Exploring the Impact of User and System Factors on Human-AI Interactions in Head-Worn Displays

Feiyu Lu[1,2] * Yan Xu[1] † Xuhai Xu[1,3] ‡ Brennan Jones[1] § Laird Malamed[1,4] ¶

[1]Reality Labs Research at Meta, Redmond, WA, USA
[2]Virginia Tech, Blacksburg, VA, USA
[3]University of Washington, Seattle, WA, USA
[4]University of Southern California, Los Angeles, CA, USA

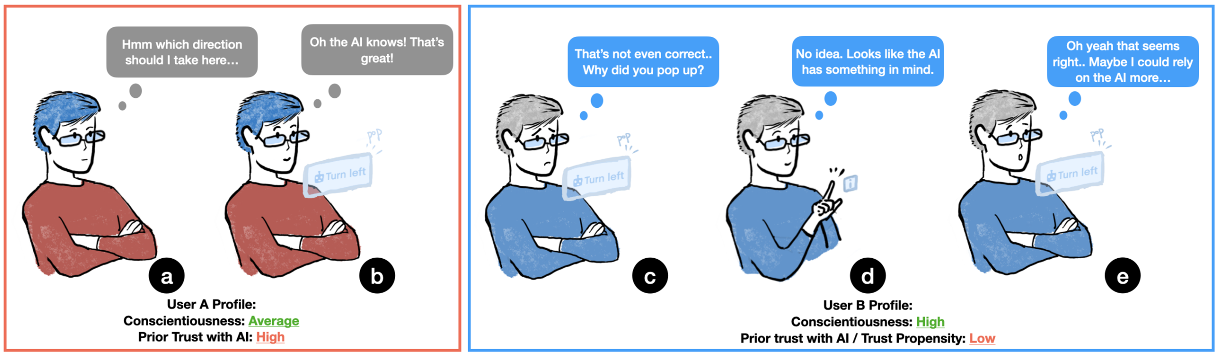

Figure 1: Illustrations of the impacts that user and system factors may pose on human-AI experiences in HWDs. In this scenario, a user wears a pair of AR glasses walking down the street, and tries to decide which direction to go next. Depending on the different user factors, the system provides the suggestions differently: (a-b) for user A with average *Conscientiousness* personality trait and high *Prior Trust with AI*, the system initiates the assistance once it detects that the user is in need of help. (c-e) For user B with high *Conscientiousness* and low *Prior Trust with AI* or *Trust Propensity*, (c) system-initiated assistance may risk lowering user trust and intent to use the system; (d-e) Instead, the system gives users the option to use the AI (i.e., (d) the user presses the inquiry button, then (e) the AI suggestion pops up), user's trust and adoption rate of the AI's suggestions could be improved.

Abstract

Empowered by the rich sensory capabilities and the advancements in artificial intelligence (AI), head-worn displays (HWD) could understand the user's contexts and provide just-in-time assistance to users' tasks to augment their everyday lives. However, there has been limited understanding of how users perceive interacting with AI services, and how different factors impact the user experience in HWD applications. In this research, we investigated broadly what user and system factors play important roles in human-AI experiences during an AI-assisted spatial task. We conducted a user study to simulate an everyday scenario where augmented reality (AR) glasses could provide suggestions/assistance. We researched three AI system factors (performance, initiation, transparency) with multiple user factors (personality traits, trust propensity, and prior trust with AI). We not only identified the impact of user traits such as the levels of conscientiousness and prior trust with the AI, but also found interesting interactions between them and system factors such as AI's performance and initiation strategy. Based on the findings, we suggest that future AI assistance on HWD needs to take users' individual characteristics into account and customize the system design accordingly.

