as *design principles*. In interaction design, design principles are *"guidelines for design of useful and desirable products"* [20].

5.1 Complementary Naturalness

> Consider complementing implicit explanations with rationales in natural language.

Why: Implicit visual explanations can accurately depict the inner workings of an AI but are often inaccessible to non-experts. In contrast, rationales in natural language are post-hoc explanations *"that are meant to sound like what a human [explainer] would say in the same situation"* [30]. Relaying facts through text may *"reassure users when system status might be uncertain or [...] obscure"* [80]. Combining visual cues with textual rationales can facilitate understanding and communicative effectiveness [30].

How: Kim et al. [46] outline a method that automatically generates explanations from visualizations through a template-based approach. Robb et al. [80] elaborate design recommendations on how to incorporate chat-based XUI for autonomous vehicle operators. For example, Yu et al. [108] provide users with a switch to change a visual explanation into verbose explicit sentences. Schaekermann et al. [81] complement quantitative low-confidence predictions with arguments in natural language to attract the attention of physicians. Sklar et al. [88] explain the reasoning behind an AI agent's actions through a chat-interface.

5.2 Responsiveness through Progressive Disclosure

> Consider offering hierarchical or iterative functionalities that allow follow-ups on initial explanations.

Why: Prior research indicated that there is a fine line between no explanation and too much explanation [61]. A user's individual need for cognition influences this threshold. Providing overly detailed explanations overwhelms users who may operate on a simpler mental model of the underlying AI.

How: Springer and Whittaker [89] recommend applying the interaction design pattern of *progressive disclosure*. It is about providing users only with high-level information and offering follow-up operations in case they are interested in further details[6] It resembles the *"progressive-step-by-step process"* demanded by [85]. As such, an XUI should (i) provide information on demand, (ii) hierarchically organize explanatory information, and (iii) keep track of the interaction with a user. For example, Millecamp et al. [61] provide a *Why?* button next to a recommendation. Clicking it provides a one-dimensional visual explanation in the form of a bar chart. If users are interested in additional details, they can click another button to receive a multi-dimensional visual explanation that compares multiple attributes of multiple recommendations in the form of a scatter plot. Krause et al. [52] use tooltips to summarize the most influential features and their sensitivity. If interested, users can drill down and freely explore these with partial dependence plots. Bock et al. [10] visualize a convolutional neural network in virtual reality. Progressive disclosure is realized through spatial distance. As the user approaches the network, more layers with finer granularity become visible. This design principle can also be implicitly implemented by enabling users to repeatedly adjust controls of the ML model [11, 108] or input parameters [18, 76] to progressively disclose local insights step-by-step.

5.3 Flexibility through Multiple Ways to Explain

> Consider offering multiple explanation methods and modalities to enable explainees to triangulate insights.

Why: Humans gain understanding in many ways. Paez [74] outlines them along a spectrum between understanding why (gained through observations and exemplifications) and objectual understanding (gained through idealizations and simplified models). In practice, there is often no best way to explain. For instance, a physician's *"differential diagnosis seldom relies on a single type of data"* [103]. In this way, explanation methods and modalities can complement each other.

How: This principle builds around the interaction design pattern of *multiple ways*[7] , which is about *"providing an opportunity to navigate [...] in more than one manner"*. Multiple publications recommend addressing local and global explanation paradigms within one XUI [24, 41, 104]. This enables users to get an overview of the overall AI behavior and scrutiny of individual cases at the same time. To facilitate this navigation, Liao et al. [54] present a catalog of natural

[6] nngroup.com/articles/progressive-disclosure/
[7] w3.org/tr/understanding-wcag20/navigation-mechanisms-mult-loc.html

language questions that can technically be answered by current XAI methods. Covering multiple of them under a *"holistic approach"* [54] allows users to triangulate insights. For example, Xie et al. [103] present a three-stage explanation workflow that supports physicians in top-down or bottom-up reasoning. Their XUI can *"connects the dots"* and highlight how explanations at each stage relate to one another. Wang et al. [97] present a XUI that provides feature attributions and counterfactual rules in parallel to support multiple ways of reasoning. Hohman et al. [41] provide highly interconnected visual model-level and instance-level explanations side by side to *"flexibly support people's differing processes"*. Chen et al. [17] provide different explanatory views that allow users to examine recommended products from different angles.

