XAIR: A Framework of Explainable AI in Augmented Reality

Xuhai Xu
Meta Reality Labs & UW
xuhaixu@uw.edu

Mengjie Yu
Meta Reality Labs
annaymj@meta.com

Tanya Jonker
Meta Reality Labs
tanya.jonker@meta.com

Kashyap Todi
Meta Reality Labs
kashyap.todi@gmail.com

Feiyu Lu
Meta Reality Labs & VT
feiyulu@vt.edu

Xun Qian
Meta Reality Labs & Purdue
qian85@purdue.edu

João Belo
Meta Reality Labs & Aarhus Univ
joaobelo@cs.au.dk

Tianyi Wang
Meta Reality Labs
tianyiwang@meta.com

Michelle Li
Meta Reality Labs
michelleli@meta.com

Aran Mun
Meta Reality Labs
aranmun@meta.com

Te-Yen Wu
Meta Reality Labs & Dartmouth
Te-yen.Wu.GR@dartmouth.edu

Junxiao Shen
Meta Reality Labs & Cambridge
js2283@cam.ac.uk

Ting Zhang
Meta Reality Labs
tingzhang@meta.com

Narine Kokhlikyan
Meta Reality Labs
narine@meta.com

Fulton Wang
Meta Reality Labs
fultonwang@meta.com

Paul Sorenson
Meta Reality Labs
pfsorenson52@meta.com

Sophie Kahyun Kim
Meta Reality Labs
sophiekkim@meta.com

Hrvoje Benko
Meta Reality Labs
benko@meta.com

ABSTRACT

Explainable AI (XAI) has established itself as an important component of AI-driven interactive systems. With Augmented Reality (AR) becoming more integrated in daily lives, the role of XAI also becomes essential in AR because end-users will frequently interact with intelligent services. However, it is unclear how to design effective XAI experiences for AR. We propose XAIR, a design framework that addresses *when*, *what*, and *how* to provide explanations of AI output in AR. The framework was based on a multi-disciplinary literature review of XAI and HCI research, a large-scale survey probing 500+ end-users' preferences for AR-based explanations, and three workshops with 12 experts collecting their insights about XAI design in AR. XAIR's utility and effectiveness was verified via a study with 10 designers and another study with 12 end-users. XAIR can provide guidelines for designers, inspiring them to identify new design opportunities and achieve effective XAI designs in AR.

KEYWORDS

Explainable AI, Augmented Reality, Design Framework

1 INTRODUCTION

Breakthroughs in Artificial Intelligence (AI) and Machine Learning (ML) have considerably advanced the degree to which interactive systems can augment our lives [127, 182]. As black-box ML models are increasingly being employed, concerns about humans misusing AI and losing control have led to the need to make AI and ML algorithms easier for users to understand [25, 155]. This, in turn, has spurred rapidly growing interest into *Explainable AI (XAI)* within academia [11, 74, 130] and industry [2, 6, 22], and by regulatory entities [1, 93, 94]. Earlier XAI research aims to help AI/ML developers on model debugging (*e.g.*, [106, 140, 178, 179, 240]) or assist domain experts such as clinicians by revealing more information such as causality and certainty (*e.g.*, [75, 129, 217, 222]). Recently, there has been a growing amount of XAI research focusing on the non-expert end-users [31, 74, 102]. Existing studies have found that XAI can help end-users resolve confusion and build trust [67, 170]. Industrial practitioners have started to integrate XAI into everyday scenarios and improve user experiences, *e.g.*, by displaying the match rate of point-of-interest suggestions on map applications [135].

Alongside the surge of interest into XAI, *Augmented Reality (AR)* is another technology making its way into everyday living [5, 8].
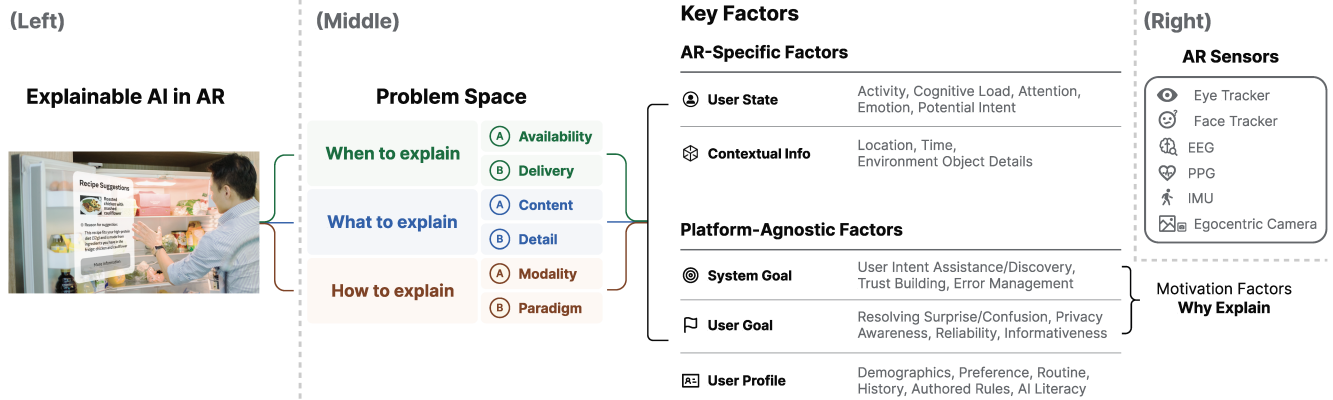
Figure 1: An Overview of XAIR Framework. (Left) An example of the AR interface with explanations. (Middle) The main structure of XAIR: the problem space and the key factors. (Right) Sensors that are integrated into AR.

Advances in more lightweight, powerful, and battery-efficient Head-Mounted Displays (HMDs) have brought us closer to the vision of pervasive AR [91]. As AI techniques are needed to enable context-aware, intelligent, everyday AR [12, 58, 177], XAI will be essential because end-users will interact with outcomes of AI systems. XAI could be used to make intelligent AR behavior interpretable, resolve confusion or surprise when encountering unexpected AI outcomes, promote privacy awareness, and build trust. Therefore, we aim to answer the following research question: **How do we create effective XAI experiences for AR in everyday scenarios?**

Researchers have developed several design spaces and frameworks to guide the design of XAI outside the context of AR [25, 67, 155, 217]. However, most previous work focused on identifying a taxonomy of explanation types or generation techniques. They did not consider everyday AR-specific factors such as the rich sensory information that AR technologies have about users and contexts, and its always-on, adaptive nature. These factors can not only support more personalized explanations, but also affect the design of an explanation interface. For example, one could render in-place explanations on related objects (*e.g.*, explaining a recipe recommendation by highlighting ingredients in the fridge). In this paper, we provide a framework to guide the design of XAI in AR.

To answer the aforementioned research question, a design space analysis [149] was used to break down the main research question into three sub-questions: 1) *When to explain?*, 2) *What to explain?*, and 3) *How to explain?* Previous research from the XAI and HCI communities has focused on one or two of these sub-questions (*e.g.*, when [164, 183], what [25, 130]). Although not within the context of AR, many of these findings can inform the design of XAI in AR. Therefore, we first summarized related literature to identify the most important dimensions under each sub-question, as well as the factors that determine the answers to these questions, such as users' goals for having explanations (*i.e.*, why explain). Then, we conducted two complementary studies to obtain insights from the perspectives of end-users and experts. Specifically, we carried out a large-scale survey including over 500 end-users with different levels of knowledge of AI to collect user preferences about the timing (related to *When*), content (related to *What*), and modality (related to *How*) of explanations in multiple AR scenarios. In addition, we ran three workshops with twelve experts (*i.e.*, four experts

per workshop) from different backgrounds, including algorithm developers, designers, UX professionals, and HCI researchers to iterate on the dimensions and generate guidelines to answer the *When/What/How* questions.

Merging the insights obtained from these two studies, we developed the **XAIR** (e**X**plainable **AI** for Augmented **R**eality) framework (Fig. 1). The framework can serve as a comprehensive reference that connects multiple disciplines across XAI and HCI. It also provides a set of guidelines to assist in the development of XAI designs in AR. XAI researchers and designers can use the guidelines to enhance their design intuition and propose more effective and rigorous XAI designs for AR scenarios.

We further conducted two user studies to evaluate XAIR. To verify its utility to support designers, the first study focused on designers' perspectives. Ten designers were invited to use XAIR and design XAI experiences for two real-life AR scenarios. To demonstrate its effectiveness in guiding the design of an actual AR system, a second study was conducted from the perspective of end-users. We implemented a real-time intelligent AR system based on the designers' proposals in the previous study using XAIR. The study measured the usability of the AR system with 12 end-users. The results indicated that XAIR could provide meaningful and insightful support for designers to propose effective designs for XAI in AR, and that XAIR could lead to an easy-to-use AR system that was transparent and trustworthy.

The contributions of this research are:

- We summarized literature from multiple domains and identified the important dimensions for the when/what/how questions in the problem space when designing XAI in AR.

- Drawing the results from a large-scale survey with over 500 users and an iterative workshop study with 12 experts, we developed XAIR, the first framework for XAI design in AR scenarios. We also proposed a set of guidelines to support designers in their design thinking process.

- The results of design workshops with 10 designers indicated that XAIR could provide meaningful and insightful creativity support for designers. The study with 12 end-users who used a real-time AR system showed that XAIR led to the design of AR systems that were transparent and trustworthy.

2 BACKGROUND

In this section, we first introduce more background about XAI (Sec. 2.1). We then summarize existing XAI design frameworks and demonstrate the need for a new XAI framework that is specifically applicable to AR scenarios (Sec. 2.2).

2.1 What is XAI?

The notion of XAI can be traced back more than four decades [223], where expert systems would explain output via a set of decision rules [195, 207]. This concept has been brought back into focus by the success of black-box AI/ML models [59]. The working definition of XAI used in this paper is: *"given an audience, an explainable AI is one that produces details or reasons to make its functioning clear or easy to understand"* [25].

With the increasing prevalence of advanced black-box models that make more critical predictions and decisions, the interpretability and transparency of AI systems has attracted increasing attention from various academic, industrial, and regulatory stakeholders [1, 90, 94, 169]. Addressing the broad vision of making AI more understandable for humans involves multidisciplinary research efforts. ML researchers have developed algorithms that results in transparent models (*e.g.*, decision trees, Bayesian models [47, 125]) or used post-hoc explanation techniques (*e.g.*, feature importance, visual explanation, [146, 196, 199]) to generate explanations for users. HCI researchers, on the other hand, have focused on improving user trust [98, 170] and understanding [132, 134] of machine generated explanations. Psychology researchers have approached XAI from a more fundamental perspective and studied how people generate, communicate, and understand explanations [209, 234].

By providing more transparency and interpretability, XAI can offer different target audiences different benefits. For instance, for algorithm developers and data scientists, XAI can provide more details for model debugging and improvement [137] and increase production efficiency and robustness [188, 237]. For domain experts, XAI can reveal insights about causality [143], transferability [49, 210], confidence [27, 168], and also enhance the reliability of model output [70, 178, 225, 226]. Early XAI research only focused on these two groups of users. Recently, there have been an increasing number of XAI studies that have focused on non-expert end-users who represent a large potential audience of XAI [74, 102]. XAI has been found to improve reliance and build trust with non-experts [170], especially when users encounter unexpected AI outcomes [67], have privacy concerns [72], or seek additional information [42, 69]. Some companies have integrated XAI into products used by the general population [135, 242], *e.g.*, visualizing the match rate of restaurant suggestions in a map application [135] or showing reasons for product recommendations on a shopping website [242]. However, these efforts are still at an early stage.

2.2 Why do we need XAI in Everyday AR?

Since the first AR HMD was built in 1968 [206], researchers and engineers have been striving to integrate AR HMDs into everyday living. Recent examples include simple head-mounted cameras and displays (*e.g.*, Google Glass Enterprise [5] and Snap Spectacles [10]), as well as more advanced HMDs with 3D-depth sensing

(*e.g.*, Microsoft Hololens [8] and Magic Leap [3]). As hardware improves, it is foreseeable that AR will become an integral aspect of everyday living for general consumers and support a wide range of applications in the near future [58, 177].

2.2.1 The Importance of AI and XAI in AR. The role of AI will be critical for AR devices if they are to provide intelligent services. The integration of sensors enables AR systems to understand users' current states [99, 194, 204] and their environment [139, 153] to provide a variety of intelligent functionalities. For example, AR could infer user intent [14] and provide contextual recommendations for daily activities (*e.g.*, recipe recommendations when a user opens the fridge during lunch) [15, 118, 122]. The rich interaction between the outcomes of AI and end-users requires effectively designed XAI that can support users in a variety of contexts, such as when users are confused or surprised while encountering an unexpected AI outcome, or when they want to make sure that an AI outcome is reliable and trustworthy [17, 155]. Recent work has started to explore the application of XAI in AR [16]. For instance, Wintersberger *et al.* found that showing traffic-related information in AR while driving can provide much needed explanation to users and improve user trust [220]. Danry *et al.* explored the use of an explainable AI assistant integrated within wearable glasses to enhance human rationality [63]. Zimmermann *et al.* found that introducing XAI during an AR-based shopping process could improve user experiences [247]. However, these studies proposed their own case-by-case XAI designs. In this research, we aggregated the major factors identified in the literature and studied the when/what/how questions systematically.

2.2.2 The Need for A New XAI Framework for AR. Researchers have proposed several XAI design spaces and frameworks for AI systems, *e.g.*, knowledge-based systems [88], decision support systems [18], and recommendation systems [96]. For instance, Wang *et al.* proposed a conceptual framework for building user-centric XAI systems and put it into practice by implementing an explainable clinical diagnostic tool [217]. Eiband *et al.* presented a stage-based participatory design process for designers to integrate transparency into systems [75]. They evaluated the process using a commercial AI fitness coach. Zhu *et al.* proposed a co-creation design space between game designers using ML techniques and investigated the usability of XAI algorithms to support game designers [246]. Liao *et al.* developed an algorithm-informed XAI question bank to support design practices for AI systems [130]. Ehsan *et al.* investigated how social transparency in AI systems supported sellers from technology companies and developed a conceptual framework to address what, who, why, when questions [73]. Wolf proposed the concept of scenario-based XAI design and highlighted researchers' need to understand AI systems in specific scenarios such as when researchers are not uncertain or they want to explain data limitations [221]. These existing frameworks aim to guide XAI design for developers or domain experts for specific applications. Focusing on non-expert end-users, Lim and Dey systematically investigated end-users' opinions and preferences about different types of explanations in multiple context-aware applications, and provided an XAI framework for intelligible context-aware systems [132]. Moreover, recent industry practitioners have also made efforts towards a designing framework for end-user-facing explanations [114].

	Augmented Reality HMD 📹	Existing Non-AR Devices 📱 ⌚ 💻 🖥
👤 **User Status**	Rich *e.g.,* cognitive load, attention	Limited *e.g.,* some activities
🎁 **Contextual Info**	High-resolution *e.g.,* env objects in details	Low-resolution *e.g.,* location
📡 **Interface**	Always-On, Spatially Adaptive World/Object/Body-Based	Episodic, Superficial Device-Based

Figure 2: The Uniqueness of AR that Distinguishes XAI Design from Other Platforms.

Such XAI frameworks focused on the content design of XAI, which is mostly visualized on laptops or mobile phones, thus making them insufficient for the myriad of AR contexts. There are several factors that distinguish AR from other platforms and necessitate the need for a new XAI design framework (see Fig. 2). First, AR has a much deeper real-time understanding of a user's current state via the sensors within an HMD [35, 186]. Second, compared to other platforms, AR systems can develop a more fine-grained understanding of a user's context [62, 89, 139]. This richer information not only provides new types of information that can be integrated into AR-based XAI explanations, but also influences the design of XAI as explanations need to be tailored to a user's state and context. Third, from an interface perspective, the ability to be always-on and 3D-aware enables AR to present information at any time [23, 144, 245], and spatially adapt explanations to the physical world [81, 176]. These factors influence the design of XAI in AR, as they need to be presented to users in an appropriate, efficient, and effective way. Overall, these unique factors demonstrate how current frameworks are insufficient and there is a need for a new XAI framework specifically designed for AR scenarios.

3 XAIR PROBLEM SPACE AND KEY FACTORS

Determining the way to create effective XAI experiences in AI is a complex challenge. Thus, it is important to first identify the problem space to bound the scope of our investigation. We first summarize over 100 papers from the ML and HCI literature to identify the problem space and the main dimensions within each problem (Sec. 3.1). Then, we outline the key factors that determine the answers to the problems (Sec. 3.2).

The problem space and key factors define the structure of XAIR (Fig. 1 middle). In Sec. 4, we present two studies conducted to obtain insights from end-users and expert stakeholders about how to design XAI in AR. Then, combining the structure and insights, we show how these factors are connected with the problem space, and provide design guidelines in Sec. 5.

3.1 Problem Space

Following the design space analysis method [149], the research question was divided into three sub-questions: when to explain, what to explain, and how to explain [76, 159].

3.1.1 When to Explain? The literature review revealed two aspects of "when" that were important to consider: the *availability* of explanations (*i.e.,* whether to prepare explanations?), and the timing of the explanation's *delivery* (*i.e.,* when to show explanations?).

Availability. Previous research has found that to maintain a positive user experience, supporting user agency and control is important during human-AI interaction [46, 121]. Having explanations that are available and accessible is in line with the goal of supporting user agency.

Delivery. With the ability to show information at any time, AR can employ various timing strategies to present explanations. Thus, it is important to find the appropriate method to deliver explanations to users. Generally, there are two approaches, manual-trigger (*i.e.,* initiated by users) and auto-trigger (*i.e.,* initiated by the system) [57, 235]. On the one hand, researchers have found that explanations should not always be presented to users, because they can introduce unnecessary cognitive load and become overwhelming for non-expert end-users [41, 51, 183, 205, 215]. This is especially important in AR, as users' cognitive capacity tends to be limited [40]. Moreover, adopting manual triggers would enable users to choose to see explanations as needed, thus enabling them to exercise agency over their experience [145, 184]. On the other hand, existing findings on just-in-time intelligent systems (*e.g.,* just-in-time recommendations [105, 148] and just-in-time interventions [159, 189, 232]) have suggested that automatically delivering explanations at the right time based on user intent and need (as detected via AR sensing that identifies a user's state and context) can provide a better user experience [32, 152].

3.1.2 What to Explain? The literature review also found two important aspects of "what" to consider: First, the *content* of the explanations (*i.e.,* what type of content to include?). Second, the level of *detail* of the explanations (*i.e.,* how much detail should be explained?).

Content. Previous literature in XAI has identified several explanation content types [25, 155]. The seven types are:

(1) Input/Output. This type explains the details of input (*e.g.,* data sources, coverage, capabilities) or output (*e.g.,* additional details, options that the system could produce) of a model [132, 133].

(2) Why/Why-Not. This type explains the features in the input data [178] or the model logic [180] that have led or not led to a given AI outcome [158] (also known as contrastive explanations). Showing feature importance is another commonly used technique to generate these explanations [48, 190].

(3) How. This type provides a holistic view to explain the overall logic of an algorithm or a model and illustrate how the AI model works. Typical techniques include model graphs [116], decision boundaries [141], or natural language explanations [29].