*e-mail: feiyulu@vt.edu
†e-mail: yanx@meta.com
‡e-mail: xuhaixu@uw.edu
§e-mail: brennanj@meta.com
¶e-mail: lairdo@cinema.usc.edu

Index Terms: Human-centered computing—Interaction paradigms—Mixed/augmented reality; Human-centered computing—Interaction techniques

1 Introduction

One notable difference between spatial head-worn displays (HWD) and traditional mobile devices such as phones and smartwatches, lies in its potential for artificial intelligence (AI)-infused intelligent services to flourish. These devices come with rich on-device sensory capabilities (e.g., hand/face/eye trackers, world-facing cameras for spatial mapping and boundary/plane detection), which allow them to possess a deeper understanding of the user's physiological status, tasks, and environments more than any other mobile devices on the market. Combined with state-of-the-art AI systems, future HWDs would be able to sense, access and stream large amounts of contextual data, which can be fed into the AI models in order to deliver information spatially to the users at the right time and place, to assist their tasks anytime and anywhere, right when the user needs them [21, 37, 56]. Such context-aware and intelligent future has been credited in recent work of augmented/mixed reality (AR/MR) HWDs, that researchers believe the merging of spatial computing and AI would lead to a low-friction, non-intrusive, adaptive, and relevant information display future to assist the user's tasks. For example, based on the contextual data, the AI could decide where to place information [10, 34], how to present information [14, 35, 36], and how to offer assistance and explanation to users [21, 56].

Recent work in intelligent interfaces focuses primarily on developing algorithmic models that optimize the objective costs of the users' tasks [10, 26, 34]. For example, Cheng et al. studied methods to layout virtual content automatically with HWDs by minimizing a pre-defined cost function [9]. However, prior research in human-

AI (HAI) partnerships have shown that AI systems that maximize objective performance may not be considered more useful by the users [42, 51]. As people gain more exposure to AI in all facets of everyday tasks, researchers should care not only about the utilitarian values of AI systems, but also about how AI services make users feel, and how to form trustworthy, partnered, controllable, and favorable HAI experiences [2]. We argue that this would be especially true for future HWDs, in which intelligent assistance could be delivered directly in front of the user's eyes without spatial or temporal constraints. When, where, and how to show the assistance need to be more carefully considered so users do not feel intimidated, offended, confused, or annoyed by them [56].

Existing research in HAI has identified the impacts of a set of individual user/system factors on HAI experiences (see Sect. 2.3). However, most work only investigated the main effects of system factors, without taking into account the characteristics of different users, without investigating the interplay between multiple factors, without producing implications to spatial wearable computing devices such as HWDs and AR/MR systems.

In this research, we attempt to fill these gaps, by conducting an in-depth analysis on the impacts of multiple user and system factors on the perceived human-AI interaction experiences in spatial tasks with HWDs. While AI can be broadly defined, we focus on a specific application of AI in HWDs, where the AI assists the user by delivering recommendations based on its spatial understanding and predictions of user status, activity, and needs. This falls into the categories of spatial, wearable, and context-aware computings. For user factors, we studied the uses' *personality traits*, *trust propensity level (tendency to trust other people/digital services)*, and *prior trust levels with AI*. For system factors, we studied the *AI's performance*, *initiation strategy*, and *transparency level*. Through an VR-simulation study with eighty participants, our results shed light on how to deliver intelligent AI services in HWDs for everyday uses. To our knowledge, our work is one of the first that seeks to provide empirical evidence on how to deliver favorable human-AI interaction experiences in spatial wearable displays such as AR/MR HWDs covering a wide variety of user and system factors.

Our contributions in this paper are three-fold: (1) an in-depth exploration of the impacts of multiple system- and user-oriented factors on HAI experiences in HWDs-empowered spatial tasks; (2) empirical evaluation of how users perceive using AI assistance in everyday scenarios enabled by HWDs; (3) design considerations distilled to deploy favorable and trust-worthy human-AI experiences. Our findings provide valuable implications to spatial wearable computing and head-worn AR/MR applications.

2 Related Work

2.1 AI-assisted Decision Making

Users face uncountable decision-making tasks everyday. With the proliferation of AI applications and systems, humans and AIs are working collectively to conduct decision-makings tasks. These scenarios are also called human-AI teaming [60]. Such tasks span across low-stake mundane everyday activities (e.g., music/TV show recommenders, navigation) to high-stake professional workflows (e.g., clinical trials [31], court scenes [22]). In most of these activities, the AI would provide its recommendations or alternatives to address the user's needs, and it is up to the users whether or not to adopt the AI's suggestions. Either adopting or not adopting, the AI's suggestion could be part of or have an impact on the user's final decisions. With the recent development of mobile computing and especially AR/MR HWDs, future devices will be more likely to understand the user's contexts better, while at the same time surfacing the AI suggestions more frequently and just-in-time [21, 56]. It becomes increasingly critical to better understand how would users perceive different types of AI-empowered intelligent assistance in HWDs.