5.4 Sensitivity to the Mind and Context

> Consider offering functionalities to adjust explanations to explainees' mental models and contexts.

Why: Explanation needs of user evolve *"as one builds understanding and trust during the interaction process"* [54]. Further, prior beliefs and biases of users influence how they respond to different styles of explanations. This calls for *"a personalized approach to explaining ML systems"* [23].

How: This principle builds around the concept of *mixed-initiative interaction* [43], which emphasizes an interaction in which the human and the computer work towards the shared goal – fostering human understanding in the case of XAI. The timing of actions along the stages of grounding, listening, and interrupting is important for a successful interaction. To adapt its operations, an XUI needs to construct a computer model (or theory of mind [91]) of the user's mental model [65]. Despite its complexity, we found first examples. Tabrez et al. [91] estimate a human collaborator's beliefs in a collaborative game to identify explanation points. Other works [15, 19, 17], elicit preferences or beliefs to estimate a user's expected AI predictions (so called foils), e.g., so that counterfactual explanations can argument only regarding these. Wenskovitch et al. [100] present a method to infer user intent from interactions with visual explanations. Xie et al. [104] implement an *"urgent"* mode that can be toggled by physicians in a hurry to only see high confidence explanations with little system complexity.

6 Limitations and Outlook

Our review excluded publications outside the *ACM Digital Library* and the *SIGCHI* community. We are confident that our review covers many publications that emphasize the interaction design perspective of XAI. However, we probably have missed relevant applied research from adjacent XAI communities inside (e.g., FAccT) and outside (e.g., AIS) of *ACM*. Future work could extend our work with their learnings. Another promising direction for future research is constructive research that encompasses all presented design principles.

None of the survey publications considered all design principles in one XUI. This makes sense as researchers try to limit and control variables for a rigorous evaluation of their research questions. However, with the emergence of open-source explanation-generating toolkits it would be a logical next step to explore reusable and customizable XUI frameworks. These could integrate multiple explanation methods under a human-centric interaction concept.

7 Summary

Interaction design has been discussed as an important aspect for effective explainability in XAI. Yet, so far, it has not been systematically analyzed. Starting from a systematically obtained set of XAI publications that mention user interfaces or user interaction, we derived seven concepts of human-XAI interaction. Further, we analyzed the presented XUI and consolidated proposed recommendations as design principles encompassing four recurring themes: naturalness, responsiveness, flexibility, and sensitivity. We contribute a categorization to describe XAI work not only by the intended target audience or domain of application, but also through the pursued interaction concept. Our survey provides a starting point for researchers and practitioners planning and designing human-centric XAI systems.

References

1. Abdul, A., Vermeulen, J., Wang, D., Lim, B.Y., Kankanhalli, M.: Trends and Trajectories for Explainable, Accountable and Intelligible Systems. CHI '18 (2018)
2. Abdul, A., von der Weth, C., Kankanhalli, M., Lim, B.Y.: COGAM: Measuring and Moderating Cognitive Load in ML Model Explanations. CHI '20 (2020)
3. Adadi, A., Berrada, M.: Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access (2018)
4. Alqaraawi, A., Schuessler, M., Weiss, P., Costanza, E., Berthouze, N.: Evaluating Saliency Map Explanations for Convolutional Neural Networks. IUI '20 (2020)
5. Amershi, S., et al.: Guidelines for human-AI interaction. CHI'19 (2019)
6. Andres, J., et al.: Introducing Peripheral Awareness as a Neurological State for Human-Computer Integration. CHI'20 (2020)
7. Arya, V., et al.: One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques. arXiv (2019)
8. Barria-Pineda, J., Brusilovsky, P.: Explaining Educational Recommendations through a Concept-Level Knowledge Visualization. IUI '19 (2019)
9. Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., Shadbolt, N.: 'It's Reducing a Human Being to a Percentage'. CHI '18 (2018)
10. Bock, M., Schreiber, A.: Visualization of Neural Networks in Virtual Reality Using Unreal Engine. VRST '18 (2018)
11. Bostandjiev, S., O'Donovan, J., Höllerer, T.: TasteWeights: A Visual Interactive Hybrid Recommender System. RecSys '12 (2012)
12. Buçinca, Z., Lin, P., Gajos, K.Z., Glassman, E.L.: Proxy Tasks and Subjective Measures Can Be Misleading in Evaluating XAI Systems. IUI '20 (2020)