(4) Certainty. This type describes the confidence level of the model with its input (*e.g.,* for models whose input is not deterministic, explain how accurate the input of the model is) or output (*e.g.,* explain how accurate, or reliable the AI outcomes are) [135, 193]. Scores based on softmax [38] or calibration [167] are commonly used as the confidence/certainty score for ML models.

(5) Example. This type presents similar input-output pairs from a model, *e.g.,* similar input that lead to the same output or similar output examples given the same input [45, 107]. This is also known as the What-Else explanation. Example methods include influence functions [112] and Bayesian case modelling [110].

(6) What-If. This type demonstrates how changing input or applying new input can affect model output [44, 134].

(7) How-To. In contrast to What-If, this type explains how to change input to achieve a target output [130, 217], *e.g.*, how to change the output from X to Y? Common methods for What-If/How-To content include rule generation [92], feature description [214], and input perturbation [241].

Moreover, another aspect that is independent of the explanation content type is global *vs.* local explanations (explaining the general decision-making process *vs.* a single instance) [156]. In general, non-expert end-users were found to prefer local explanations [67, 117].

Detail. Displaying every relevant explanation content type to an end-user can be overwhelming, especially with the limited cognitive capacity they have in AR [26, 40]. Explanations that extend a user's prior knowledge or fulfill their immediate needs should be prioritized [60]. Moreover, previous research has suggested that presenting detailed and personalized explanations is useful for better understanding AI outcomes [78, 101, 113, 192, 228].

Our focus on *content* and *detail* is about choosing appropriate explanation content types and proper levels of detail, but not on picking which techniques to generate explanations. From a technical perspective, there are interpretable models (*i.e.*, the model being transparent, such as linear regression or decision trees) and ad-hoc explainers (*i.e.*, generating explanations for complex, black-box models) [146, 178]. The latter can further be divided into model-specific and model-agnostic explanation methods [25]. We refer readers to other surveys and toolkits for developing or selecting explanation generation algorithms [7, 9, 13, 22, 131].

3.1.3 How to Explain? The last sub-question *how* focuses on the visual representation of the content in AR. Two dimensions emerged from the literature review, *i.e.*, modality and paradigm.

Modality. The multi-modal nature of AR enables it to support AI outcomes via various modalities (*e.g.*, visual, audio, or haptic) [53, 161]. Explanations are hard to convey using modalities with limited bandwidth (*e.g.*, haptic, olfactory, or even gustatory). Therefore, visual and audio are the two major modalities that should be employed for explanations.

Paradigm. If explanations are presented using audio, the design space is relatively limited (*e.g.*, volume, speed). We refer readers to existing literature on audio design (*e.g.*, [83, 109]). The design space of the visual paradigm for explanations, however, is much larger. First, from a formatting perspective, explanation content can be presented in a textual format (*e.g.*, narrative, dialogue) [116, 158], graphical format (*e.g.*, icons, images, heatmaps) [200, 240], or a combination of both. Second, from a pattern perspective, an explanation can be displayed either in an implicit way (*i.e.*, embedded in the environment, such as a boundary highlight of an object) or explicit way (*i.e.*, distinguished from the environment, such as a pop-up dialogue window) [68, 136, 208]. The pattern is closely related to the adaptiveness of the AR interface [61, 217]. With 3D sensing capabilities, the location of an explanation can be body-based (mostly explicit), object-based (implicit or explicit), or world-based (implicit or explicit) [35, 120, 145, 227]. Prior AR research has explored adaptive interface locations [147, 157], *e.g.*, interfaces should be adaptive based on the semantic understanding of the ongoing interaction [54, 171, 185] and ergonomic metrics [79].

3.2 Key Factors

These three questions, and their dimensions, form the overall problem space of XAIR. Another important aspect of XAIR is the factors that determine the answers to these questions. We summarize these factors from two perspectives, one specific to AR platforms (Sec. 3.2.1), and the other agnostic to any platform (Sec. 3.2.2).

3.2.1 AR-Specific Factors. Fig. 2 summarizes the three main features that distinguish AR from other platforms: *User State*, *Contextual Information*, and *Interface*. As *Interface* is an integral property of an AR platform, it remains invariant to external changes. In contrast, the other two aspects are dynamic and would alter the design of XAI in AR.

User State. The sensors that could be integrated within future HMDs would empower an AR system to have a rich, instant understanding of user's state, such as activities (IMU [86, 219], camera [80, 128, 194, 201], microphone [103, 218, 229, 230]), cognitive load (eye tracking [71, 104, 238], EEG [20, 224]), attention (eye tracking [56, 99, 204, 231], IMU [123], EEG [213]), emotion (facial tracking [233, 236], EEG [202, 216]) and potential intent (the fusion of multiple sensors and low-level intelligence [14, 111, 211]). Depending on a user's state, the design of explanations could be different. For example, as identified in previous research on ambient interfaces [82, 164], when users engage in activities with a high cognitive load, explanations should not show up automatically to interrupt them (related to *when*).

Contextual Information. Compared to devices such as smartphones, AR HMDs have more context awareness. Other than having an awareness of location and time [66], an egocentric camera and LiDAR, combined with other sensors (*e.g.*, Bluetooth, WiFi, RFID), can identify details about digital and non-digital objects in the environment [139, 163, 175], and have a better understanding of the semantics of a scene [34, 89, 153, 162]. Such contextual information would also influence the design of XAI. For instance, an explanation visualization about recipe recommendations that appears when users open the fridge may look differently from explanations about podcast suggestions that are shown while driving (related to *how*).

3.2.2 Platform-Agnostic Factors. There are also other factors that are platform agnostic such as the motivation to present explanations (*i.e.*, *why explain?*). We view this factor from two perspectives, one from the system side (*i.e.*, what are the *system's goals* when presenting explanations?), and the other from the non-expert end-user side (*i.e.*, what are *users' goals* when they want to see explanations?) [187]. The *user profile* (*i.e.*, individual details) is another important factor related to personalized explanations [113, 192].

System Goal. Based on prior literature, we summarize four system goals that are desired when an AR system provides explanations for AI outcomes:

(1) User Intent Discovery. When an AI model generates suggestions for a new topic, the system seeks to help users discover new intent [87, 170, 187]. For example, when a user is traveling in a city, the system recommends several attractions and local restaurants to visit. Both the recommendation and explanations help the user explore new things that they were not aware of.

(2) User Intent Assistance. When the target task has been already initiated by users, then the goal of generating AI outcomes and

explanations assists users with existing intent [28, 39, 64]. For instance, when a user is making dinner, intelligent instructions and explanations would suggest alternative ingredients based on what a user has in their space.

(3) Error Management. When a system has low confidence about input/output or makes a mistake, explanations can serve as error management and explain the process so that users can understand where an error comes from if it appears [13, 223], how they might better collaborate with the system [64], or when to adjust their expectation of the system's intelligence [30, 243].

(4) Trust Building. Various studies have found that explanations can help systems build user trust by offering transparency and increasing intelligibility [19, 150, 198]. As a result, users' trust in models leads them to rely on the system [29, 43].

These four types of system goals are not exclusive. A system can seek to achieve multiple goals simultaneously. Depending on the subset of system goals, the appropriate explanation timing and content types can differ [132, 158] (related to *when* and *what*).

User Goal. While a system has varying reasons to provide explanations, end-users also have varying reasons to have explanations. We summarize four types of user goals from literature.

(1) Resolving Confusion/Surprise. Expectation mismatch is one of the main reasons to need explanations [37, 67, 119, 181]. Users can become confused or surprised when AI outcomes are different from what users are expecting, and having explanations can help to resolve concerns [85, 174].

(2) Privacy Awareness. As AI influences more aspects of daily living, concerns about invasion of one's privacy are also growing [151]. Explanations could disclose which data is being used in a model's decision-making process to end-users [65, 77, 172]. Researchers and designers are recommended to follow an existing privacy framework, such as contextual integrity [160], to make privacy explanations more robust.

(3) Reliability. Ensuring the reliability of AI outcomes is essential for non-trivial decision-making processes so that users can rely on a trustworthy system [102, 124, 178], *e.g.*, daily activity recommendations for personal health management or automatic emergency service contacting in safety-threatening incidents.

(4) Informativeness. End-users can be curious about the reason or process behind an AI outcome [97, 126]. Explanations can fulfill users' curiosity by providing more information [33, 115, 172].

Similar to the system goals, these user goals are not exclusive and users can have multiple goals at the same time. Different goals can require different explanation timings and content (*when* and *what*).

User Profile. This factor covers a range of individual details that influence the design of XAI. For example, information such as demographics and user preferences is necessary to generate personalized explanations [84, 113, 192]. End-users' familiarity with system outcomes is related to the need for explanations and *when* to provide them [60]. Users' digital literacy with AI also affects *what* types of explanations are appropriate and would serve users' purposes [74, 114, 142]. Moreover, users may have individual preferences about explanation visualizations, which may be closely related to *how*. This factor takes these considerations into account.

It is worth noting that XAIR is proposed as a design framework. In a context that AR can detect robustly, designers can use the framework to infer end-users' latent factors, such as *User State* and *User Goal*, based on their design expertise [75]. For example, when users are driving (which can be easily detected by AR), designers can assess users' cognitive load to be high (*User State*). For more complex factors such as *User Goal*, designers can propose a set of potential goals in a given scenario and then refer to the framework to propose a set of designs. As sensing and AI technology are maturing, the framework could be coupled with the automatic inference of these factors [14, 86, 111, 211, 233].

4 METHODS

We conducted two studies after outlining the problem space, one from end-users' perspectives (Sec. 4.1), and the other from XAI/design/AR expert stakeholders' perspectives (Sec. 4.2). The findings from the studies are complementary and provided insights that guided the development of the framework.

4.1 Study 1: Large-Scale End-User Survey

In spite of the existing studies on XAI for end-users, it is unclear whether these findings hold for AR scenarios due to the unique features of AR systems. Thus, we conducted a large-scale survey with end-users to collect their preferences on various aspects of XAI experiences for everyday AR.

4.1.1 Participants. We recruited 506 participants from a third-party online user study platform (age 18 - 54, average 37±10), with a balanced gender distribution (Female 260, Male 241, Non-binary 5). Participants' digital literacy with AI varied, thus they were split into six groups: 1) unfamiliar with AI (12.2%, 62), 2) heard of AI but never used AI-based products (23.5%, 119), 3) used AI products occasionally a few times (23.1%, 117), 4) used AI products on a regular basis (12.8%, 65), 5) used AI products frequently (20.0%, 101), and 6) worked on AI products (8.3%, 42). Participants were familiar with the concept of AR. Among these groups, we further randomly sampled 20 participants (age 18 - 53, average 37±9, 11 Female, 9 Male) for a semi-structured interview to collect a more in-depth understanding about their preferences for XAI in AR.

4.1.2 Design and Procedure. We prepared five sets of proof-of-concept descriptions and images with intelligent everyday AR services that represented five scenes in a typical weekday (*i.e.*, one set per scene). They included 1) music recommendations for the morning when users would be brushing their teeth, 2) podcast recommendations for when users would be driving to work, 3) music recommendations for when users would be working out, 4) recipe recommendations for when users would be making dinner, and 5) additional spice recommendations for when users would be making dinner. In this study, we chose recommendations as the main AI service category, since it is arguably one of the most common AI applications in everyday AR [50, 118] and users could easily contextualize these scenes in their mind.

For the AI outcome in each scene, participants were asked whether they wanted explanations (*i.e.*, yes, no, neutral). If their answer was yes, they would be directed to answer when they wanted it (*i.e.*, always/frequent, contextually dependent, rare/never), their preferred

length of explanation (*i.e.*, concise *vs.* detailed) and the presenting modality (*e.g.*, visual, audio, neutral). After viewing these scenes, they were asked to choose the explanation content types that they found useful. Participants were compensated $5 USD for the task.

We randomly sampled 20 respondents who were willing to participate in a one-hour interview about the detailed reasons behind their survey responses. These participants were compensated $10 for the interview. The interviews were video-recorded and manually transcribed. Two researchers collectively summarized and coded the data using a thematic analysis [36]. Specifically, they first met to establish an agreement on the themes and independently coded all the data. Then, they gathered to discuss and refine the coded data to resolve differences. Their inter-rater reliability (κ) was over 90% after the refinement.

4.1.3 Results. The survey found that respondents had specific preferences for the timing, content, and modality of explanations.

Finding 1: Most users wanted explanations of AI outputs in AR. (related to *when - availability*). A large proportion of respondents wanted explanations (89.7%), motivating the need for XAI in everyday AR scenarios (see Fig. 3a). Our findings were consistent with previous work on end-users' needs for XAI outside AR [74, 114]. The results indicated that if respondents had at least heard of AI, they were more likely to express a need for XAI in AR compared to those who were not familiar with AI. Our interviews found that respondents with little knowledge of AI didn't realize what explanations could be used for. Interestingly, around 10% of respondents who worked on AI indicated that they didn't want explanations. Our interviews revealed the main reason being that some users were *"familiar enough... with the algorithm"* (P2).

Finding 2: The majority of users wanted explanations to be occasional and contextual, especially when they saw anomalies (related to *when - delivery*). Although most respondents wanted explanations, only 13.8% indicated that they needed explanations all the time. The majority of respondents (63.4%) preferred for explanations to be presented contextually only when they have the need. The results of the interviews indicated that the need for explanations was mainly in cases where AI outcomes were new or anomalous to respondents. This finding is also in line with previous studies' findings outside AR [67, 102].

Finding 3: Users generally preferred specific types of explanations (related to *what - content*). Four explanation content types stood out as useful: Input/Output (41.5%), How (37.1%), Why/Why-Not (31.6%), and Certainty (30.6%). The first three types were highlighted in previous findings about context-aware systems [132, 133], while the last type has been adopted by industrial practitioners [135, 203]. As shown in Fig. 3c, respondents with more knowledge of AI would prefer having these explanation types more than those with less AI knowledge.

Finding 4: Users found detailed and personalized explanations useful (related to *what - detail*). Although showing more explanation content can introduce additional cognitive costs, 48.3% of respondents reported that they would find detailed explanations with multiple content types to be useful. Moreover, respondents indicated that explanations that included personal preferences would be more convincing, *e.g.*, *"more personable, more upbeat"* (P13). These results suggest that there is a need to provide options to modulate the level of explanation detail (see Sec. 3.1.2) and the *User Profile* factor in the framework.

Finding 5: Users' preferences for modalities depended on the cognitive load in an AR scenario (related to *how - modality*). The five scenes introduced different levels of cognitive load, which led respondents' preferences for XAI modality to vary. We found that for scenes with complex visual stimuli such as driving, respondents tended to prefer audio explanations over visual ones by 40%, as they were *"more easy and convenient"* (P8). This suggests that it is necessary to take modality bandwidths into account when choosing *how* to present XAI in different AR scenarios [40].

Overall, these findings motivated the need for XAI in AR (**Finding 1**). Moreover, these results (**Finding 2-5**) also provided guidance for design XAI for end-users in AR.

4.2 Study 2: Iteration with Expert Workshops

Based on the existing literature and the end-user survey results, we created an early draft of the framework. Since XAIR aims to support designers and researchers during their design process, we utilized our draft within three workshops with expert stakeholders to collect their insights and finalize the framework.
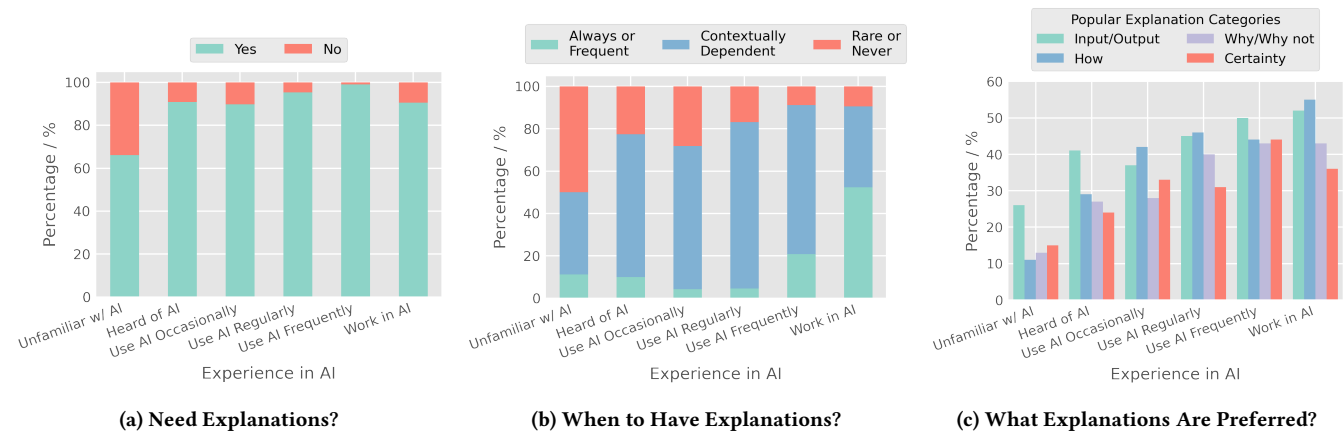

(a) Need Explanations? (b) When to Have Explanations? (c) What Explanations Are Preferred?

Figure 3: Highlight of Survey Results with 506 End-Users about Their Needs and Preferences of XAI in everyday AR scenarios.

4.2.1 Participants. Twelve participants (7 Female, 5 Male, Age 35 ± 6) from a technology company volunteered to participate in the study. They came from four backgrounds, *i.e.*, 3 XAI algorithm developers, 3 designers, 3 UX professionals, and 3 HCI/AR researchers. Participants worked in their domains for at least five years. All participants were familiar with the concept of AI and AR. Participants were randomly assigned into three groups, with each group containing one expert from each domain.

4.2.2 Design and Procedure. We proposed a draft of the framework combining the summary of literature and the results of end-user study. It was an early version of XAIR that is introduced in Sec. 5 and can be found in Appendix A. We also prepared a set of everyday AR scenarios similar to the ones used in the end-user survey (Sec 4.1) to provide more context and stimulate more insights from experts. We utilized a Figma board to show images of the framework and experts could add in-place feedback to different areas of the framework.

We adopted an iterative process using three sequential workshops. All workshops lasted about 90 minutes and were video-recorded. After each workshop, two researchers went through a similar coding and refining process as Sec. 4.1.2, to make sure the result achieved a inter-rater reliability (κ) over 90%. We summarized experts' feedback, iterated on the framework, and presented the new version in the next workshop.

4.2.3 Results. Overall, experts found the framework to be "*useful*" (P2, P6, P7) and that it would "*serve as a very good reference for design*" (P11). Our framework converged as the workshops proceeded, with us receiving rich feedback during the first workshop, and participants in the last workshop only offering small suggestions. We briefly highlight the major comments that were made.

Suggestion 1: Add Missing Pieces. Participants found a few factors missing in the early version of the framework. For example, they pointed out that *User Goal* and *User Profile* needed to be considered for the *what* part, and that the modality of AI output in AR needed to be taken into account for the *how* part. They also provided suggestions on appropriate explanation content types with different system/user goals (*what - content*).