2.2 Intelligent User Interfaces in HWDs

One application of AI in HCI is intelligent user interfaces (IUI), which refer to interfaces that aim to enhance "*the efficiency, effectiveness, and naturalness of human-machine interaction by representing, reasoning, and acting on models of the user, domain, task, discourse, and media* [39]." There has been extensive research in deploying IUIs in mobile devices for tasks such as text-entry [59], navigation [46], context exploration [28, 57], and everyday recommendations [7]. IUIs serve as a medium for users to interact with AI-infused systems, and for AIs to understand user intent, make decisions, and act on behalf of the users.

More recently, enabled by the rich sensory capabilities of spatial wearable displays such as AR/MR HWDs, research starts to investigate how to deploy IUI to facilitate tasks in these devices. For example, Yu et al. explored interfaces that facilitate pointing/selection tasks in VR by recommending potential targets [58]. Lindlbauer et al. explored adapting the UI's level of detail and frame of reference based on user's workload and tasks [34]. Davari et al. explored alternating AR content's transparency and location to facilitate social interactions [14]. Lu and Xu explored methods to automatically transition UIs to different locations [37]. Cheng et al. proposed optimization methods that adapt the layout of AR windows when users move from one location to another [10]. Lages et al. explored user-triggered methods to change the layout and placements of AR windows in walking scenarios [29].

There are two common gaps in deploying HAI experiences in HWDs: (1) existing research focuses heavily on maximizing the objective performance (e.g., [34, 58]), with less emphasis on the *human* part of the HAI experiences; (2) existing research seems to indicate an *one-size-fits-all* solution, without taking into consideration the different characteristics of the users and how that may impact the user's perception of using intelligent services in spatial wearable computing devices (e.g., [13, 14, 29, 37]). In this research, we attempt to fill both gaps by exploring the impacts of both user profile and system design choices on user experience of using intelligent AI systems in HWDs during a spatial task. We focus on one specific application of IUI in HWDs, in which AI is used to offer intelligent assistance during a spatial task by making predictions about user status and needs.

2.3 Review of User & System Factors in HAI

Previous research, not in the wearable computing or AR/MR domain, but in the broad HAI domain, has identified some factors that may play a role in the user's experience of interacting with AI services. This serves as inspiration for our explorations. In this section, we detail the rationale of why we chose to study the specific factors, and how we believe these factors may play a role in HAI experiences in HWDs.

2.3.1 User factors

Prior study showed that the user's *personality* and *trust propensity* were likely to impact user's experience of interacting with AI.

Personality traits. According to Diener and Lucas, personality reflects *"people's characteristic patterns of thoughts, feelings, and behaviours [16]."* It reflects the basic dimensions in which people differ [38]. The Big-Five personality model has been widely used to determine the user's personality in five dimensions (*O-C-E-A-N*): *(O)penness*, *(C)onscientiousness*, *(E)xtroversion*, *(A)greeableness*, and *(N)euroticism* [40] (please see Table 1 for definitions). Research has indicated that some of these personality traits could influence the user's sense of trust and intention to use an AI system [8, 62]. For example, Cai et al. found that user's conscientiousness and neuroticism traits are likely to influence their trust with music recommenders [8]. However, it remains unclear whether such impacts would also apply to spatial wearable computing use cases with HWDs.

Table 1: A description of the seven user factors studied in the work.