13. Bunt, A., Lount, M., Lauzon, C.: Are Explanations Always Important? A Study of Deployed, Low-Cost Intelligent Interactive Systems. IUI '12 (2012)
14. Cai, C.J., Jongejan, J., Holbrook, J.: The Effects of Example-Based Explanations in a Machine Learning Interface. IUI '19 (2019)
15. Chakraborti, T., Sreedharan, S., Grover, S., Kambhampati, S.: Plan Explanations as Model Reconciliation: An Empirical Study. HRI '19 (2019)
16. Chen, L.: Adaptive Tradeoff Explanations in Conversational Recommenders. RecSys '09 (2009)
17. Chen, L., Wang, F.: Explaining Recommendations Based on Feature Sentiments in Product Reviews. IUI '17 (2017)
18. Cheng, H.F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F.M., Zhu, H.: Explaining Decision-Making Algorithms through UI. CHI '19 (2019)
19. Chromik, M., Fincke, F., Butz, A.: Mind the (Persuasion) Gap: Contrasting Predictions of Intelligent DSS with User Beliefs. EICS '20 Companion (2020)
20. Cooper, A., Reimann, R., Cronin, D.: About Face 3: The Essentials of Interaction Design (2007)
21. Danilevsky, M., Qian, K., Aharonov, R., Katsis, Y., Kawas, B., Sen, P.: A Survey of the State of Explainable AI for Natural Language Processing. arXiv (2020)
22. Das, D., Chernova, S.: Leveraging Rationales to Improve Human Task Performance. IUI '20 (2020)
23. Dodge, J., Liao, Q.V., Zhang, Y., Bellamy, R.K.E., Dugan, C.: Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment. IUI '19 (2019)
24. Dodge, J., Penney, S., Hilderbrand, C., Anderson, A., Burnett, M.: How the Experts Do It: Assessing and Explaining Agent Behaviors in Real-Time Strategy Games. CHI '18 (2018)
25. Dominguez, V., Messina, P., Donoso-Guzmán, I., Parra, D.: The Effect of Explanations and Algorithmic Accuracy on Visual Recommender Systems of Artistic Images. IUI '19 (2019)
26. Donkers, T., Kleemann, T., Ziegler, J.: Explaining Recommendations by Means of Aspect-Based Transparent Memories. IUI '20 (2020)
27. Doshi-Velez, F., Kim, B.: Towards A Rigorous Science of Interpretable Machine Learning. arXiv (2017)
28. Douglas, N., Yim, D., Kartal, B., Hernandez-Leal, P., Maurer, F., Taylor, M.E.: Towers of Saliency: A Reinforcement Learning Visualization Using Immersive Environments. ISS '19 (2019)
29. Dudley, J.J., Kristensson, P.O.: A Review of User Interface Design for Interactive Machine Learning. ACM Trans. Interact. Intell. Syst. (2018)
30. Ehsan, U., Tambwekar, P., Chan, L., Harrison, B., Riedl, M.O.: Automated Rationale Generation: A Technique for Explainable AI and Its Effects on Human Perceptions. IUI '19 (2019)
31. Eiband, M., Buschek, D., Kremer, A., Hussmann, H.: The Impact of Placebic Explanations on Trust in Intelligent Systems. CHI EA '19 (2019)
32. Feng, S., Boyd-Graber, J.: What Can AI Do for Me? Evaluating Machine Learning Interpretations in Cooperative Play. IUI '19 (2019)
33. Ferreira, J.J., Monteiro, M.S.: What Are People Doing About XAI User Experience? A Survey on AI Explainability Research and Practice. LNCS '20 (2020)
34. Fulton, L.B., Lee, J.Y., Wang, Q., Yuan, Z., Hammer, J., Perer, A.: Getting Playful with Explainable AI. CHI EA '20 (2020)
35. Garg, N., Schiebinger, L., Jurafsky, D., Zou, J.: Word embeddings quantify 100 years of gender and ethnic stereotypes (2018)

36. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. ACM Surveys (2018)
37. Gunning, D.: DARPA's XAI Program. IUI '19 (2019)
38. Guzdial, M., Liao, N., Chen, J., Chen, S.Y., Shah, S., Shah, V., Reno, J., Smith, G., Riedl, M.O.: Friend, Collaborator, Student, Manager: How Design of an AI-Driven Game Level Editor Affects Creators. CHI '19 (2019)
39. Hastie, H., Chiyah Garcia, F.J., Robb, D.A., Laskov, A., Patron, P.: MIRIAM: A Multimodal Interface for Explaining the Reasoning Behind Actions of Remote Autonomous Systems. ICMI '18 (2018)
40. Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining Collaborative Filtering Recommendations. CSCW '00 (2000)
41. Hohman, F., Head, A., Caruana, R., DeLine, R., Drucker, S.M.: Gamut: A Design Probe to Understand How Data Scientists Understand Machine Learning Models. CHI '19 (2019)
42. Hornbaek, K., Oulasvirta, A.: What Is Interaction? CHI '17 (2017)
43. Horvitz, E.: Principles of Mixed-Initiative User Interfaces. CHI '99 (1999)
44. Ishibashi, T., Nakao, Y., Sugano, Y.: Investigating Audio Data Visualization for Interactive Sound Recognition. IUI '20 (2020)
45. Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., Wortman Vaughan, J.: Interpreting Interpretability. CHI '20 (2020)
46. Kim, D.H., Hoque, E., Agrawala, M.: Answering Questions about Charts and Generating Visual Explanations. CHI '20 (2020)
47. Kitchenham, B., Charters, S.: Guidelines for performing Systematic Literature Reviews in Software Engineering (2007)
48. Kleinerman, A., Rosenfeld, A., Kraus, S.: Providing Explanations for Recommendations in Reciprocal Environments. RecSys '18 (2018)
49. Knijnenburg, B.P., Bostandjiev, S., O'Donovan, J., Kobsa, A.: Inspectability and Control in Social Recommenders. RecSys '12 (2012)
50. Kocaballi, A.B., Coiera, E., Berkovsky, S.: Revisiting Habitability in Conversational Systems. CHI EA '20 (2020)
51. Koch, J., Lucero, A., Hegemann, L., Oulasvirta, A.: May AI? Design Ideation with Cooperative Contextual Bandits. CHI '19 (2019)
52. Krause, J., Perer, A., Ng, K.: Interacting with Predictions: Visual Inspection of Black-Box Machine Learning Models. CHI '16 (2016)
53. Law, E.L.C., Roto, V., Hassenzahl, M., Vermeeren, A.P.O.S., Kort, J.: Understanding, Scoping and Defining User Experience. CHI '09 (2009)
54. Liao, Q.V., Gruen, D., Miller, S.: Questioning the AI: Informing Design Practices for Explainable AI User Experiences. CHI '20 (2020)
55. Licklider, J.: Man-Computer Symbiosis (1960)
56. Lim, B.Y., Dey, A.K.: Weights of Evidence for Intelligible Smart Environments. UbiComp '12 (2012)
57. Linardatos, P., Papastefanopoulos, V., Kotsiantis, S.: Explainable AI: A Review of Machine Learning Interpretability Methods (2020)
58. Ludwig, J., Geiselman, E.: Intelligent Pairing Assistant for Air Operation Centers. IUI '12 (2012)
59. Mai, T., Khanna, R., Dodge, J., Irvine, J., Lam, K.H., Lin, Z., Kiddle, N., Newman, E., Raja, S., Matthews, C., Perdriau, C., Burnett, M., Fern, A.: Keeping It "Organized and Logical". IUI '20 (2020)
60. Mikhail, M., Roegiest, A., Anello, K., Wei, W.: Dancing with the AI Devil: Investigating the Partnership Between Lawyers and AI. CHIIR '20 (2020)