Suggestion 2: Remove Redundancy. Participants also found some parts unnecessarily complex. For example, four experts suggested removing the interface location from *how* part (*i.e.*, where to explain, mentioned in Sec. 3.1.3), because the location needed to be optimized with the whole interface including AI outcomes.

Suggestion 3: Add Default Options. Participants provided advice for default options of different dimensions. For instance, they recommended using the manual-trigger as the default delivery method (*when*) due to users' limited cognitive capacity in AR.

Suggestion 4: Connect across Sub-questions. Participants came to the consensus that the three sub-questions were interwoven. For example, the choice of *what* to explain would influence the design of *how* to explain, and the framework should capture and emphasize such connection.

Suggestion 5: Improve Visual Structure. Finally, participants also offered several suggestions about the visual simplification, clarification, and color choices. The figures in Appendix A show the evolution of the visual structure.

The results of the end-user study and expert workshops are complementary and guided the final version of the framework.

5 XAIR FRAMEWORK

We introduced the structure of XAIR framework in Sec. 3 (*i.e.*, problem space and key factors), and summarized insights from end-users and experts in Sec. 4. Connecting the literature survey and studies' results (**Findings 1-5** in Sec. 4.1.3, and **Suggestions 1-5** in Sec. 4.2.3), we introduce the details of XAIR, identify how the key factors determine the design choices for each dimension in the when/what/how questions, and present a set of guidelines.

5.1 When to Explain?

We first introduce the *when* part and discuss how to make a choice for delivery options. Fig. 4 presents an overview.

5.1.1 Ⓐ *Availability.* The end-user survey results suggested the need for explanations in AR for the majority end-users (**Finding 1**). A system should always generate explanations with AI outcomes and make them accessible for users, so that they can have a better sense of agency whenever they need explanations [121, 138, 239].

G1. Make explanations always accessible to provide user agency.

5.1.2 Ⓑ *Delivery.* Aligned with previous work [40, 100, 165], experts also mentioned the risk of cognitive overload in AR (**Suggestion 3**). The default option should be to wait until users manually request explanations. An example could be a button with an information icon that enables users to click on it to see an explanation.

However, there are cases where automatically presenting just-in-time explanations is beneficial [32, 152]. We summarize the three cases based on our two studies (**Finding 2** about the importance of contextual explanations, and **Suggestion 1** about the need of considering *User Goal* and *User Profile*):
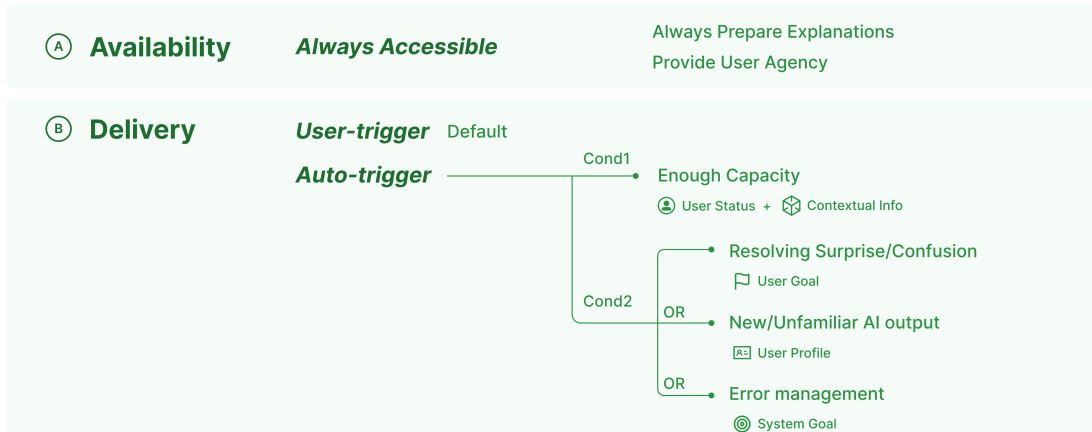
1) Cases when users have an expectation mismatch and become surprised/confused about AI outcomes [37, 67], *i.e.*, *User Goal* as Resolving Surprise/Confusion (also reflected by *User State*, which could be detected by AR HMDs using facial expressions and gaze patterns [21, 212]). An example could be an intelligent reminder to bring umbrella when users are leaving home on a sunny morning (but it will rain in the afternoon). Automatic explanations of the weather forecast could help resolve users' confusion.

2) Cases when users are unfamiliar with new AI outcomes (indicated via history information of *User Profile*), *e.g.*, users receive a recommendation of a song that they have never heard before. Just-in-time explanations of the reason can help users to better understand the recommendation.

3) Cases when the model's input or output confidence is low and the model may make mistakes [30, 108], *i.e.*, *System Goal* as Error Management. For instance, a system turning on a do-not-disturb mode when it detects a user working on a laptop in an office when the AR-based activity recognition confidence was low (*e.g.*, 80%). Explanations could be a gatekeeper if the detection was wrong and users could calibrate their expectations or adjust the system to improve the detection [64, 243].

All of these cases have the prerequisite that users have enough capacity to consume explanations [166, 191], *e.g.*, users' cognitive load is not high (could be detected via gaze or EEG on wearable AR devices [224, 238]), and users have enough time to do so (inferred based on context).

Figure 4: The "When" Part of XAIR. It contains two major dimensions: (A) Availability and (B) Delivery, highlighted in bolded texts. The design choice of dimensions are in *italic* texts (same below for Fig. 5 and Fig. 6). For example, Delivery can be either *User-trigger* or *Auto-trigger*. Each dimension has factors that should be considered for explanation designs. The guidelines G1 and G2 provide advice on these design choices.

G2. By default, don't trigger explanations automatically, wait until users' request. Only trigger explanations automatically when both conditions are met:
(1) Users have enough capacity (e.g., cognitive load, urgency);
(2) Users are surprised/confused, or unfamiliar with the outcome, or the model is uncertain.

5.2 What to Explain?

In Sec. 3.1.2, we identified that *content* and *detail* were two dimensions of the *what* part of the framework. We introduce how to choose among all explanation content types in Fig. 5.

5.2.1 Ⓐ *Content.* In AR systems, the AI outcomes are based on factors such as *User State* (*e.g.*, user activity), *Contextual Information* (*e.g.*, the current environment), and *User Profile* (*e.g.*, user preference). These factors also determine the content of different explanation content types. To choose the right types, the framework lists three factors to consider and provides recommendations of personalized explanation content types based on the literature (shown as solid check marks in the top table in Fig. 5), end-user survey, or expert advice (based on **Finding 3** and **Suggestion 1**, shown as hollow check marks).

1) *System Goal.* Different system goals need different explanations. For example, when a system recommends that users check out a new clothing store (User Intent Discovery), presenting Examples of similar stores that users are interested in and Why this store is attractive to users can be helpful. When a system wants to calibrate users' expectations about uncertain recipe recommendations (Error Management), showing Examples is less meaningful than presenting How and Why the system recommended this recipe,

and How To change output if users want to. We leverage some literature on contextualized explanation content types to support our recommendations in the framework [64, 132, 133].

2) *User Goal.* Similarly, different user goals also require different explanations. For instance, Certainty explanations are helpful when users want to make sure an exercise recommendation fits their health plan (Reliability), while such explanations would be not useful when users want to be more aware of which data an AR system uses (Privacy Awareness). Most of these recommendations are supported by previous studies [25, 119, 132, 155, 174, 217]. Regarding how to identify the user goals to choose explanations, designers can use their expertise to infer them in the context determined by AR systems. In the future, it is also possible for AR systems to combine a range of sensor signals to detect/predict users' goals [14, 111, 211].

3) *User Profile*, specifically user literacy with AI. For the majority of end-users who are unfamiliar with the AI techniques, we recommend only considering the four content types that users indicated that they would find useful: Input/Output, Why/Why-Not, How, and Certainty (as shown in **Findings 3** and Fig. 3c). If users have high AI literacy, then all types could be considered [74, 114].

Elements in *System/User Goal* are not exclusive to each other. If there is more than one goal, these columns can be merged within each factor section to find the union (*i.e.*, content types checked in at least one column). Then, one can find the intersection among the three factors' content type sets (*i.e.*, overlapping types in all sets) to ensure that these explanations can fulfill all factors simultaneously. We show complete examples in later sections (Sec. 6 and Sec. 7).

G3. To determine personalized explanation content, consider three factors: system goal, user goal, and user profile.

A — Content

Explanation Types	System Goal				User Goal				User Profile	
	User Intent Discovery	User Intent Assistance	Error Management	Trust Building	Resolving Surprise /Confusion	Privacy Awareness	Reliability	Informative-ness	Low AI Literacy	High AI Literacy
Input/Output	○	✓	✓	○	○	✓	✓	✓	✓	✓
Why/Why-Not	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
How			✓	✓	✓	✓	✓	✓	✓	✓
Certainty			✓	○	○		✓	✓	✓	✓
Example	○	○		○			✓	✓		✓
What-If		○		○			✓	✓		✓
How-To			✓	○	○			✓		✓

Content Source:
- 👤 User Status
- 🔷 Contextual Info
- 🆔 User Profile

Select based on Scenarios & Union — *Select & Union* — *Select*

↓ Intersection

Multiple Types

B — Detail

- 🎯 System Goal
- 🚩 User Goal
- 🆔 User Profile

Highlight Top Types — Prioritize "Why"

Detailed Explanation ← "More" Option upon User Request

Concise Explanation (Default)

G3 Ⓐ: To determine personalized explanation contents, consider three factors: system goal, user goal, and user profile.
G4 Ⓑ: By default, display concise explanations with top types. Prioritize "Why", and choose other types based on the context.
G5 Ⓑ: Always provide users opportunities for agency with the option to explore more detailed explanations upon request.

Figure 5: The "What" Part of XAIR. (A) Content and (B) Detail are the two major dimensions. Combining the literature summary (shown as solid check marks) and findings from Study 1 & 2 (shown as hollow check marks), G3-G5 provide guidelines on choosing appropriate explanation content types and length.

5.2.2 Ⓑ *Detail.* After selecting the appropriate content, default explanations need to be concise and can be further simplified by highlighting the most important types [26, 40]. General end-users are primarily interested in Why, which is in line with experts' advice (**Suggestion 3** about default options) and previous literature [102, 132, 133]. Designers' can leverage their expertise to determine whether other types should be omitted or combined with Why in a specific context.

G4. By default, display concise explanations with top types. Prioritize Why, and choose other types based on the context.

As a large proportion of Study 1 participants indicated that detailed explanations could be useful (**Finding 4**), AR systems need to provide an easy portal (an interface widget such as a button) for end-users to explore more details. This can also provide user agency [145].

G5. Always provide users opportunities for agency with the option to explore more detailed explanations upon request.

5.3 How to Explain?

Finally, we introduce the *how* part and elaborate from the *modality* and *paradigm* perspective (see Fig. 6).

5.3.1 Ⓐ *Modality.* Considering channel bandwidth, the visual and audio modalities are the two most feasible modalities for AR. Since explanations usually come during or after AI outcomes, to maintain consistency, the default modality of an explanation should be the same(**Finding 5** and **Suggestion 3**). In cases when outcomes use a haptic modality (*e.g.*, vibration as a reminder), audio channels could be used as necessary (although this should be rare), since the choice of the haptic channel already conveys the need to be subtle.

However, there are also cases where one modality could be overloaded (based on *User State* and *Contextual Information*). For example, when users are driving and a navigation app suggests an alternate detour route, although the AI outcome is visual, the explanation should be audio to avoid visual overload. When users are in a loud environment, a vibration-based AI outcome needs to use the visual modality for explanations. These scenarios can be easily detected by AR HMDs.

G6. By default, adopt the same explanation modality as that of the AI output (except for haptic→audio). When one modality's load is high, use another modality.

Note that the modality choice also applies to the manual-trigger case when explanations are not automatically delivered (*G2*), *e.g.*, a button icon for visual modality, a voice trigger for audio modality.
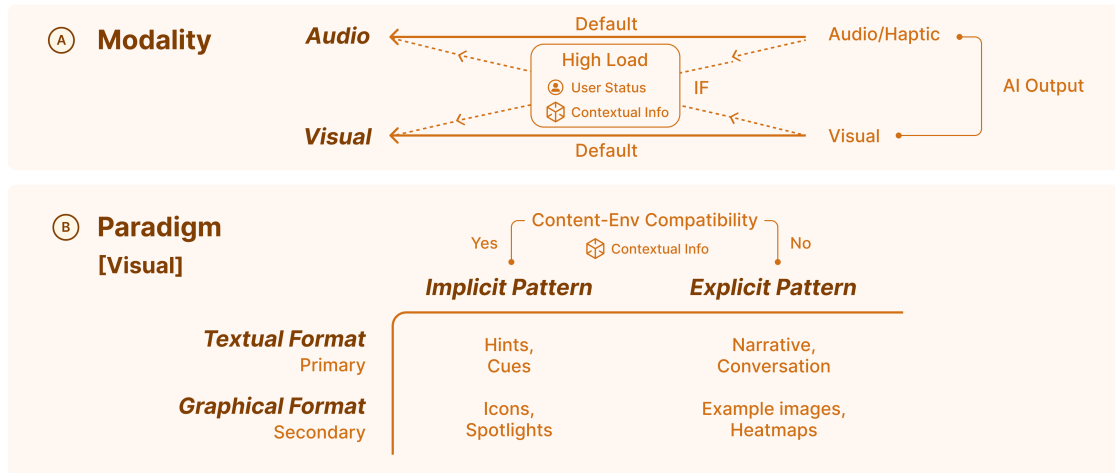
G6 Ⓐ : By default, adopt the same modality as that of AI output (except haptic → audio).
When one modality's load is high, use the other modality.

G7 Ⓑ : [Visual] Use text as the primary format. Only use graphics if they are easy to understand.

G8 Ⓑ : [Visual] Use implicit patterns if content can be part of the env. Otherwise, use explicit patterns.

Figure 6: The "How" Part of XAIR. (A) Modality and (B) Paradigm are the major dimensions. Note that Paradigm is only for the visual modality, and it is further broken down into two perspectives: format and pattern. G6-G8 provides guidelines on making the proper design choices.

5.3.2 Ⓑ *Paradigm.* Experts agreed that the audio design space does not belong within this framework. For visual design, after removing the location from our framework (**Suggestion 2** of redundancy removal), we mainly focused on two aspects: format and pattern. Depending on the content (*G3*), the explanation format can be textual [116, 158], graphical [200, 240], or both. Based on the consensus of experts in Study 2, text should be the primary format. Experts suggested several reasons for this. Text takes up less space in a limited AR interface, and can introduce relatively less cognitive load. Moreover, the textual format is more universal and can cover all types. Graphics can be used as the secondary format. For default explanations (*G4*), in addition to displaying a short and concise textual paragraph, simple graphics such as icons can be used to provide additional information. For detailed explanations (*G5*), more complex graphical formats (*e.g.*, example images or heatmaps) can be used as long as they are easy for end-users to understand.

G7. [Visual] Use text as the primary format. Only use graphics if they are easy to understand.

Independent of the format, explanations can be presented in an implicit or explicit pattern [136, 208]. Given the capability of depth sensing and 3D registration in AR, we recommend using the implicit pattern when the explanation content is compatible with the environment (*i.e.*, can be naturally embedded as a part of the environment). For example, for book recommendations, a text cue or a small icon can float on the book to indicate the book's topic that users like (belonging to the Why explanation content type). When explanations and the environment are not compatible, using an explicit pattern (*e.g.*, a dialogue window) can be the backup option. With regard to what explanation content is compatible with the environment, designers can leverage their expertise and

intuition to propose appropriate embedding patterns for given a context. Future AR systems may first understand the environment using object detection and context recognition algorithms, and then utilize techniques such as knowledge graphs (*i.e.*, networks of real-world entities and their relationships) [52] to assess the compatibility between the content and the environment.

G8. [Visual] Use implicit patterns if content can be embedded in the environment. Otherwise, use explicit patterns.

XAIR can not only serve as a summary of the study findings and the multidisciplinary literature across XAI and HCI, but also guide effective XAI design in AR. In the next two sections, we provide examples of XAIR-supported applications (Sec. 6), and evaluate XAIR from both designers' and end-users' perspectives (Sec. 7).

6 APPLICATIONS

To demonstrate how to leverage XAIR for XAI design, we present two examples that showcase potential workflows that use XAIR for everyday AR applications (Fig. 7). More details can be found in Appendix B.1. After determining the key factors for a given scenario, we used the framework (Fig. 4-Fig. 6) to make design choices based on the factors.

6.1 Scenario 1: Route Suggestion while Jogging

Scene. Nancy (AI expert, high AI literacy) is jogging in the morning on a quiet trail. Since it is the cherry-blossom season and Nancy loves cherries, her AR glasses display a map beside her and recommend a detour. Nancy is surprised since this route is different from

(a) Scenario 1: Route Suggestion when Jogging

(b) Scenario 2: Plant Fertilization Reminder.

Figure 7: Application of XAIR on Two Everyday AR Scenarios. In the second scenario, the hand icon indicates that explanations are manually triggered (the same below). Figures only present the default, concise explanations. Detailed explanations are described in the main text of Sec. 6.

her regular one, but she is happy to explore it. She is also curious to know the reason this new route was recommended.

When. Delivery. Nancy has enough cognitive capacity in this scenario. Her *User Goal* is Resolving Surprise. Therefore, an explanation is automatically triggered because the two conditions are met (*G2*).

What. Content. Other than the *User Goal*, the *System Goal* is User Intent Discovery (exploring a new route to see cherry blossom). Considering Nancy's *User Profile*, she is an expert in AI, so the appropriate explanation content types (*G3*) are Input/Output (*e.g.*, "This route is recommended based on seasons, your routine, and preferences.") and Why/Why-Not (*e.g.*, "The route has cherry blossom trees that you can enjoy. The length of the route is appropriate and fits your morning schedule."). Examples for all seven explanation content types can be found in Appendix B.1.
Detail. The AR interface shows the Why as default (*G4*), and can be expanded to show both types in detail (*G5*). Nancy can slow down and click the "More" button to see more detailed explanations while standing or walking.

How. Modality. The explanation is presented visually, the same as the recommendation (*G6*).
Format. The default explanation uses text, while the detailed explanation contains cherry-blossom pictures of the new route to help explain the Why (*G7*).
Pattern. The explanation is shown explicitly within the route recommendation window (*G8*).

6.2 Scenario 2: Plant Fertilization Reminder

Scene. Sarah (general end-user, low AI literacy) was chatting with her neighbor about gardening. After she returned home and sat on the sofa, her AR glasses recommended instructions about plant fertilization by showing a care icon on the plant. Sarah is concerned about technology invading her privacy, and wants to know the reason behind the recommendation.

When. *Delivery*. Although Sarah has enough cognitive capacity, none of the three cases in the second condition of *G2* are met (*i.e.*, she was familiar with the recommendation and not confused, and

the model didn't make a mistake). Therefore, the explanation needs to be manually triggered (*G2*).

What. Content. In this case, the *System Goal* is Trust Building (clarifying the usage of data), and the *User Goal* is Privacy Awareness. Sarah's *User Profile* indicates that she is not an expert in AI. According to *G3*, the explanation content type list contains Input/Output, Why/Why-Not, and How.
Detail. Considering Sarah's concern, the default explanation merges Why and How: "The system scans the plant's visual appearance. It has abnormal spots on the leaves, which indicate fungi or bacteria infection." (*G4*). For the detailed explanation, the full content of the three types is presented in a drop-down list upon her request (*G5*).

How. Modality. Following *G6*, the visual modality is used for both the explanation and the manual trigger (a button beside the plant care icon).
Format. Other than using text as the primary format, the abnormal spots on the leaves are also highlighted via circles to provide an in-situ explanation (*G7*).
Pattern. Since the highlighting of spots is compatible with the environment (shown on leaves), it adopts the implicit pattern (*G8*). The rest of the texts of the explanation uses the explicit pattern.

Our two examples demonstrate XAIR's ability to guide XAI design in AR in various scenarios. In Appendix B.2, we provide additional everyday AR scenarios to further illustrate its practicality.

7 EVALUATION

In addition to showing examples to illustrate the use case of XAIR, we also conducted two user studies to evaluate XAIR. The first study was from the perspective of designers (as XAIR users) to evaluate XAIR's ability to assist designers during their design processes (Sec. 7.1). The second study was from an end-user perspective and evaluated XAIR's effectiveness at achieving a user-friendly XAI experience in AR. We measured the usability of the real-time AR experiences that were developed based on the design examples proposed by designers (Sec. 7.2).

7.1 Study 3: Design Workshops

We conducted one-on-one design workshops with designers to investigate whether the framework could support their design processes, inspire them to identify new design opportunities, and achieve effective designs.

7.1.1 Participants. Future XAI and AR designers can come from various backgrounds, so we recruited 10 participants (4 Female, 6 Male, Age 32 ± 6) from a technology company as volunteers. Three were XAI algorithm researchers, four were product designers, and three were HCI/AR researchers. All participants were familiar with AI and AR, and none had participated in previous studies.

7.1.2 Design and Procedure. We prepared two AR scenarios, both related to recipe recommendations while preparing meals.

Case 1: Reliable Recipe Recommendation. Michael works in a sales company (general end-user, low digital literacy). He recently started a high-protein diet due to his workout routine. He opens the fridge and wants to make lunch. His AR glasses present a window on the fridge door and recommend an option that Michael usually has, but Michael wants to make sure that this option fits his recent diet changes.

Case 2: Wrong Recipe Recommendation. Mary works in an AI company (high AI literacy) and has friends coming over for dinner, who are beef lovers. She opens the fridge and sees steak. However, her AR glasses mistakenly recognize steak as salmon with a medium level of confidence [1], and recommends a few recipes that use salmon. She is confused and wonders how she can correct the recommendations.

Since generating explanations is not the focus of the framework, we prepared examples for the seven explanation content types (Appendix B.3). Participants were free to use our examples, or propose their own (without the need to design how an algorithm could generate them).

Participants first used their expertise and intuition to propose XAI designs for the two cases before being shown the framework. They spent 10 minutes on each case. Participants were encouraged to think aloud and describe their design via text and simple sketches. Then, after XAIR was introduced, they spent another 10 minutes following the three parts and eight guidelines and applied them to the two cases, resulting in another version of the design. The order of the two cases was counterbalanced.

To quantify the utility of XAIR, we employed the Creativity Support Index (CSI, 1-10 Likert scale) [55] and System Usability Scale (SUS) [24]. Since both scales were originally designed for tools or systems, the language was modified from "tools" and "system" to "framework" and "guidelines". At the end of the workshop, we conducted a semi-structured interview that began with the question: "Do you think the framework and guidelines are helpful? If so, in what aspects they are helpful?" Each workshop lasted 90 minutes. Two researchers independently coded the qualitative data using thematic analysis and discussed it to reach an agreement.

---

[1] If the system has low-level confidence, the expected cost of making mistakes will be higher than the cost of asking for users' input, so the system should ask for users' confirmation about the ingredients they have on hand before presenting recommendations (*e.g.*, asking "Is this salmon or steak?"). In this scenario, the confidence is at the medium level, thus the system provides recommendations, but is still aware of the potential to make mistakes.

7.1.3 Design Results. After using XAIR, nine out of ten participants modified their designs and preferred the updated version. One participant (P7) liked the design as it was and thought that the framework *"perfectly supported the design"*. Consistency was found among the designs, which indicated that XAIR could effectively guide users through the design process. For example, Tab. 1 presents two designers' designs (images are rendered based on their proposals) of the reliable recommendation case. Their designs of the *when* part and most of the *what* part were the same. Tab. 2 presents another two designers' designs of the wrong recommendation case (Case 2). Similarly, we also found consistent design choices between the two examples.

Meanwhile, we also found variance across participants' designs. For instance, in Case 1, P6 had a different consideration of *User State* than P2, in which P6 brought up a case where the user could hold something in their hand. In this case, P6 adopted the audio modality for manual trigger (the rightmost column of Tab. 1). Moreover, as shown in the rightmost column of Tab. 2, P9 proposed an interesting tweak that always highlighted ingredients (Input explanation type). Her reason was that it introduced *"ultra-low cognitive cost"*, thus there was no need to check the second auto-trigger condition. *"I don't think it is a violation of the guideline. Instead, I was inspired by the framework to consider this case."* This reveals that XAIR is flexible and can support the diverse creativity of users.

7.1.4 Feedback Results. Participants provided positive feedback about the framework. Eight participants explicitly commented that XAIR was *"useful/helpful"*. The results of the CSI scores (Fig. 8) and the SUS scores (74 ± 6 out of 100, indicating good usability) both illustrate the good utility of XAIR. Four themes emerged in participants' feedback.

The Framework as a Useful and Comprehensive Reference. Consistent with the feedback from the experts in Study 2 (Sec. 4.2), participants also found that the framework was a valuable handbook. For example, "*This framework is an excellent reference point for people getting started designing XAI experiences... to check if they have missed things*" (P4) and "*I may not use it for every design decision, but I would refer to it when I want to make sure that I have considered everything.*" (P7) The comprehensiveness of XAIR thus helped participants perform a sanity check of their designs.
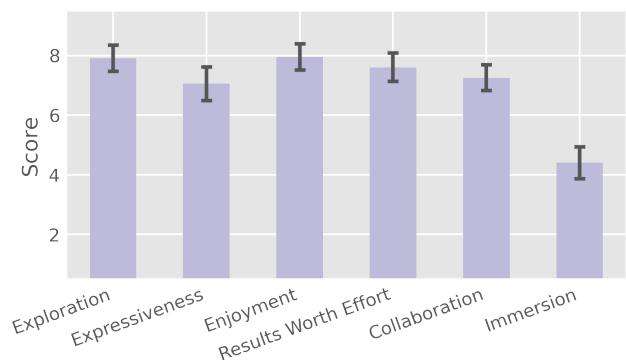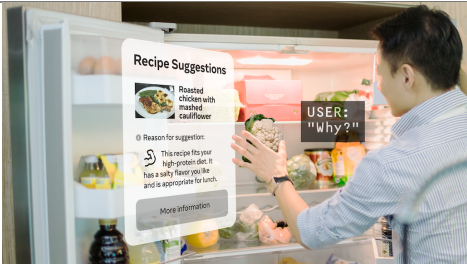

Figure 8: CSI Scores of Design Workshops in Study 3

Designer		P2, Product Designer	P6, XAI Researcher
Platform-Agnostic Key Factors	**System Goal**	**User Intent Assistance** (to find a good recipe)	
	User Goal	**Reliability** (to make sure the recipe fits the diet)	
	User Profile	**User Preference**: High protein food; **History**: Know these recommended recipes; **AI Literacy**: General end-user, low	
AR-Specific Key Factors	**Contextual Info**	**Location**: Kitchen; **Time**: Noon; **Environment**: Various ingredients in the fridge	
	User State	**Activity**: Opening the fridge to make lunch; **Cognitive Load**: Low	**Activity**: Opening the fridge to make lunch, possibly holding something **Cognitive Load**: Low
XAI Designs in AR: *When*	Availability (*G1*)	Always available	Same as P2
	Delivery (*G2*)	Manual-trigger, because the second condition of auto-trigger was not met given the *System Goal, User Goal,* and *User Profile.*	Same as P2
XAI Designs in AR: *What*	Content (*G3*)	Input/Output & Why/Why-Not based on Fig. 5's table	Same as P2
	Detail - Concise (*G4*)	An explanation merging the Why and Input content types, as explaining *"showing ingredients is also important"*	An explanation of the Why part as *"it needs to be prioritized"*
	Detail - Detailed (*G5*)	A list of the two explanation types in detail	Same as P2; Cherry flower pictures to support the Why explanation
XAI Designs in AR: *How*	Modality (*G6*)	Visual modality to ensure consistency with the recommendation interface	Visual modality for explanations; Audio/visual modality for manual trigger if the user is/isn't holding something
	Paradigm - Format (*G7*)	Textual format	Textual format as the primary format; Graphic format (protein icon) to support explanations
	Paradigm - Pattern (*G8*)	Explicit pattern, presenting texts in the same window as the recommendations	Same as P2

Table 1: Two Design Examples of Case 1: Reliable Recipe Recommendation. Participants' quotes were presented in italic font. Among key factors, P2 and P6 had different thoughts on *User State,* which leads to different design choices of how - modality. The comparison indicates both consistency and variance between two designers' examples.

Design Opportunity Inspiration. Participants also leveraged XAIR to inspire new ideas. P6's original design did not consider the case where users' hands could be busily holding ingredients. But the modality in the how part inspired him, *i.e., "The framework reminded me to realize potential alternatives. It inspired me to think about not just one design, but a set of designs."* Moreover, participants found that XAIR could help generate baseline designs. *"I could then further customize it for various scenarios."* (P8) The high scores for exploration (7.9±0.4 out of 10) and expressiveness (7.1±0.6) on the CSI also support this observation.

Backing Up Design Intuitions. Some participants also found that the guidelines in XAIR could support their intuition. For instance, P7 did not change her design after using XAIR, but was very excited to see the alignment, *e.g., "Sometimes I am not sure whether my design*

intuition is right. It feels great that the framework can support it." This could be part of the reason for the positive enjoyment score on the CSI (8.0±0.4).

Time to Learn The Framework. Participants also commented that XAIR incorporates a lot of information and that they needed time to digest it, *e.g.,"I need to go back and forth between the visual diagrams"* (P10) and *"the table [in Fig. 5] is useful but also pretty complex"* (P4). This may explain the relatively low immersion score (4.4±0.5) on the CSI. Moreover, six participants Agreed or Strongly Agreed in response to the question *"Need to learn a lot..."* on the SUS. On the one hand, this shows XAIR's comprehensiveness (covering multiple research domains), whereas on the other hand, this illuminates future directions to convert XAIR into a design tool.

	Designer	**P5, HCI/AR Researcher**	**P9, Product Designer**
Platform-Agnostic Key Factors	**System Goal**	**User Intent Assistance** (to find a good recipe for friends) **Error Management** (to calibrate the user's trust for mid-level recognition confidence)	
	User Goal	**Resolve Confusion** (to understand why the recommendations are wrong)	
	User Profile	**User Preference**: Meet-lovers (friends); **AI Literacy**: Expert, high	
AR-Specific Key Factors	**Contextual Info**	**Location**: Kitchen; **Time**: Evening; **Environment**: Various ingredients in the fridge	
	User State	**Activity**: Opening the fridge to make dinner; **Cognitive Load**: Low	
XAI Designs in AR: *When*	Availability (*G1*)	Always available	Same as P5
	Delivery (*G2*)	Auto-trigger, because both conditions were met given the *System Goal* and *User Goal*	Auto-trigger; Besides, a new tweak to always spotlight ingredients automatically, since it introduced *"ultra-low cognitive cost"*
XAI Designs in AR: *What*	Content (*G3*)	Five Types: Input/Output, Why/Why-Not, How-To, Certainty, and How	Same as P5
	Detail - Concise (*G4*)	An explanation merging Why, Input, Certainty (color-coding to show ingredient with a mid-level confidence), and How-To (selecting ingredients to change)	An explanation Why and How-To; Besides, Input explanations were shown by spotlighting ingredients, which can be selected and changed (How-To)
	Detail - Detailed (*G5*)	A drop down menu of the five types	Same as P5
XAI Designs in AR: *How*	Modality (*G6*)	Visual modality	Same as P5
	Paradigm - Format (*G7*)	Textual format	Textual format as the primary format; Graphic format (spotlighting boundaries) to denote ingredients
	Paradigm - Pattern (*G8*)	Explicit pattern, presenting texts in the same window as the recommendations	Explicit pattern for texts (same as P5); Implicit pattern for graphic spotlights

Table 2: Two Design Examples of Case 2: Wrong Recipe Recommendation.

7.2 Study 4: Intelligent AR System Evaluation

To demonstrate XAIR's effectiveness, we show that the designers' proposals using XAIR could achieve a positive XAI user experience in AR for end-users. Based on the designs proposed in Study 3, we took one example from each case and implemented a real-time intelligent AR system. We then evaluated the system's usability.

7.2.1 System Implementation. We selected one reliable recipe recommendation example from the left of Tab. 1 and one wrong recipe recommendation example from the left of Tab. 2. We then instantiated the examples by implementing a real-time system on a Microsoft Hololens V2. The system had three major modules: a recognition module, a recommendation module, and an interface module.

For ingredient recognition, we trained a vision-based object detection model that was a variant of the Vision Transformer from CLIP [173] on the LVIS [95] and Objects365 [197] datasets. We then added ImageNet22k and performed weakly-supervised training with both box and image level annotations [244]. The top 50 ingredient-related classes from LVIS were retained, with an average

F1 score of 81.1%. The model was run on Hololens' egocentric camera stream at 5 FPS to recognize ingredients. The model was used in Case 1, while in Case 2, misrecognition (*i.e.*, recognizing steak as salmon) was manually inserted to create the designed experience.

For recipe recommendation, the Spoonacular Food API [4] was used to obtain potential recipes given a set of ingredients. We then implemented an algorithm to rank the recipes based on user preference and recommend the top recipes (*e.g.*, if a user prefers food that is fast to prepare, the recipes are sorted based on the cooking time). For the explanations, we developed a template-based explanation generation technique [242] to cover different types.

Finally, the interface followed the designs in Tab. 1 and Tab. 2. Clicking on one recipe's image would show the detailed instructions. An icon button under each recipe could be triggered to present short default explanations, followed by another button to display detailed explanations as a list of content types.

7.2.2 Participants and Apparatus. Twelve participants (5 Female, 7 Male, Age 32 ± 3) volunteered to join the study. None of the them had participated in previous studies. The two cases had the same

(a) Evaluation Setup

(b) Evaluation Scores

Figure 9: End-user Evaluation of The AR System in Study 4. (a) Study Setup. (b) Evaluation Scores. Users had positive experience in both tasks. Note that tasks were evaluated separately and not meant to compare against each other.

setup (except for the recognition error). We prepared a number of food ingredients on a shelf (including steak, but no salmon) to simulate the opening-a-fridge moment, as shown in Fig. 9a.

7.2.3 Design and Procedure. Since there is no existing XAI design for AR systems, we compared the design examples with a baseline condition that only presented recommendations without explanations. Note that for Case 2's baseline condition, participants could still change the output by clicking a button that said "Doesn't seem right? Click to see the next batch." to ensure a fair comparison [2].

We used a within-subject design. Participants started with one case and completed both conditions. They took a break and completed a questionnaire to compare the two conditions. Then, they completed the two conditions in the second case and completed a similar questionnaire. The case order was counterbalanced. The study took about 30 minutes and ended with a brief interview.

The questionnaire contained six questions (1-7 Likert scale) comparing the two conditions. Three were from the XAI literature and measured the explanations' effect on the system's intelligibility, transparency, and trustworthiness. The other three questions asked about participants' preferences towards the design choices of *when*,

---

[2]Another baseline could have been to compare against designers' old designs before using XAIR. However, we did not include this baseline since designers already explicitly preferred the new version that they created after using XAIR.

what, and *how*[3]. The SUS was also administered to measure the usability of the system with explanations.

7.2.4 Results. Participants strongly preferred the condition with explanations in both cases, especially Case 2, *e.g.,*"*Seeing the explanation automatically when the AR system makes mistakes is very helpful. It lets me know when I should adjust my expectation*" (P2) and "*the mistake [in Case 2] is understandable... salmon and steak can have similar colors and shapes. But if I didn't see the explanation, I would be very confused.*" (P9) This sentiment was also reflected in participants' high rating of the system's intelligibility, transparency, and trustworthiness with the explanation (Fig. 9b). Moreover, the AR system received high SUS scores: 86 ± 3 in Case 1, and 80 ± 3 in Case 2, both indicating excellent usability of the system. Participants also liked the design of the system, which was supported by the positive ratings for the when/what/how questions (see Fig. 9b). These results demonstrated that compared to the baseline, XAI design using XAIR can effectively improve the transparency and trustworthiness of AR systems for end-users.

8 DISCUSSION

XAIR defines the problem space structure of XAI design in AR and details the relationship that exists between the factors and the problem space. By highlighting the key factors that designers need to consider and providing a set of design guidelines for XAI in AR, XAIR not only serves as a reference for researchers, but also assists designers by helping them propose more effective XAI designs in AR scenarios. The two evaluation studies in Sec. 7 illustrated that XAIR can inspire designers with more design opportunities and lead to transparent and trustworthy AR systems. In this section, we discuss how researchers and designers can apply XAIR, as well as potential future directions of the framework inspired by our studies. We also summarize the limitations of this work.

8.1 Applying XAIR to XAI Design for AR

Researchers and designers can make use of XAIR in their XAI design for AR scenarios by initially using their intuition to propose an initial set of designs. Then, they can follow the framework to identify five key factors: *User State, Contextual Information, System Goal, User Goal,* and *User Profile.* The example scenarios in Sec. 6 and Sec. 7 indicate how these factors can be specified. Based on these factors, they would then work through the eight guidelines of *when, what,* and *how,* using Fig. 4-Fig. 6 to inspect their initial design and make modifications if there is anything inappropriate or missing. Low-fidelity storyboards or prototypes of the designs can be tested via small-scale end-user evaluation studies. This would be an iterative process. In the future, when sensing and AI technologies are more advanced, it is promising that the procedures of identifying factors and checking guidelines could be automated.

---

[3]Since a factorial study design to compare all XAIR design options would involve a large number of conditions (*i.e.,* 2 options of *when* × at least 2 options of *what* × 2 options of *how*), asking participants to undergo several scenarios would be too costly. Order effects would also be hard to counterbalance. So the three questions about when/what/how described other design choices by showing examples and asked about participants' preferences. For instance, in Case 1's *when* part, participants rated how much they agreed with the claim "I prefer to have explanations triggered manually by me, compared to being triggered automatically.", or vice-versa in Case 2.