61. Millecamp, M., Htun, N.N., Conati, C., Verbert, K.: To Explain or Not to Explain: The Effects of Personal Characteristics When Explaining Music Recommendations. IUI '19 (2019)
62. Miller, T.: Explanation in Artificial Intelligence: Insights From the Social Sciences. Artificial Intelligence (2019)
63. Misztal-Radecka, J., Indurkhya, B.: Persona Prototypes for Improving the Qualitative Evaluation of Recommendation Systems. UMAP '20 Adjunct (2020)
64. Moore, J.D., Paris, C.: Requirements for an expert system explanation facility. Computational Intelligence (1991)
65. Mueller, S.T., Hoffman, R.R., Clancey, W., Emrey, A., Klein Macrocognition, G.: Explanation in Human-AI Systems. arXiv (2019)
66. Muhammad, K.I., Lawlor, A., Smyth, B.: A Live-User Study of Opinionated Explanations for Recommender Systems. IUI '16 (2016)
67. Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B.: Definitions, Methods, and Applications in Interpretable Machine Learning (2019)
68. Musto, C., Lops, P., de Gemmis, M., Semeraro, G.: Justifying Recommendations through Aspect-Based Sentiment Analysis of Users Reviews. UMAP '19 (2019)
69. Norman, D., Draper, S.: User Centered System Design: New Perspectives on Human-Computer Interaction (1986)
70. Nourani, M., Honeycutt, D.R., Block, J.E., Roy, C., Rahman, T., Ragan, E.D., Gogate, V.: Investigating the Importance of First Impressions and Explainable AI with Interactive Video Analysis. CHI EA '20 (2020)
71. O'Donovan, J., Smyth, B., Gretarsson, B., Bostandjiev, S., Höllerer, T.: PeerChooser: Visual Interactive Recommendation. CHI '08 (2008)
72. Oh, C., Kim, S., Choi, J., Eun, J., Kim, S., Kim, J., Lee, J., Suh, B.: Understanding How People Reason about Aesthetic Evaluations of AI. DIS '20 (2020)
73. Oulasvirta, A., Hornbaek, K.: HCI Research as Problem-Solving. CHI '16 (2016)
74. Páez, A.: The Pragmatic Turn in Explainable Artificial Intelligence (XAI). Minds and Machines (2019)
75. Patel, K., Bancroft, N., Drucker, S.M., Fogarty, J., Ko, A.J., Landay, J.: Gestalt: Integrated Support for Implementation and Analysis in ML. UIST '10 (2010)
76. Paudyal, P., Banerjee, A., Gupta, S.: On Evaluating the Effects of Feedback for Sign Language Learning Using Explainable AI. IUI '20 (2020)
77. Pilling, F., Akmal, H., Coulton, P., Lindley, J.: The Process of Gaining an AI Legibility Mark. CHI EA '20 (2020)
78. Poltrock, S.E., Steiner, D.D., Tarlton, P.N.: Graphic Interfaces for Knowledge-Based System Development (1986)
79. Pu, P., Chen, L.: Trust Building with Explanation Interfaces. IUI '06 (2006)
80. Robb, D.A., Lopes, J., Padilla, S., Laskov, A., Chiyah Garcia, F.J., Liu, X., Scharff Willners, J., Valeyrie, N., Lohan, K., Lane, D., Patron, P., Petillot, Y., Chantler, M.J., Hastie, H.: Exploring Interaction with Remote Autonomous Systems Using Conversational Agents. DIS '19 (2019)
81. Schaekermann, M., Beaton, G., Sanoubari, E., Lim, A., Larson, K., Law, E.: Ambiguity-Aware AI Assistants for Medical Data Analysis. CHI '20 (2020)
82. Schneeberger, T., Gebhard, P., Baur, T., André, E.: PARLEY: A Transparent Virtual Social Agent Training Interface. IUI '19 (2019)
83. Schuessler, M., Wei\ss{}, P.: Minimalistic Explanations: Capturing the Essence of Decisions. CHI EA '19 (2019)
84. Shannon, C.E.: A mathematical theory of communication. The Bell system technical journal (1948)