8.2 Towards An Automatic Design Recommendation Toolkit

In Study 3, more than one user mentioned the possibility of converting the framework into an automatic toolkit. For example, P3 was thinking aloud when using XAIR in the study, *"If this framework is described as an algorithm, the five key factors can be viewed as the input of the algorithm... and the output is the design of the three questions."* There are a few decision-making steps in the current framework that involve human intelligence. For example, when designing the default explanations in what - detail, designers need to consider users' priority under a given context to determine which explanation content type to highlight. When picking the appropriate visual paradigm, designers need to determine whether the explanation content is more appropriate in a textual or graphical format, as well as whether the content can be naturally embedded within the environment. Assuming future intelligent models can assist with these decisions, XAIR could be transformed into a design recommendation tool that could enable designers and researchers to experiment with a set of *User State, Contexts, System/User Goals*, and so on. This could achieve a more advanced version of XAIR, where XAIR are fully automated as an end-to-end model: determining the optimal XAI experience by inferring the five key factors in real time. This is an appealing direction. However, although factors such as *Context* and *System Goal* are easier to predict with a system, the inference of *User State/Goal* is still at an early research stage [21, 71, 99]. Moreover, extensive research is needed to validate the adequacy and comprehensiveness of the end-to-end algorithm. This also introduces the challenge of nested explanations in XAIR (*i.e.*, explaining explanations) [154], which calls for further study.

8.3 The Customized Configuration of XAI Experiences in AR

The experts in Study 2 and the designers in Study 3 brought up the need for end-user to control XAI experiences in AR, *e.g.*,*"XAIR can provide a set of default design solutions, and users could further customize the system"* (P12, Study 2) and *"I personally agree with the guidelines, but I can also imagine some users may want different design options. So there should be some way that allows them to select when/what/how... For example, a user may want the interface to be in an explicit dialogue window all the time [related to how]. We should support this."* (P8, Study 3) This need for control suggests that to achieve a personalized AR system, designers should provide users with methods to configure their system, so that they can set up specific design choices to customize their XAI experience. Such personalization capabilities may also be used to support people with accessibility needs (also mentioned by P2 in Study 3), *e.g.*, visually impaired users can choose to always use the audio modality.

8.4 User-in-The-Loop and Co-Learning

During the iterative expert workshops (Study 2, Sec. 4.2), experts mentioned an interesting long-term co-learning process between the AR system and a user. On the one hand, based on a user's reactions to AI outcomes and explanations, a system can learn from the data and adapt to the user. Ideally, as the AR system better understands the user, the AI models would be more accurate, thus reducing the need for mistake-related explanations (*e.g.*, cases

where *System Goal* as Error Management). On the other hand, the user is also learning from the system. "*Users' understanding of the system and AI literacy may change as they learn from explanations*" (P4, Study 2). This may also affect the user's need for explanations. For example, the user may have less confusion (*User Goal* as Resolving Surprise/Confusion) as they become more familiar with the system. Meanwhile, they may become more interested in exploring additional explanation types (*User Goal* as Informativeness). Such a long-term and co-learning process is an interesting research question worth more exploration.

8.5 Limitations

There are a few limitations to this research. First, although we highlighted promising technical paths within the framework in Sec. 5, XAIR does not involve specific AR techniques. The real-time AR system in Study 4 implemented the ingredient recognition and recipe recommendation modules, but the detection of user state/goal was omitted. Second, our studies might have some intrinsic biases. For example, Study 1 only involved AR recommendation cases. Since everyday AR HMDs are still not widely adopted in daily life, we grouped 500+ participants only based on AI experience instead of AR experience. The experts and designers of our studies were all employees of a technology company. Study 4 only evaluated two specific proposals from designers. Moreover, as there is no previous XAI design in AR, we were only able to compare our XAIR-based system against a baseline without explanation. Third, other than when, what, and how, there could be more aspects in the problem space, *e.g.*, who and where to explain. Moreover, XAIR mainly focuses on non-expert end-users. Other potential users, such as developers or domain experts, were not included. The scope of the five key factors may also not be comprehensive. For example, we do not consider user trust in AI, which is a part of *User Profile* that may be dynamic along with user-system interaction. These could limited the generalizability of our framework, but also suggests a few potential future work directions to expand and enhance XAIR.

9 CONCLUSION

In this paper, we propose XAIR, a framework to guide XAI design in AR. Based on a literature review of multiple domains, we identified the problem space using three main questions, *i.e.*, when to explain, what to explain, and how to explain. We combined the results from a large-scale survey with over 500 end-users (Study 1) and iterative workshops with 12 experts (Study 2) to develop XAIR and a set of eight design guidelines. Using our framework, we walked through example XAI designs in two everyday AR scenarios. To evaluate XAIR's utility, we conducted a study with 10 designers (Study 3). The study revealed that designers found XAIR to be a helpful, comprehensive reference that could inspire new design thoughts and provide a backup of designer intuitions. Moreover, to demonstrate the effectiveness of XAIR, we instantiated two design examples in a real-time AR system and conducted another user study with 12 end-users (Study 4). The results indicated excellent usability of the AR system. XAIR can thus help future designers and researchers achieve effective XAI designs in AR and help them explore new design opportunities.

REFERENCES

[1] 2019. General Data Protection Regulation (GDPR). https://gdpr-info.eu/
[2] 2022. AI system cards. https://ai.facebook.com/tools/system-cards/
[3] 2022. Enterprise augmented reality (AR) platform designed for business: Magic leap. https://www.magicleap.com/en-us/
[4] 2022. Free meal planner, food tracker, and Recipe Saver. https://spoonacular.com/
[5] 2022. Glass Glass. https://www.google.com/glass/start/
[6] 2022. Google Cloud Model Cards. https://modelcards.withgoogle.com/
[7] 2022. H2O driverless AI. https://h2o.ai/platform/ai-cloud/make/h2o-driverless-ai/
[8] 2022. Microsoft hololens: Mixed Reality Technology for Business. https://www.microsoft.com/en-us/hololens
[9] 2022. Model interpretability with DALEX. http://uc-r.github.io/dalex
[10] 2022. Spectacles by snap inc. - the next generation of spectacles. https://www.spectacles.com/
[11] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–18. https://doi.org/10.1145/3173574.3174156
[12] Michael Abrash. 2021. Creating the future: augmented reality, the next human-machine interface. In *2021 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 1–2.
[13] Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052 Conference Name: IEEE Access.
[14] Henny Admoni and Siddhartha Srinivasa. 2016. Predicting user intent through eye gaze for shared autonomy. In *2016 AAAI Fall Symposium Series*.
[15] Gediminas Adomavicius and Alexander Tuzhilin. 2011. Context-aware recommender systems. In *Recommender systems handbook*. Springer, 217–253.
[16] Imran Ahmed, Gwanggil Jeon, and Francesco Piccialli. 2022. From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where. *IEEE Transactions on Industrial Informatics* 18, 8 (2022), 5031–5042.
[17] Saleema Amershi, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, Eric Horvitz, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, and Paul N. Bennett. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, New York, USA, 1–13. https://doi.org/10.1145/3290605.3300233
[18] Mostafa Amini, Ali Bagheri, and Dursun Delen. 2022. Discovering injury severity risk factors in automobile crashes: A hybrid explainable AI framework for decision support. *Reliability Engineering & System Safety* 226 (2022), 108720.
[19] Stavros Antifakos, Nicky Kern, Bernt Schiele, and Adrian Schwaninger. 2005. Towards improving trust in context-aware systems by displaying system confidence. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*. 9–14.
[20] Pavlo Antonenko, Fred Paas, Roland Grabner, and Tamara Van Gog. 2010. Using electroencephalography to measure cognitive load. *Educational psychology review* 22, 4 (2010), 425–438.
[21] Amaël Arguel, Lori Lockyer, Ottmar V Lipp, Jason M Lodge, and Gregor Kennedy. 2017. Inside out: detecting learners' confusion to improve interactive digital learning environments. *Journal of Educational Computing Research* 55, 4 (2017), 526–551.
[22] Vijay Arya, Rachel KE Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C Hoffman, Stephanie Houde, Q Vera Liao, Ronny Luss, Aleksandra Mojsilović, et al. 2019. One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques. *arXiv preprint arXiv:1909.03012* (2019).
[23] Ronald T Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* (1997), 48.
[24] Aaron Bangor, Philip T Kortum, and James T Miller. 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction* 24, 6 (2008), 574–594.
[25] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (June 2020), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
[26] James Baumeister, Seung Youb Ssin, Neven AM ElSayed, Jillian Dorrian, David P Webb, James A Walsh, Timothy M Simon, Andrew Irlitti, Ross T Smith, Mark Kohler, et al. 2017. Cognitive cost of using augmented reality displays. *IEEE transactions on visualization and computer graphics* 23, 11 (2017), 2378–2388.
[27] Vaishak Belle. 2017. Logic meets Probability: Towards Explainable AI Systems for Uncertain Worlds.. In *IJCAI*. 5116–5120.
[28] Pascal Bercher, Susanne Biundo, Thomas Geier, Thilo Hoernle, Florian Nothdurft, Felix Richter, and Bernd Schattenberg. 2014. Plan, repair, execute, explain—how planning helps to assemble your home theater. In *Proceedings of the International Conference on Automated Planning and Scheduling*, Vol. 24. 386–394.
[29] Shlomo Berkovsky, Ronnie Taib, and Dan Conway. 2017. How to recommend? User trust factors in movie recommender systems. In *Proceedings of the 22nd international conference on intelligent user interfaces*. 287–300.
[30] Leopoldo Bertossi and Floris Geerts. 2020. Data quality and explainable AI. *Journal of Data and Information Quality (JDIQ)* 12, 2 (2020), 1–9.
[31] Umang Bhatt, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José MF Moura, and Peter Eckersley. 2020. Explainable machine learning in deployment. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 648–657.
[32] Biswarup Bhattacharya, Iftikhar Burhanuddin, Abhilasha Sancheti, and Kushal Satya. 2017. Intent-aware contextual recommendation system. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 1–8.
[33] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 Chi conference on human factors in computing systems*. 1–14.
[34] Marc Bolanos, Mariella Dimiccoli, and Petia Radeva. 2016. Toward storytelling from visual lifelogging: An overview. *IEEE Transactions on Human-Machine Systems* 47, 1 (2016), 77–90.
[35] Leonardo Bonanni, Chia-Hsun Lee, and Ted Selker. 2005. Attention-based design of augmented reality interfaces. In *CHI'05 extended abstracts on Human factors in computing systems*. 1228–1231.
[36] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.
[37] Andrea Brennen. 2020. What Do People Really Want When They Say They Want" Explainable AI?" We Asked 60 Stakeholders.. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–7.
[38] John Bridle. 1989. Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. *Advances in neural information processing systems* 2 (1989).
[39] Scott M Brown, Eugene Santos, and Sheila B Banks. 1998. Utility theory-based user models for intelligent interface agents. In *Conference of the Canadian Society for Computational Studies of Intelligence*. Springer, 378–392.
[40] Josef Buchner, Katja Buntins, and Michael Kerres. 2022. The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning* 38, 1 (2022), 285–303.
[41] Andrea Bunt, Matthew Lount, and Catherine Lauzon. 2012. Are explanations always important? A study of deployed, low-cost intelligent interactive systems. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*. 169–178.
[42] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society* 3, 1 (2016), 2053951715622512.
[43] Adrian Bussone, Simone Stumpf, and Dympna O'Sullivan. 2015. The role of explanations on trust and reliance in clinical decision support systems. In *2015 international conference on healthcare informatics*. IEEE, 160–169.
[44] Carrie J Cai, Jonas Jongejan, and Jess Holbrook. 2019. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th international conference on intelligent user interfaces*. 258–262.
[45] Carrie J Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S Corrado, Martin C Stumpe, et al. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–14.
[46] Wanling Cai, Yucheng Jin, and Li Chen. 2022. Impacts of Personal Characteristics on User Trust in Conversational Recommender Systems. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–14. https://doi.org/10.1145/3491102.3517471
[47] Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. 2015. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 1721–1730.
[48] Giuseppe Casalicchio, Christoph Molnar, and Bernd Bischl. 2018. Visualizing the feature importance for black box models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 655–670.
[49] Ajay Chander, Ramya Srinivasan, Suhas Chelian, Jun Wang, and Kanji Uchino. 2018. Working with beliefs: AI transparency in the enterprise. In *IUI Workshops*.
[50] Dimitris Chatzopoulos and Pan Hui. 2016. Readme: A real-time recommendation system for mobile augmented reality ecosystems. In *Proceedings of the 24th ACM international conference on Multimedia*. 312–316.
[51] Larissa Chazette, Oliver Karras, and Kurt Schneider. 2019. Do end-users want explanations? Analyzing the role of explainability as an emerging aspect of non-functional requirements. In *2019 IEEE 27th International Requirements Engineering Conference (RE)*. IEEE, 223–233.

[52] Xiaojun Chen, Shengbin Jia, and Yang Xiang. 2020. A review: Knowledge reasoning over knowledge graph. *Expert Systems with Applications* 141 (2020), 112948.

[53] Zhaorui Chen, Jinzhu Li, Yifan Hua, Rui Shen, and Anup Basu. 2017. Multimodal interaction in augmented reality. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 206–209.

[54] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. 2021. SemanticAdapt: Optimization-based Adaptation of Mixed Reality Layouts Leveraging Virtual-Physical Semantic Connections. In *Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology*. 16.

[55] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 4 (2014), 1–25.

[56] Eunji Chong, Nataniel Ruiz, Yongxin Wang, Yun Zhang, Agata Rozga, and James M Rehg. 2018. Connecting gaze, scene, and attention: Generalized attention estimation via joint modeling of gaze and scene saliency. In *Proceedings of the European conference on computer vision (ECCV)*. 383–398.

[57] Gabriele Cimolino and T.C. Nicholas Graham. 2022. Two Heads Are Better Than One: A Dimension Space for Unifying Human and Artificial Intelligence in Shared Control. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–21. https://doi.org/10.1145/3491102.3517610

[58] Pietro Cipresso, Irene Alice Chicchi Giglioli, Mariano Alcañiz Raya, and Giuseppe Riva. 2018. The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Frontiers in psychology* (2018), 2086.

[59] Roberto Confalonieri, Ludovik Coba, Benedikt Wagner, and Tarek R Besold. 2021. A historical perspective of explainable Artificial Intelligence. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 11, 1 (2021), e1391.

[60] Sven Coppers, Jan Van den Bergh, Kris Luyten, Karin Coninx, Iulianna Van der Lek-Ciudin, Tom Vanallemeersch, and Vincent Vandeghinste. 2018. Intellingo: An intelligible translation environment. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[61] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5828–5839.

[62] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Antonino Furnari, Evangelos Kazakos, Jian Ma, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, et al. 2022. Rescaling egocentric vision: collection, pipeline and challenges for epic-kitchens-100. *International Journal of Computer Vision* 130, 1 (2022), 33–55.

[63] Valdemar Danry, Pat Pataranutaporn, Yaoli Mao, and Pattie Maes. 2020. Wearable Reasoner: towards enhanced human rationality through a wearable device with an explainable AI assistant. In *Proceedings of the Augmented Humans International Conference*. 1–12.

[64] Devleena Das, Siddhartha Banerjee, and Sonia Chernova. 2021. Explainable ai for robot failures: Generating explanations that improve user assistance in fault recovery. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 351–360.

[65] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2014. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. *arXiv preprint arXiv:1408.6491* (2014).

[66] Anind K. Dey. 2001. Understanding and Using Context. *Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing* (2001), 304–307.

[67] Shipi Dhanorkar, Christine T. Wolf, Kun Qian, Anbang Xu, Lucian Popa, and Yunyao Li. 2021. Who needs to know what, when?: Broadening the Explainable AI (XAI) Design Space by Looking at Explanations Across the AI Lifecycle. In *Designing Interactive Systems Conference 2021*. ACM, Virtual Event USA, 1591–1602. https://doi.org/10.1145/3461778.3462131

[68] Stephen DiVerdi, Tobias Hollerer, and Richard Schreyer. 2004. Level of detail interfaces. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, 300–301.

[69] Finale Doshi-Velez and Been Kim. 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* (2017).

[70] Filip Karlo Došilović, Mario Brčić, and Nikica Hlupić. 2018. Explainable artificial intelligence: A survey. In *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*. IEEE, 0210–0215.

[71] Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. 2018. The Index of Pupillary Activity: Measuring Cognitive Load *vis-à-vis* Task Difficulty with Pupil Oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. https://doi.org/10.1145/3173574.3173856

[72] Lilian Edwards and Michael Veale. 2017. Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L. & Tech. Rev.* 16 (2017), 18.

[73] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19.

[74] Upol Ehsan, Samir Passi, Q Vera Liao, Larry Chan, I Lee, Michael Muller, Mark O Riedl, et al. 2021. The who in explainable ai: How ai background shapes perceptions of ai explanations. *arXiv preprint arXiv:2107.13509* (2021).

[75] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *23rd International Conference on Intelligent User Interfaces*. ACM, Tokyo Japan, 211–223. https://doi.org/10.1145/3172944.3172961

[76] Julian H Elliott, Anneliese Synnot, Tari Turner, Mark Simmonds, Elie A Akl, Steve McDonald, Georgia Salanti, Joerg Meerpohl, Harriet MacLehose, John Hilton, et al. 2017. Living systematic review: 1. Introduction—the why, what, when, and how. *Journal of clinical epidemiology* 91 (2017), 23–30.

[77] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. " I always assumed that close to [her]" Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 153–162.

[78] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *nature* 542, 7639 (2017), 115–118.

[79] João Marcelo Evangelista Belo, Anna Maria Feit, Tiare Feuchtner, and Kaj Grønbæk. 2021. Xrgonomics: Facilitating the creation of ergonomic 3d interfaces. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.

[80] Alireza Fathi, Ali Farhadi, and James M Rehg. 2011. Understanding egocentric activities. In *2011 international conference on computer vision*. IEEE, 407–414.

[81] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the world: 2D windows for 3D augmented reality. In *Proceedings of the 6th annual ACM symposium on User interface software and technology*. 145–155.

[82] James Fogarty, Scott E Hudson, Christopher G Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny C Lee, and Jie Yang. 2005. Predicting human interruptibility with sensors. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 1 (2005), 119–146.

[83] Christopher Frauenberger, Tony Stockman, and Marie-Luce Bourguet. 2007. A survey on common practice in designing audio in the user interface. In *Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK 21*. 1–9.

[84] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. 2014. How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies* 72, 4 (2014), 367–382.

[85] Melinda T Gervasio, Karen L Myers, Eric Yeh, and Boone Adkins. 2018. Explanation to Avert Surprise.. In *IUI Workshops*, Vol. 2068.

[86] Hristijan Gjoreski, Ivana Kiprijanovska, Simon Stankoski, Stefan Kalabakov, John Broulidakis, Charles Nduka, and Martin Gjoreski. 2021. Head-ar: Human activity recognition with head-mounted imu using weighted ensemble learning. In *Activity and Behavior Computing*. Springer, 153–167.

[87] David Gotz and Zhen Wen. 2009. Behavior-driven visualization recommendation. In *Proceedings of the 14th international conference on Intelligent user interfaces*. 315–324.

[88] Arthur C Graesser, Natalie Person, and John Huber. 1992. Mechanisms that generate questions. *Questions and information systems* 2 (1992), 167–187.

[89] Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, et al. 2022. Ego4d: Around the world in 3,000 hours of egocentric video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18995–19012.

[90] Shirley Gregor and Izak Benbasat. 1999. Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS quarterly* (1999), 497–530.

[91] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2016. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE transactions on visualization and computer graphics* 23, 6 (2016), 1706–1724.

[92] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Dino Pedreschi, Franco Turini, and Fosca Giannotti. 2018. Local rule-based explanations of black box decision systems. *arXiv preprint arXiv:1805.10820* (2018).

[93] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense advanced research projects agency (DARPA), nd Web* 2, 2 (2017), 1.

[94] David Gunning and David Aha. 2019. DARPA's explainable artificial intelligence (XAI) program. *AI magazine* 40, 2 (2019), 44–58.

[95] Agrim Gupta, Piotr Dollar, and Ross Girshick. 2019. Lvis: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5356–5364.

[96] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference

on Computer supported cooperative work. 241–250.

[97] Robert R Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. 2018. Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608* (2018).

[98] Daniel Holliday, Stephanie Wilson, and Simone Stumpf. 2016. User trust in intelligent systems: A journey over time. In *Proceedings of the 21st international conference on intelligent user interfaces*. 164–168.

[99] Yifei Huang, Minjie Cai, Zhenqiang Li, and Yoichi Sato. 2018. Predicting gaze in egocentric video by learning task-dependent attention transition. In *Proceedings of the European conference on computer vision (ECCV)*. 754–769.

[100] Emin İbili. 2019. Effect of Augmented Reality Environments on Cognitive Load: Pedagogical Effect, Instructional Design, Motivation and Interaction Interfaces. *International Journal of Progressive Education* 15, 5 (2019), 42–57.

[101] Farnaz Jahanbakhsh, Ahmed Hassan Awadallah, Susan T. Dumais, and Xuhai Xu. 2020. Effects of Past Interactions on User Experience with Recommended Documents. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. ACM, Vancouver BC Canada, 153–162. https://doi.org/10.1145/3343413.3377977

[102] Jinglu Jiang, Surinder Kahai, and Ming Yang. 2022. Who needs explanation and when? Juggling explainable AI and user epistemic uncertainty. *International Journal of Human-Computer Studies* 165 (Sept. 2022), 102839. https://doi.org/10.1016/j.ijhcs.2022.102839

[103] Yincheng Jin, Yang Gao, Xuhai Xu, Seokmin Choi, Jiyang Li, Feng Liu, Zhengxiong Li, and Zhanpeng Jin. 2022. EarCommand: "Hearing" Your Silent Speech Commands In Ear. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (July 2022), 1–28. https://doi.org/10.1145/3534613

[104] Antony William Joseph and Ramaswamy Murugesh. 2020. Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research. *J. Sci. Res* 64, 1 (2020), 168–175.

[105] Komal Kapoor, Karthik Subbian, Jaideep Srivastava, and Paul Schrater. 2015. Just in time recommendations: Modeling the dynamics of boredom in activity streams. In *Proceedings of the eighth ACM international conference on web search and data mining*. 233–242.

[106] Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. 2020. Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376219

[107] Mark T Keane and Eoin M Kenny. 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 *International Conference on Case-Based Reasoning*. Springer, 155–171.

[108] Eoin M Kenny, Courtney Ford, Molly Quinn, and Mark T Keane. 2021. Explaining black-box classifiers using post-hoc explanations-by-example: The effect of explanations and error-rates in XAI user studies. *Artificial Intelligence* 294 (2021), 103459.

[109] Dagmar Kern and Albrecht Schmidt. 2009. Design space for driver-based automotive user interfaces. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 3–10.

[110] Been Kim, Cynthia Rudin, and Julie A Shah. 2014. The bayesian case model: A generative approach for case-based reasoning and prototype classification. *Advances in neural information processing systems* 27 (2014).

[111] Keesung Kim, Jiyeon Hwang, Hangjung Zo, and Hwansoo Lee. 2016. Understanding users' continuance intention toward smartphone augmented reality applications. *Information development* 32, 2 (2016), 161–174.

[112] Pang Wei Koh and Percy Liang. 2017. Understanding Black-box Predictions via Influence Functions. In *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 1885–1894. https://proceedings.mlr.press/v70/koh17a.html ISSN: 2640-3498.

[113] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2019. Personalized explanations for hybrid recommender systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 379–390.

[114] TTC Labs. 2022. People-centric approaches to algorithmic explainability. https://www.ttclabs.net/report/people-centric-approaches-to-algorithmic-explainability

[115] Isaac Lage, Emily Chen, Jeffrey He, Menaka Narayanan, Been Kim, Samuel J Gershman, and Finale Doshi-Velez. 2019. Human evaluation of models built for interpretability. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. 59–67.

[116] Himabindu Lakkaraju, Stephen H Bach, and Jure Leskovec. 2016. Interpretable decision sets: A joint framework for description and prediction. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1675–1684.

[117] Himabindu Lakkaraju, Ece Kamar, Rich Caruana, and Jure Leskovec. 2019. Faithful and customizable explanations of black box models. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 131–138.

[118] Kit Yung Lam, Lik Hang Lee, and Pan Hui. 2021. A2W: Context-Aware Recommendation System for Mobile Augmented Reality Web Browser. In *Proceedings of the 29th ACM International Conference on Multimedia*. ACM, Virtual Event China, 2447–2455. https://doi.org/10.1145/3474085.3475413

[119] Markus Langer, Daniel Oster, Timo Speith, Holger Hermanns, Lena Kästner, Eva Schmidt, Andreas Sesing, and Kevin Baum. 2021. What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence* 296 (July 2021), 103473. https://doi.org/10.1016/j.artint.2021.103473

[120] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional.

[121] John Lee and Neville Moray. 1992. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35, 10 (1992), 1243–1270.

[122] Lik-Hang Lee, Tristan Braud, Pengyuan Zhou, Lin Wang, Dianlei Xu, Zijun Lin, Abhishek Kumar, Carlos Bermejo, and Pan Hui. 2021. All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda. http://arxiv.org/abs/2110.05352 Number: arXiv:2110.05352 arXiv:2110.05352 [cs].

[123] Teesid Leelasawassuk, Dima Damen, and Walterio W Mayol-Cuevas. 2015. Estimating visual attention from a head mounted IMU. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*. 147–150.

[124] Bruno Lepri, Nuria Oliver, Emmanuel Letouzé, Alex Pentland, and Patrick Vinck. 2018. Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology* 31, 4 (2018), 611–627.

[125] Benjamin Letham, Cynthia Rudin, Tyler H McCormick, and David Madigan. 2015. Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. *The Annals of Applied Statistics* 9, 3 (2015), 1350–1371.

[126] Xiao-Hui Li, Caleb Chen Cao, Yuhan Shi, Wei Bai, Han Gao, Luyu Qiu, Cong Wang, Yuanyuan Gao, Shenjia Zhang, Xun Xue, et al. 2020. A survey of data-driven and knowledge-aware explainable ai. *IEEE Transactions on Knowledge and Data Engineering* 34, 1 (2020), 29–49.

[127] Yang Li, Ranjitha Kumar, Walter S Lasecki, and Otmar Hilliges. 2020. Artificial intelligence for HCI: a modern approach. In *Extended Abstracts of the 2020 CHI conference on human factors in computing systems*. 1–8.

[128] Chen Liang, Chun Yu, Xiaoying Wei, Xuhai Xu, Yongquan Hu, Yuntao Wang, and Yuanchun Shi. 2021. Auth+Track: Enabling Authentication Free Interaction on Smartphone by Continuous User Tracking. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–16. https://doi.org/10.1145/3411764.3445624

[129] Peng Liao, Kristjan Greenewald, Predrag Klasnja, and Susan Murphy. 2020. Personalized HeartSteps: A Reinforcement Learning Algorithm for Optimizing Physical Activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (March 2020), 1–22. https://doi.org/10.1145/3381007

[130] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15. https://doi.org/10.1145/3313831.3376590 arXiv:2001.02478 [cs].

[131] Q Vera Liao and Kush R Varshney. 2021. Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790* (2021).

[132] Brian Y. Lim and Anind K. Dey. 2009. Assessing demand for intelligibility in context-aware applications. In *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, Orlando Florida USA, 195–204. https://doi.org/10.1145/1620545.1620576

[133] Brian Y. Lim and Anind K. Dey. 2010. Toolkit to support intelligibility in context-aware applications. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, Copenhagen Denmark, 13–22. https://doi.org/10.1145/1864349.1864353

[134] Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Boston MA USA, 2119–2128. https://doi.org/10.1145/1518701.1519023

[135] Sophia Lin. 2018. Explore and eat your way around town with google maps. https://www.blog.google/products/maps/explore-around-town-google-maps/

[136] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-Aware Online Adaptation of Mixed Reality Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, New Orleans LA USA, 147–160. https://doi.org/10.1145/3332165.3347945

[137] Zachary C Lipton. 2018. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16, 3 (2018), 31–57.

[138] Bingjie Liu. 2021. In AI we trust? Effects of agency locus and transparency on uncertainty reduction in human–AI interaction. *Journal of Computer-Mediated Communication* 26, 6 (2021), 384–402.

[139] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. 2020. Deep learning for generic object detection: A survey. *International journal of computer vision* 128, 2 (2020), 261–318.

[140] Mengchen Liu, Jiaxin Shi, Kelei Cao, Jun Zhu, and Shixia Liu. 2017. Analyzing the training processes of deep generative models. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 77–87.

[141] Tania Lombrozo. 2009. Explanation and categorization: How "why?" informs "what?". *Cognition* 110, 2 (2009), 248–253.

[142] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–16. https://doi.org/10.1145/3313831.3376727

[143] Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. 2017. Causal effect inference with deep latent-variable models. *Advances in neural information processing systems* 30 (2017).

[144] Feiyu Lu. 2021. Glanceable AR: Towards an Always-on Augmented Reality Future. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 717–718. https://doi.org/10.1109/VRW52623.2021.00241

[145] Feiyu Lu and Yan Xu. 2022. Exploring Spatial UI Transition Mechanisms with Head-Worn Augmented Reality. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–16. https://doi.org/10.1145/3491102.3517723

[146] Scott M. Lundberg and Su In Lee. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 2017-Decem, Section 2 (2017), 4766–4775.

[147] Weizhou Luo, Anke Lehmann, Hjalmar Widengren, and Raimund Dachselt. 2022. Where Should We Put It? Layout and Placement Strategies of Documents in Augmented Reality for Collaborative Sensemaking. In *CHI Conference on Human Factors in Computing Systems*. 1–16.

[148] Yifei Ma, Balakrishnan Narayanaswamy, Haibin Lin, and Hao Ding. 2020. Temporal-contextual recommendation in real-time. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2291–2299.

[149] Allan MacLean, Richard M. Young, Victoria M. E. Bellotti, and Thomas P. Moran. 1991. Questions, Options, and Criteria: Elements of Design Space Analysis. *Hum.-Comput. Interact.* 6, 3 (sep 1991), 201–250. https://doi.org/10.1207/s15327051hci0603%264_2

[150] Basim Mahbooba, Mohan Timilsina, Radhya Sahal, and Martin Serrano. 2021. Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model. *Complexity* 2021 (2021).

[151] Karl Manheim and Lyric Kaplan. 2019. Artificial intelligence: Risks to privacy and democracy. *Yale JL & Tech.* 21 (2019), 106.

[152] Rishabh Mehrotra, Mounia Lalmas, Doug Kenney, Thomas Lim-Meng, and Golli Hashemian. 2019. Jointly leveraging intent and interaction signals to predict user satisfaction with slate recommendations. In *The World Wide Web Conference*. 1256–1267.

[153] Antoine Miech, Jean-Baptiste Alayrac, Lucas Smaira, Ivan Laptev, Josef Sivic, and Andrew Zisserman. 2020. End-to-end learning of visual representations from uncurated instructional videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9879–9889.

[154] Brent Mittelstadt, Chris Russell, and Sandra Wachter. 2019. Explaining explanations in AI. In *Proceedings of the conference on fairness, accountability, and transparency*. 279–288.

[155] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2021. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Transactions on Interactive Intelligent Systems* 11, 3-4 (Dec. 2021), 1–45. https://doi.org/10.1145/3387166

[156] Christoph Molnar. 2020. *Interpretable machine learning*. Lulu. com.

[157] Tobias Müller and Ralf Dauenhauer. 2016. A taxonomy for information linking in augmented reality. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 368–387.

[158] Brad A Myers, David A Weitzman, Amy J Ko, and Duen H Chau. 2006. Answering why and why not questions in user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 397–406.

[159] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2018. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine* 52, 6 (May 2018), 446–462. https://doi.org/10.1007/s12160-016-9830-8

[160] Helen Nissenbaum. 2009. Privacy in context. In *Privacy in Context*. Stanford University Press.

[161] SS Muhammad Nizam, Rimaniza Zainal Abidin, Nurhazarifah Che Hashim, Meng Chun Lam, Haslina Arshad, and NAA Majid. 2018. A review of multimodal interaction technique in augmented reality environment. *Int. J. Adv. Sci. Eng. Inf. Technol* 8, 4-2 (2018), 1460.

[162] Xingjia Pan, Fan Tang, Weiming Dong, Chongyang Ma, Yiping Meng, Feiyue Huang, Tong-Yee Lee, and Changsheng Xu. 2019. Content-based visual summarization for image collections. *IEEE transactions on visualization and computer graphics* 27, 4 (2019), 2298–2312.

[163] Kihong Park, Seungryong Kim, and Kwanghoon Sohn. 2018. High-precision depth estimation with the 3d lidar and stereo fusion. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2156–2163.

[164] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond interruptibility: Predicting opportune moments to engage mobile phone users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–25.

[165] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond Interruptibility: Predicting Opportune Moments to Engage Mobile Phone Users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–25. https://doi.org/10.1145/3130956

[166] Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When attention is not scarce - detecting boredom from mobile phone usage. *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2015), 825–836. https://doi.org/10.1145/2750858.2804252 ISBN: 9781450335744.

[167] John Platt. 1998. Sequential minimal optimization: A fast algorithm for training support vector machines. (1998).

[168] Phillip E Pope, Soheil Kolouri, Mohammad Rostami, Charles E Martin, and Heiko Hoffmann. 2019. Explainability methods for graph convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10772–10781.

[169] Alun Preece, Dan Harborne, Dave Braines, Richard Tomsett, and Supriyo Chakraborty. 2018. Stakeholders in explainable AI. *arXiv preprint arXiv:1810.00184* (2018).

[170] Pearl Pu and Li Chen. 2006. Trust building with explanation interfaces. In *Proceedings of the 11th international conference on Intelligent user interfaces*. 93–100.

[171] Xun Qian, Fengming He, Xiyun Hu, Tianyi Wang, Ananya Ipsita, and Karthik Ramani. 2022. ScalAR: Authoring Semantically Adaptive Augmented Reality Experiences in Virtual Reality. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–18. https://doi.org/10.1145/3491102.3517665

[172] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[173] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*. PMLR, 8748–8763.

[174] Arun Rai. 2020. Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science* 48, 1 (2020), 137–141.

[175] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

[176] Gerhard Reitmayr and Dieter Schmalstieg. 2001. Mobile collaborative augmented reality. In *Proceedings IEEE and ACM International Symposium on Augmented Reality*. IEEE, 114–123.

[177] Alexandra Rese, Daniel Baier, Andreas Geyer-Schulz, and Stefanie Schreiber. 2017. How augmented reality apps are accepted by consumers: A comparative analysis using scales and opinions. *Technological Forecasting and Social Change* 124 (2017), 306–319.

[178] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why should i trust you?" Explaining the predictions of any classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 13-17-Augu (2016), 1135–1144. https://doi.org/10.1145/2939672.2939778 ISBN: 9781450342322.

[179] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2018. Anchors: High-precision model-agnostic explanations. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.

[180] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2018. Anchors: High-Precision Model-Agnostic Explanations. *Proceedings of the AAAI Conference on Artificial Intelligence* 32, 1 (April 2018). https://doi.org/10.1609/aaai.v32i1.11491

[181] Mireia Ribera and Agata Lapedriza. 2019. Can we do better explanations? A proposal of user-centered explainable AI.. In *IUI Workshops*, Vol. 2327. 38.

[182] Mark O Riedl. 2019. Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies* 1, 1 (2019), 33–36.

[183] Scott Robbins. 2019. A misdirected principle with a catch: explicability for AI. *Minds and Machines* 29, 4 (2019), 495–514.

[184] Quentin Roy, Futian Zhang, and Daniel Vogel. 2019. Automation Accuracy Is Good, but High Controllability May Be Better. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–8. https://doi.org/10.1145/3290605.3300750

[185] Rufat Rzayev, Susanne Korbely, Milena Maul, Alina Schark, Valentin Schwind, and Niels Henze. 2020. Effects of position and alignment of notifications on AR glasses during social interaction. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–11.

[186] Najmeh Samadiani, Guangyan Huang, Borui Cai, Wei Luo, Chi-Hung Chi, Yong Xiang, and Jing He. 2019. A review on automatic facial expression recognition systems assisted by multimodal sensor data. *Sensors* 19, 8 (2019), 1863.
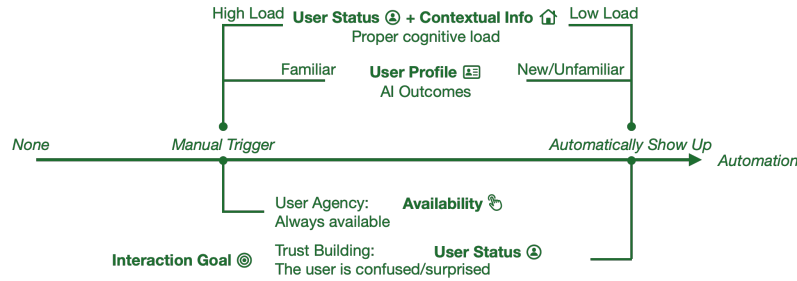
[187] Wojciech Samek and Klaus-Robert Müller. 2019. Towards explainable artificial intelligence. In *Explainable AI: interpreting, explaining and visualizing deep learning*. Springer, 5–22.

[188] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. 2017. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296* (2017).

[189] Hillol Sarker, Moushumi Sharmin, Amin Ahsan Ali, Md Mahbubur Rahman, Rummana Bari, Syed Monowar Hossain, and Santosh Kumar. 2014. Assessing the availability of users to engage in just-in-time intervention in the natural environment. *UbiComp 2014 - Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2014), 909–920. https://doi.org/10.1145/2632048.2636082 ISBN: 9781450329682.

[190] Udo Schlegel, Hiba Arnout, Mennatallah El-Assady, Daniela Oelke, and Daniel A Keim. 2019. Towards a rigorous evaluation of xai methods on time series. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 4197–4201.

[191] Philipp Schmidt, Felix Biessmann, and Timm Teubner. 2020. Transparency and trust in artificial intelligence systems. *Journal of Decision Systems* 29, 4 (Oct. 2020), 260–278. https://doi.org/10.1080/12460125.2020.1819094

[192] Johanes Schneider and Joshua Handali. 2019. Personalized explanation in machine learning: A conceptualization. *arXiv preprint arXiv:1901.00770* (2019).