85. Shneiderman, B.: Bridging the Gap Between Ethics and Practice. ACM Transactions on Interactive Intelligent Systems (2020)
86. Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., Diakopoulos, N.: Confessions: Grand challenges for HCI researchers. Interactions (2016)
87. Simon, H.A.: Models of bounded rationality: Empirically grounded economic reason, vol. 3. MIT press (1997)
88. Sklar, E.I., Azhar, M.Q.: Explanation through Argumentation. HAI '18 (2018)
89. Springer, A., Whittaker, S.: Progressive Disclosure. ACM Transactions on Interactive Intelligent Systems (2020)
90. Stolterman, E., Wiltse, H., Chen, S., Lewandowski, V., Pak, L.: Analyzing artifact interaction complexity (2012)
91. Tabrez, A., Agrawal, S., Hayes, B.: Explanation-Based Reward Coaching to Improve Human Performance via Reinforcement Learning. HRI '19 (2019)
92. Tintarev, N.: Explanations of Recommendations. RecSys '07 (2007)
93. Tsai, C.H., Brusilovsky, P.: Evaluating Visual Explanations for Similarity-Based Recommendations: User Perception and Performance. UMAP '19 (2019)
94. Vig, J., Sen, S., Riedl, J.: Tagsplanations: Explaining Recommendations Using Tags. IUI '09 (2009)
95. Vilone, G., Longo, L.: Explainable Artificial Intelligence: a Systematic Review. arXiv (2020)
96. Wang, D., Churchill, E., Maes, P., Fan, X., Shneiderman, B., Shi, Y., Wang, Q.: From Human-Human Collaboration to Human-AI Collaboration. CHI EA '20 (2020)
97. Wang, D., Yang, Q., Abdul, A., Lim, B.Y.: Designing Theory-Driven User-Centric Explainable AI. CHI '19 (2019)
98. Wang, N., Pynadath, D.V., Hill, S.G.: Trust Calibration within a Human-Robot Team: Comparing Automatically Generated Explanations. HRI '16 (2016)
99. Weisz, J.D., Jain, M., Joshi, N.N., Johnson, J., Lange, I.: BigBlueBot: Teaching Strategies for Successful Human-Agent Interactions. IUI '19 (2019)
100. Wenskovitch, J., Dowling, M., North, C.: With Respect to What? Simultaneous Interaction with Dimension Reduction and Clustering Projections. IUI '20 (2020)
101. Wiegand, G., Schmidmaier, M., Weber, T., Liu, Y., Hussmann, H.: I Drive - You Trust: Explaining Driving Behavior Of Autonomous Cars. CHI EA '19 (2019)
102. Wolf, C.T.: Explainability Scenarios: Towards Scenario-Based XAI Design. IUI '19 (2019)
103. Xie, J., Myers, C.M., Zhu, J.: Interactive Visualizer to Facilitate Game Designers in Understanding Machine Learning. CHI EA '19 (2019)
104. Xie, Y., Chen, M., Kao, D., Gao, G., Chen, X.A.: CheXplain: Enabling Physicians to Explore and Understand Data-Driven Medical Imaging Analysis. CHI '20 (2020)
105. Xu, W.: Toward human-centered AI: A perspective from human-computer interaction. Interactions (2019)
106. Yang, F., Huang, Z., Scholtz, J., Arendt, D.L.: How Do Visual Explanations Foster End Users' Appropriate Trust in Machine Learning? IUI '20 (2020)
107. Yin, M., Wortman Vaughan, J., Wallach, H.: Understanding the Effect of Accuracy on Trust in Machine Learning Models. CHI '19 (2019)
108. Yu, B., Yuan, Y., Terveen, L., Wu, Z.S., Forlizzi, J., Zhu, H.: Keeping Designers in the Loop: Communicating Inherent Algorithmic Trade-Offs Across Multiple Objectives. DIS '20 (2020)
109. Zanker, M.: The Influence of Knowledgeable Explanations on Users' Perception of a Recommender System. RecSys '12 (2012)