[193] Tjeerd AJ Schoonderwoerd, Wiard Jorritsma, Mark A Neerincx, and Karel Van Den Bosch. 2021. Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human-Computer Studies* 154 (2021), 102684.

[194] Matthias Schröder and Helge Ritter. 2017. Deep learning for action recognition in augmented reality assistance systems. In *ACM SIGGRAPH 2017 Posters*. 1–2.

[195] A Carlisle Scott, William J Clancey, Randall Davis, and Edward H Shortliffe. 1977. *Explanation capabilities of production-based consultation systems*. Technical Report. STANFORD UNIV CA DEPT OF COMPUTER SCIENCE.

[196] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626.

[197] Shuai Shao, Zeming Li, Tianyuan Zhang, Chao Peng, Gang Yu, Xiangyu Zhang, Jing Li, and Jian Sun. 2019. Objects365: A large-scale, high-quality dataset for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*. 8430–8439.

[198] Donghee Shin. 2021. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies* 146 (2021), 102551.

[199] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. 2017. Learning important features through propagating activation differences. *34th International Conference on Machine Learning, ICML 2017* 7 (2017), 4844–4866. ISBN: 9781510855144.

[200] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034* (2013).

[201] Suriya Singh, Chetan Arora, and CV Jawahar. 2016. First person action recognition using deep learned descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2620–2628.

[202] Mohammad Soleymani, Sadjad Asghari-Esfeden, Yun Fu, and Maja Pantic. 2015. Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing* 7, 1 (2015), 17–28.

[203] Spotify. 2022. How Spotify's newest personalized experience, blend, creates a playlist for you and your bestie. https://newsroom.spotify.com/2021-08-31/how-spotifys-newest-personalized-experience-blend-creates-a-playlist-for-you-and-your-bestie/

[204] Lukas Stappen, Georgios Rizos, and Björn Schuller. 2020. X-aware: Context-aware human-environment attention fusion for driver gaze prediction in the wild. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. 858–867.

[205] Simone Stumpf, Adrian Bussone, and Dympna O'sullivan. 2016. Explanations considered harmful? user interactions with machine learning systems. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*.

[206] Ivan E Sutherland. 1968. A head-mounted three dimensional display. *Fall Joint Computer Conference* (1968), 8.

[207] William R Swartout. 1985. Explaining and justifying expert consulting programs. In *Computer-assisted medical decision making*. Springer, 254–271.

[208] Markus Tatzgern, Valeria Orso, Denis Kalkofen, Giulio Jacucci, Luciano Gamberini, and Dieter Schmalstieg. 2016. Adaptive information density for augmented reality displays. In *2016 IEEE Virtual Reality (VR)*. 83–92. https://doi.org/10.1109/VR.2016.7504691 ISSN: 2375-5334.

[209] J Eric T Taylor and Graham W Taylor. 2021. Artificial cognition: How experimental psychology can help generate explainable artificial intelligence. *Psychonomic Bulletin & Review* 28, 2 (2021), 454–475.

[210] Alan B Tickle, Robert Andrews, Mostefa Golea, and Joachim Diederich. 1998. The truth will come to light: Directions and challenges in extracting the knowledge embedded within trained artificial neural networks. *IEEE Transactions on Neural Networks* 9, 6 (1998), 1057–1068.

[211] Chung-Hsien Tsai and Jiung-Yao Huang. 2018. Augmented reality display based on user behavior. *Computer Standards & Interfaces* 55 (2018), 171–181.

[212] Hiroyuki Umemuro and Jun Yamashita. 2003. Detection of user's confusion and surprise based on pupil dilation. *The Japanese Journal of Ergonomics* 39, 4 (2003), 153–161.

[213] Lisa-Marie Vortmann, Felix Kroll, and Felix Putze. 2019. EEG-based classification of internally-and externally-directed attention in an augmented reality paradigm. *Frontiers in human neuroscience* 13 (2019), 348.

[214] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2017. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv. JL & Tech.* 31 (2017), 841.

[215] Ben Wagner, Krisztina Rozgonyi, Marie-Therese Sekwenz, Jennifer Cobbe, and Jatinder Singh. 2020. Regulating transparency? Facebook, twitter and the German network enforcement act. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 261–271.

[216] Chunting Wan, Dongyi Chen, Zhiqi Huang, and Xi Luo. 2021. A Wearable Head Mounted Display Bio-Signals Pad System for Emotion Recognition. *Sensors* 22, 1 (2021), 142.

[217] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–15. https://doi.org/10.1145/3290605.3300831

[218] Yuntao Wang, Xiyuxing Zhang, Jay M. Chakalasiya, Xuhai Xu, Yu Jiang, Yuang Li, Shwetak Patel, and Yuanchun Shi. 2022. HearCough: Enabling continuous cough event detection on edge computing hearables. *Methods* 205 (Sept. 2022), 53–62. https://doi.org/10.1016/j.ymeth.2022.05.002

[219] Jens Windau and Laurent Itti. 2016. Walking compass with head-mounted IMU sensor. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5542–5547.

[220] Philipp Wintersberger, Anna-Katharina Frison, Andreas Riener, and Tamara von Sawitzky. 2018. Fostering user acceptance and trust in fully automated vehicles: Evaluating the potential of augmented reality. *PRESENCE: Virtual and Augmented Reality* 27, 1 (2018), 46–62.

[221] Christine T. Wolf. 2019. Explainability scenarios: towards scenario-based XAI design. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. ACM, Marina del Ray California, 252–257. https://doi.org/10.1145/3301275.3302317

[222] Yao Xie, Melody Chen, David Kao, Ge Gao, and Xiang'Anthony' Chen. 2020. CheXplain: enabling physicians to explore and understand data-driven, AI-enabled medical imaging analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[223] Feiyu Xu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. 2019. Explainable AI: A brief survey on history, research areas, approaches and challenges. In *CCF international conference on natural language processing and Chinese computing*. Springer, 563–574.

[224] Jiahui Xu and Baichang Zhong. 2018. Review on portable EEG technology in educational research. *Computers in Human Behavior* 81 (2018), 340–349.

[225] Xuhai Xu, Prerna Chikersal, Afsaneh Doryab, Daniella K. Villalba, Janine M. Dutcher, Michael J. Tumminia, Tim Althoff, Sheldon Cohen, Kasey G. Creswell, J. David Creswell, Jennifer Mankoff, and Anind K. Dey. 2019. Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (Sept. 2019), 1–33. https://doi.org/10.1145/3351274

[226] Xuhai Xu, Prerna Chikersal, Janine M. Dutcher, Yasaman S. Sefidgar, Woosuk Seo, Michael J. Tumminia, Daniella K. Villalba, Sheldon Cohen, Kasey G. Creswell, J. David Creswell, Afsaneh Doryab, Paula S. Nurius, Eve Riskin, Anind K. Dey, and Jennifer Mankoff. 2021. Leveraging Collaborative-Filtering for Personalized Behavior Modeling: A Case Study of Depression Detection among College Students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (March 2021), 1–27. https://doi.org/10.1145/3448107

[227] Xuhai Xu, Alexandru Dancu, Pattie Maes, and Suranga Nanayakkara. 2018. Hand range interface: information always at hand with a body-centric mid-air input surface. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, Barcelona Spain, 1–12. https://doi.org/10.1145/3229434.3229449

[228] Xuhai Xu, Ahmed Hassan Awadallah, Susan T. Dumais, Farheen Omar, Bogdan Popp, Robert Rounthwaite, and Farnaz Jahanbakhsh. 2020. Understanding User Behavior For Document Recommendation. In *Proceedings of The Web Conference*. ACM, Taipei Taiwan, 3012–3018. https://doi.org/10.1145/3366423.3380071

[229] Xuhai Xu, Ebrahim Nemati, Korosh Vatanparvar, Viswam Nathan, Tousif Ahmed, Md Mahbubur Rahman, Daniel McCaffrey, Jilong Kuang, and Jun Alex Gao. 2021. Listen2Cough: Leveraging End-to-End Deep Learning Cough Detection Model to Enhance Lung Health Assessment Using Passively Sensed Audio. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (March 2021), 1–22. https://doi.org/10.1145/3448124

[230] Xuhai Xu, Haitian Shi, Xin Yi, Wenjia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K Dey. 2020. EarBuddy: Enabling On-Face Interaction via Wireless Earbuds. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 14.

[231] Xuhai Xu, Chun Yu, Yuntao Wang, and Yuanchun Shi. 2020. Recognizing Unintentional Touch on Interactive Tabletop. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (March 2020), 1–24. https://doi.org/10.1145/3381011

[232] Xuhai Xu, Tianyuan Zou, Han Xiao, Yanzhang Li, Ruolin Wang, Tianyi Yuan, Yuntao Wang, Yuanchun Shi, Jennifer Mankoff, and Anind K Dey. 2022. TypeOut: Leveraging Just-in-Time Self-Affirmation for Smartphone Overuse Reduction. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–17. https://doi.org/10.1145/3491102.3517476

[233] Zihan Yan, Yufei Wu, Yang Zhang, and Xiang'Anthony' Chen. 2022. EmoGlass: an End-to-End AI-Enabled Wearable Platform for Enhancing Self-Awareness of Emotional Health. In *CHI Conference on Human Factors in Computing Systems*. 1–19.

[234] Tal Yarkoni and Jacob Westfall. 2017. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science* 12, 6 (2017), 1100–1122.

[235] Su-Fang Yeh, Meng-Hsin Wu, Tze-Yu Chen, Yen-Chun Lin, XiJing Chang, You-Hsuan Chiang, and Yung-Ju Chang. 2022. How to Guide Task-oriented Chatbot Users, and When: A Mixed-methods Study of Combinations of Chatbot Guidance Types and Timings. In *CHI Conference on Human Factors in Computing Systems*.

[236] Hwanmoo Yong, Jisuk Lee, and Jongeun Choi. 2019. Emotion recognition in gamers wearing head-mounted display. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1251–1252.

[237] Xiaoyong Yuan, Pan He, Qile Zhu, and Xiaolin Li. 2019. Adversarial examples: Attacks and defenses for deep learning. *IEEE transactions on neural networks and learning systems* 30, 9 (2019), 2805–2824.

[238] Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2018. Studying eye movements as a basis for measuring cognitive load. In *Extended Abstracts of the 2018 CHI conference on human factors in computing systems*. 1–6.

[239] Fabio Massimo Zanzotto. 2019. Viewpoint: Human-in-the-loop Artificial Intelligence. *Journal of Artificial Intelligence Research* 64 (Feb. 2019), 243–252. https://doi.org/10.1613/jair.1.11345

[240] Matthew D Zeiler and Rob Fergus. 2014. Visualizing and understanding convolutional networks. In *European conference on computer vision*. Springer, 818–833.

[241] Xin Zhang, Armando Solar-Lezama, and Rishabh Singh. 2018. Interpreting neural network judgments via minimal, stable, and symbolic corrections. *Advances in neural information processing systems* 31 (2018).

[242] Yongfeng Zhang and Xu Chen. 2020. Explainable Recommendation: A Survey and New Perspectives. *Foundations and Trends® in Information Retrieval* 14, 1 (2020), 1–101. https://doi.org/10.1561/1500000066 arXiv: 1804.11192.

[243] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 295–305.

[244] Xingyi Zhou, Rohit Girdhar, Armand Joulin, Phillip Krähenbühl, and Ishan Misra. 2022. Detecting twenty-thousand classes using image-level supervision. *arXiv preprint arXiv:2201.02605* (2022).

[245] Fengyuan Zhu and Tovi Grossman. 2020. BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376233

[246] Jichen Zhu, Antonios Liapis, Sebastian Risi, Rafael Bidarra, and G. Michael Youngblood. 2018. Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*. 1–8. https://doi.org/10.1109/CIG.2018.8490433

[247] Robert Zimmermann, Daniel Mora, Douglas Cirqueira, Markus Helfert, Marija Bezbradica, Dirk Werth, Wolfgang Jonas Weitzl, René Riedl, and Andreas Auinger. 2022. Enhancing brick-and-mortar store shopping experience with an augmented reality shopping assistant application using personalized recommendations and explainable artificial intelligence. *Journal of Research in Interactive Marketing* (2022).
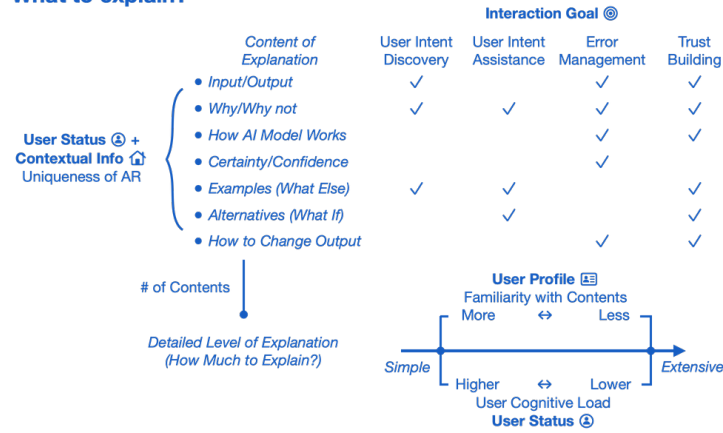
A APPENDIX A: EARLIER VERSIONS OF XAIR

We provide the initial versions of the framework that were used at the beginning of the three iterative workshops (from Fig. 10 to Fig. 12). These examples show how XAIR improved throughout the workshops.

Figure 10: Version 1 before The 1rd Iterative Expert Workshop (Study 2)

When to explain?

(A) Availability (Whether to explain?)
- Always Available (default)
- Turned off based on User Preference

System Goal ◎ User Agency
User Profile 🪪

(B) Automation (When to show-up?)

User Status ⊕ **+ Contextual Info** ⌂ Cognitive Load, Time Pressure — Enough Capacity — AND

User Profile 🪪 AI Outcomes — New/Unfamiliar — AND

User Goal 👤⊗ Resolving Confusion/Surprise, Informativeness
System Goal ◎ Error Management — OR — AND

User-trigger ●————————————● System-trigger

Default

What to explain?

		System Goal ◎				User Goal 👤⊗				User Profile 🪪	
(A) Content of Explanation (What content to include?)		Intent Discovery	Intent Assistance	Error Management	Trust Building	Resolving Confusion/Surprise	Privacy Awareness	Accountability	Informativeness	Low Digital Literacy	High Digital Literacy
• Input/Output		✓		✓	✓	✓	✓		✓	✓	✓
• Why/Why not		✓	✓	✓	✓	✓	✓	✓	✓	✓	
• How AI Model Works					✓		✓	✓	✓		✓
• Certainty/Confidence				✓	✓	✓		✓	✓	✓	✓
• Examples (What Else)		✓	✓		✓			✓	✓		✓
• Alternatives (What If)			✓		✓			✓	✓		✓
• How to Change Output				✓	✓	✓			✓		✓

User Status ⊕ **+ Contextual Info** ⌂ **Uniqueness of AI in AR**

Merge Goal Columns   Merge Goal Columns   Identify Profile Column

Find Overlapping Contents in Columns

(B) Details of Explanation (How much to explain?)

Multiple Contents in One Explanation
● Detailed

System Goal ◎ **+ User Goal** 👤⊗ User Agency & Informativeness: "More" Option upon User Request

- One Explanation
- Limited Content Ranking based on Current Contexts & Personal Preference

Contextual Info ⌂ **+ User Profile** 🪪

● High-level   Default

How to explain?

(A) Modality (Which modality to use?)
Audio
Visual
- Available Cognitive Capacity
- Noise & Visual Loading
- User preference

User Status ⊕ **Contextual Info** ⌂ **User Profile** 🪪

(B) Visual - Paradigm (In what visual paradigm?)

Format
Textual — Narrative, Conversation
Graphical — Cues, Highlights — Example Images, Heatmaps — Pattern
Contextual Info ⌂   Implicit   Explicit

(C) Visual - Position (Where to display?)

Semantic Adaptation
Contextual Info ⌂

World-referenced — e.g., event recommendation with explanation displayed on the calendar wall

Object-referenced — e.g., milk order reminder displayed on top of the milk bottle because milk is running low

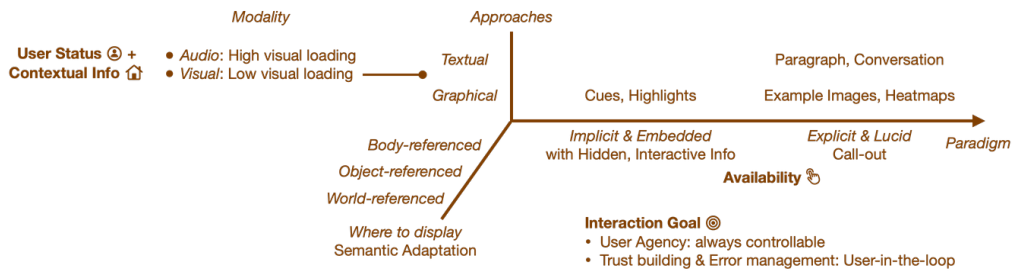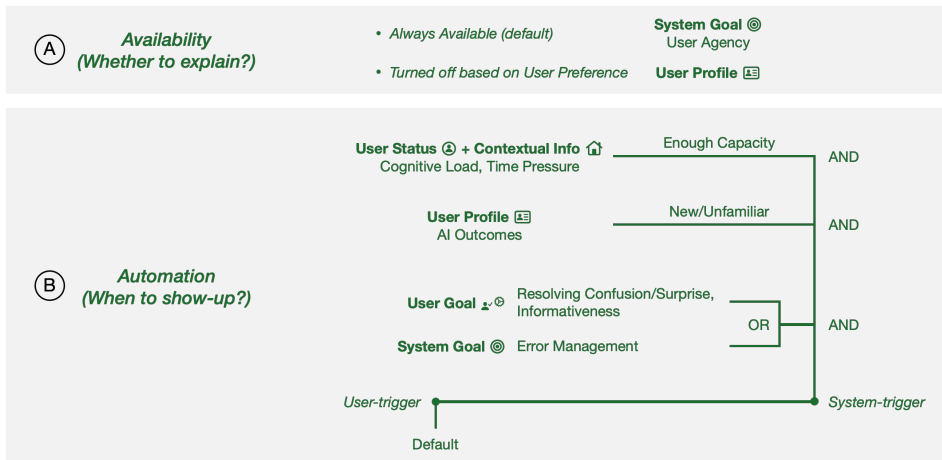Body-referenced — e.g., music suggestion and explanation pops up and floats along with the user as they move

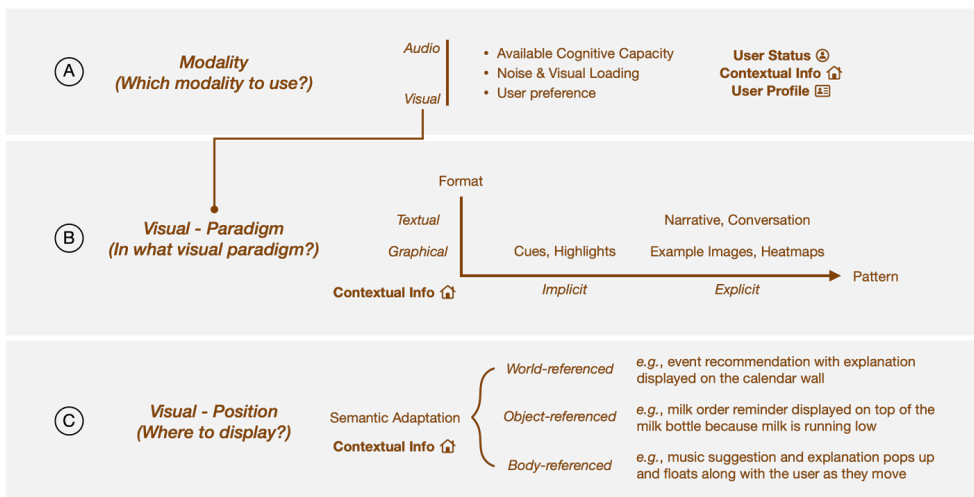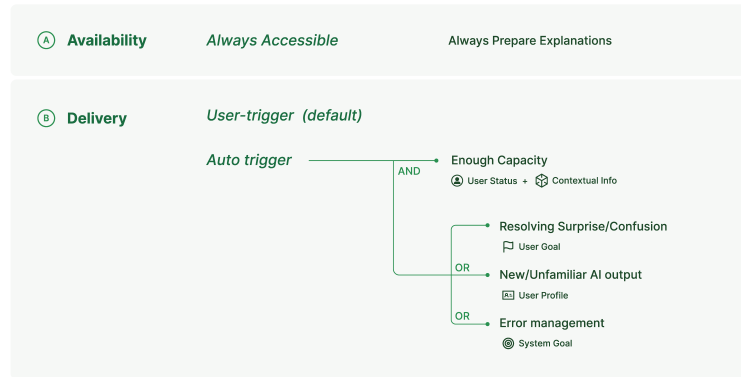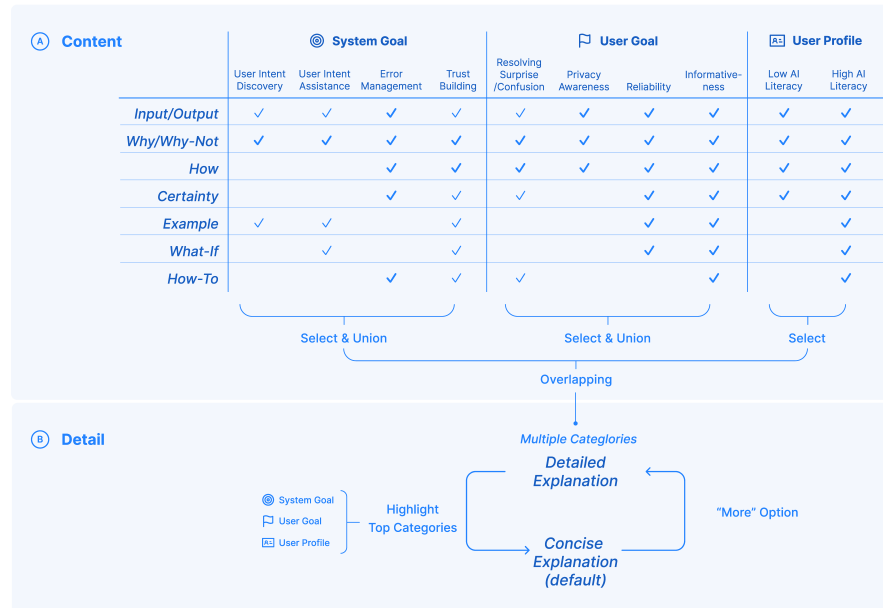**Figure 11: Version 2 before The 2rd Iterative Expert Workshop (Study 2). Main updates from Version 1: (*When*) Add dimensions and update the connection between the key factors and the dimensions. (*What*) Add *User Goal* and *User Profile* into the content type table. (*How*) Reorganize according to the dimensions and simply the structure.**

When to explain

(A) **Availability**	*Always Accessible*	Always Prepare Explanations

(B) **Delivery**	*User-trigger* (default)
	Auto trigger

AND → Enough Capacity
(user) User Status + (hex) Contextual Info

OR → Resolving Surprise/Confusion
(flag) User Goal

New/Unfamiliar AI output
(profile) User Profile

OR → Error management
(target) System Goal

What to explain

	System Goal				User Goal				User Profile	
(A) **Content**	User Intent Discovery	User Intent Assistance	Error Management	Trust Building	Resolving Surprise /Confusion	Privacy Awareness	Reliability	Informative-ness	Low AI Literacy	High AI Literacy
Input/Output	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Why/Why-Not	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
How			✓	✓	✓	✓	✓	✓	✓	✓
Certainty			✓	✓	✓		✓	✓	✓	✓
Example	✓	✓		✓			✓	✓		✓
What-If		✓		✓			✓	✓		✓
How-To			✓	✓	✓			✓		✓

Select & Union   Select & Union   Select

Overlapping

(B) **Detail**	*Multiple Categlories*

(target) System Goal
(flag) User Goal
(profile) User Profile

Highlight Top Categories

Detailed Explanation ⟷ "More" Option ⟷ *Concise Explanation (default)*

How to explain

(A) **Modality**	*Audio* ⟷ Audio/Haptic

High Load
(user) User Status
(hex) Contextual Info

Visual ⟷ *Visual*   AI Output

(B) **Paradigm [Visual]**	

Content-Env Compatibility
(hex) Contextual Info
High   Low

Implicit Pattern   *Explicit Pattern*

Textual Format   Cues   Narrative, Conversation

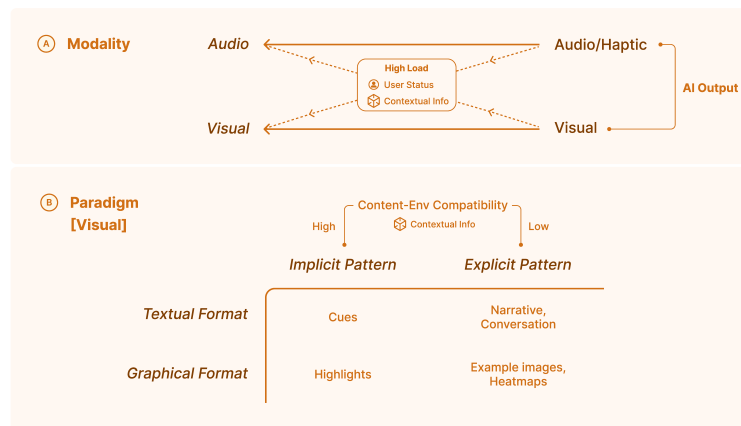*Graphical Format*   Highlights   Example images, Heatmaps

Figure 12: Version 3 before The 3rd Iterative Expert Workshop (Study 2). Main updates from Version 2: (*When*) Update the connection between the key factors and the dimensions. (*What*) Simplify the structure and provide default option as guidance. (*How*) Remove the "location" dimension and improve the visual design.

B APPENDIX B: DETAILS OF APPLICATION SCENARIOS

B.1 Explanation Details for The Two Applications

We present a more structured summary of the two scenarios in Sec. 6, together with examples of all explanation content type (Tab. 3).

Scenario Info		Scenario 1: Route Suggestion	Scenario 2: Plant Fertilization Reminder
Scenario		Nancy (AI expert, high AI literacy) is jogging in the morning on a quiet trail. Since it is the cherry-blossom season and Nancy loves cherries, her AR glasses display a map beside her and recommend a detour. Nancy is surprised since this route is different from her regular one, but she is happy to explore it. She is also curious to know the reason this new route was recommended.	Sarah (general end-user, low AI literacy) was chatting with her neighbor about gardening. After she returned home and sat on the sofa, her AR glasses recommended instructions about plant fertilization by showing a care icon on the plant. Sarah is concerned about technology invading her privacy, and wants to know the reason behind the recommendation.
Platform-Agnostic Key Factors	System Goal	**User Intent Discovery (new route)**	**Trust Building (clarification)**
	User Goal	**Resolve Surprise**	**Privacy Awareness**
	User Profile	**User Preference**: Cherry blossom tree lover; **History**: Regular jogging in the morning; **AI Literacy**: Expert, high	**User Preference**: Plant enthusiast; **History**: Did not care of the plant for a while; **AI Literacy**: General end-user, low
AR-Specific Key Factors	Contextual Info	**Location**: Outdoor; **Time**: Morning; **Environment**: Trails, streets	**Location**: Home; **Time**: Afternoon; **Environment**: Living room furniture, the plant
	User State	**Activity**: Jogging; **Cognitive Load**: Low	**Activity**: Sitting on the sofa; **Cognitive Load**: Low
Explanation Content Type Examples	Input/Output	This route is recommended based on seasons, your routine and preferences.	The system checks the plant's current status by visually scanning the plant.
	Why/Why-Not	The route has cherry blossom trees that you can enjoy. The length of the route is appropriate and fits your morning schedule.	The plant has abnormal spots on the leaves, which indicates fungi or bacteria infection.
	How	This algorithm finds and ranks possible routes based your location and other people who share similar preferences to you.	The system checks the plant's visual appearance, then searches online to find ways to cure it.
	Certainty	Match rate between this route's condition and your preference: 93%	The chance of the plant having disease is high (94%).
	Example	These photos captured memories about jogging during cherry blossom season.	These are some images of other plants with similar symptoms.
	What-If	The recommended route will be a shorter one if you jog in the evening.	N/A
	How-To	Disable the "season option" to return to the old route.	N/A

Table 3: Details of The Two Application Scenarios in Sec. 6. "N/A" indicates that this particular explanation type is not applicable for this case. The same below.

B.2 Additional Application Scenarios

We further applied XAIR to additional everyday AR scenarios to illustrate the practicability of XAIR. The four scenarios cover extra indoor & outdoor recommendations (Tab. 4), as well as AR-based intelligent instructions and automation aside from recommendations (Tab. 5).
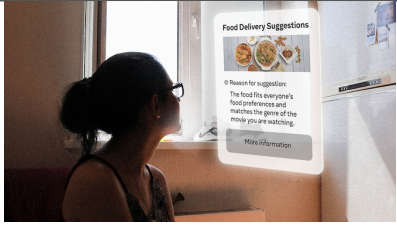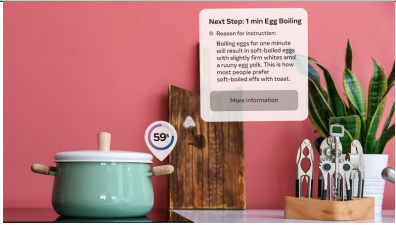

Scenario Info		Scenario 3: Food Rec for A Movie Night	Scenario 4: Podcast Rec while Driving
Scenario		Emma (general end-user, low AI literacy) has a few friends over for a small party. They decide to watch a Bollywood movie and now they are about to order food. The AR glasses recommends ordering from an Indian restaurant. Mary never heard of this restaurant before, but she loves this idea. She is also curious about the reason of this recommendation.	Jeff (general end-user, low AI literacy) is about to drive to work. The AR glasses recommends a new podcast "TEDx Shorts" that Jeff is unfamiliar with. However, the topic is interesting and Jeff wants to give it a try. Meanwhile, Jeff is curious to know the reason for this recommendation.
Platform-Agnostic Key Factors	System Goal	**User Intent Assistance** (find good food)	**User Intent Discovery** (new podcast)
	User Goal	**Reliability, Informativeness**	**Informativeness**
	User Profile	**User Preference**: Everyone's food preferences; **History**: Just decided to watch a Bollywood movie; **AI Literacy**: General end-user, low	**User Preference**: Topic interests; **History**: Morning driving routine; **AI Literacy**: General end-user, low
AR-Specific Key Factors	Contextual Info	**Location**: Indoor; **Time**: Evening; **Environment**: Home with a group of friends	**Location**: Outdoor; **Time**: Morning; **Environment**: Street conditions
	User State	**Activity**: Hanging out with friends; **Cognitive Load**: Low to Medium	**Activity**: About to Start Driving to Work; **Cognitive Load**: High
Explanation Content Type Examples	Input/Output	This restaurant is recommended based on everyone's food preference and movie genre.	The recommendation takes your playlist history and driving routine into account.
	Why/Why-Not	The food fits everyone's food preferences and matches the genre of the movie you are watching.	This podcast's topic is in line with your interest, and its length fits your expected driving time.
	How	The algorithm filters the restaurants by food preferences and then finds the best match between the food and the related activity.	The algorithm detects that it's morning and you are driving to work, then recommends the new podcast whose topic may be of interest to you.
	Certainty	Match score between the restaurant and the food preference and the movie: 90%	The podcast was liked by 85% of people with similar interests as you.
	Example	Last time, everyone enjoyed Chinese food while watching a Chinese movie.	"The Daily" and "Fresh Air" are other appropriate examples when you drove to work
	What-If	If movie genre is disabled, other cuisines would be recommended.	If the commute is longer, there are other episodes that may be of interest to you.
	How-To	N/A	To listen to previous podcasts, you can set history as the main recommendation factor.
XAI Designs in AR: *When*	Availability (*G1*)	Always available	Always available
	Delivery (*G2*)	Auto-trigger as both conditions is met (enough capacity and the user is not familiar with the recommendation).	Manual-trigger (high cognitive load during driving).
XAI Designs in AR: *What*	Content (*G3*)	Input/Output & Why/Why-Not	Input/Output & Why/Why-Not
	Detail - Concise (*G4*)	The Why part of the explanation examples.	The Why part of the explanation examples.
	Detail - Detailed (*G5*)	A list of the two explanation content types, plus images of the movie and food to support the Why part.	A list of the two explanation types.
XAI Designs in AR: *How*	Modality (*G6*)	Visual modality.	Audio modality
	Paradigm - Format (*G7*)	Textual format, plus graphical format in the detailed explanations.	N/A
	Paradigm - Pattern (*G8*)	Explicit pattern, presenting texts in the same window as the recommendations.	N/A

Table 4: Additional Application Examples of XAIR on Two Recommendation Scenarios.

Scenario Info		**Scenario 5: Cooking Instructions**	**Scenario 6: Automatic Do-Not-Disturb Mode**
Scenario		Lisa (general end-user, low AI literacy) has recently been learning how to cook. She wants to try out a new recipe for today's lunch. She picks "Poached Egg on Avacado Toast" and starts to follow the instructions. After she takes the eggs out of the fridge, the AR glasses prompts to boil the egg for 1 min. Lisa is curious about the time recommendation and wants to understand what the prompt is based on.	Jeff (general end-user, low AI literacy) is about to drive to work. The AR glasses recommends a new podcast "TEDx Shorts" that Jeff is unfamiliar with. However, the topic is interesting and Jeff wants to give it a try. Meanwhile, Jeff is curious to know the reason for this recommendation.
Platform-Agnostic Key Factors	**System Goal**	**User Intent Assistance** (learn the recipe)	**User Intent Discovery** (new podcast)
	User Goal	**Reliability**	**Informativeness**
	User Profile	**User Preference**: The purpose of learning how to cook; **History**: Following this instruction for the first time; **AI Literacy**: General end-user, low	**User Preference**: Topic interests; **History**: Morning driving routine; **AI Literacy**: General end-user, low
AR-Specific Key Factors	**Contextual Info**	**Location**: Kitchen; **Time**: Noon; **Environment**: Cookwares and ingredients	**Location**: Outdoor; **Time**: Morning; **Environment**: Street conditions
	User State	**Activity**: Cooking; **Cognitive Load**: High	**Activity**: About to Start Driving to Work; **Cognitive Load**: High
Explanation Content Type Examples	**Input/Output**	The guidance is based on the instruction and your current stage.	Last week you turned on smart do-not-disturb mode. The mode is based on your location, time, and your ongoing activity.
	Why/Why-Not	Boiling eggs for one minute will result in soft-boiled eggs with slightly firm whites and a runny egg yolk. This is how people prefer soft-boiled eggs with toast.	This setting automatically blocks notifications when you are at the office during the working hour and working on the laptop.
	How	The algorithm detects your activity and recognizes which stage you are in, then it provides the guidance for the next step.	The system detects your current context and activity, and checks whether they meet your authored settings. If so, the Do-Not-Disturb mode will be turned on.
	Certainty	The recognition of activity has a high certainty (88%).	The recognition of the time, location and current activity has a high certainty of 92%.
	Example	N/A	N/A
	What-If	Other possible ways of cooking eggs, such as scrambled eggs, if you want to explore other recipe instructions.	When you are not in the office, or it is out of working hours, or you are not working in front of the laptop, the setting will not be turned on.
	How-To	N/A	You can update any of the three conditions to change the moment the setting it's activated.
XAI Designs in AR: *When*	Availability (*G1*)	Always available	Always available
	Delivery (*G2*)	Manual-trigger (high cognitive load during cooking).	Manual-trigger (limited capacity in office).
XAI Designs in AR: *What*	Content (*G3*)	Input/Output & Why/Why-Not	Input/Output, Why/Why Not, How to, Confidence, How
	Detail - Concise (*G4*)	The Why part of the explanation examples.	A summary of Input, Why, and How-To as the user gets confused and wants to change the output.
	Detail - Detailed (*G5*)	A list of the two explanation types, plus images of the soft-boiled eggs to support the Why part.	A list of the five types.
XAI Designs in AR: *How*	Modality (*G6*)	Visual modality.	Visual modality
	Paradigm - Format (*G7*)	Textual format, plus graphical format in the detailed explanations.	Textual format
	Paradigm - Pattern (*G8*)	Explicit pattern, presenting texts in the window besides the timer.	Explicit pattern, presenting texts in front of the user.

Table 5: Additional Application Examples of XAIR on Intelligent Instructions and Automation.

B.3   Explanation Details of the Scenarios in Study 3 & 4

Tab. 6 shows the explanation examples presented to designers in Study 3.

Scenario Info		Case 1: Reliable Recipe Recommendation	Case 2: Wrong Recipe Recommendation
Explanation Content Type Examples	**Input/Output**	This recipe comes from the items detected in the fridge: egg and shrimp, and take your diet into account.	This recipe is based on friends' food preferences and the detected ingredients in your fridge: salmon and carrot.
	Why/Why-Not	This recipe fits your diet and food preference. It is recommended based on the rich amount of protein: 32g.	This recipe matches your friends' preference. It is recommended based on the popularity: 3201 people liked it.
	How	The algorithm recognizes ingredients in the fridge, finds and ranks recipes based on the available ingredients and your diet preference.	The algorithm first recognizes ingredients in the fridge, finds and ranks recipes based on the available ingredients and food preference.
	Certainty	Match rate between the recipe and the food preference & ingredients : 82%.	The recognition of salmon is uncertain (confidence 71%). It is not sure whether salmon or steak (recognition confidence 29%).
	Example	N/A	N/A
	What-If	More recipes if you want to try other cuisines.	Different recipes if your friends want to try other cuisines.
	How-To	Disable the diet option to see previous recipes before you went on the high-protein diet.	Select the right ingredients to change the recommendations: salmon or steak [clickable buttons].

Table 6: Details of The Two Cases in Sec. 7. Examples are "N/A" as they are already multiple examples in the recommendations.