Big Five Personality Traits [8,19,20]	(High ratings mean higher traits)
Openness	Routine, Conventional (Low) - Creative, Imaginative (High)
Conscientiousness	Impulsive, Disorganized (Low) - Prudent, Disciplined (High)
Extroversion	Reserved, Thoughtful (Low) - Sociable, Outgoing (High)
Agreeableness	Critical, Tough (Low) - Cooperative, Sympathetic (High)
Neuroticism	Unflappable, Confident (Low) - Emotionally sensitive, Anxious (High)
Trust Propensity & Prior Trust	(High ratings mean higher propensity/trust levels)
Trust Propensity [32]	How willing the person is to trust other people/new technologies? Hesitant, Unlikely to Trust (Low) - Inclined, Likely to Trust (High)
Prior Trust	How much does the user trust AI/computer recommendations prior to the study? Not trust AI at all (Low) - Trust AI a lot (High)

Trust propensity. Trust propensity is defined as the dispositional willingness to trust others [11]. It is a component in dispositional trust that could be influenced by the user's past experiences, cultural background, and personality traits [23]. Previous research has indicated that trust propensity could influence the user's formation of trust towards technologies [41] and the intention to use recommendation systems [8]. We hypothesize that this may also apply well in HWD applications, that the user's tendency to trust may impact their experience and willingness of using AI systems while wearing HWDs.

Prior trust with AI. Other than the propensity to trust, another important user factor is the prior trust with an AI system. The user's prior trust with the AI system determines the users attitude in the beginning of interactions with the AI, which could influence their initial trust while interacting with a new AI system. Early research has indicated the impact of prior trust attitudes on the final trust of AI assistance in high-stake scenarios [55]. To our knowledge, there has not been any research studying the potential impact of prior trust on intelligent services in spatial tasks involving HWDs.

2.3.2 System factors

AI's objective performance. As mentioned earlier, recent work in IUIs in HWD/AR/MR focuses on producing algorithmic models that optimize the objective performance or benefits of a certain outcome [9, 34, 45, 58]. The AI system's objective performance is one of the most intuitive measures that indicates its helpfulness in a given situation, but existing research also indicated that AI systems that perform better may not always be considered more useful/usable [42, 51, 63]. For example, Roy et al. found that users prioritize being in control more than high AI performance [51]. How do users perceive AI interfaces with different performance levels remains an underexplored and popular question in UI design. We decided to also dig in this direction, but in the use cases of assisting spatial tasks with HWDs. In this research, we explored two levels of performance: good performance ($\geq 75\%$ confidence) and medium performance ($\geq 50\%$ confidence). We decided to not include bad performance in our explorations, given that most available intelligent services on the market can deliver reasonable performance powered by big data and advancements on machine learning. We wanted to make our results more relevant and applicable to current situations.

Transparency. Although intelligent services often come with different performance levels, not all interfaces decide to make such information available to the users. Recent research in explainable and responsible AI has highlighted the need to increase the transparency of AI systems to make the AI's suggestions more intelligible [56]. A transparent system would allow users to understand why an AI makes a specific decision or exhibits a certain behavior [12, 54]. Confidence display (e.g., whether the confidence level of an AI's suggestion is made visible) is one strategy to increase the transparency of an AI system [6, 12]. In this research, we studied the effect of AI transparency, in particular the visibility of confidence display, on the user's perceived HAI experiences in spatial HWDs.

Initiation strategy. Intelligent AI assistance could be user-initiated or system-initiated. Initiation strategy has been extensively studied by previous research [1, 4, 24]. While user-initiated strategies may preserve agency to the greatest extent [51], it puts the interaction effort solely on the users [37]. System-initiated strategy, on the other hand, reduces the users effort by having high level of system automation at the cost of potentially reducing the user's sense of being in control [37, 51]. Most previous studies explored the potential impact of initiation strategy on the sense of control and agency [17, 37]. In this research, we would like to further study this popular factor, in the context of spatial HWD applications.

2.3.3 Implications to HWDs

Existing work that studied the impacts of user factors in HWD/AR/MR focus primarily on perceptions of avatar embodiment and presence [15, 27, 50], but rarely HAI experiences. The most relevant recent work to ours is [8] and [56]. Cai et al. studied the interplay between user factors with AI initiation strategy while interacting with a music recommender on a mobile phone [8]. However, the study did not explore other system factors such as performance and transparency, and it is unclear whether their findings would generalize to spatial wearable computing use cases with HWDs. Xu et al. explored how to present AI explanations in AR HWDs while considering user activity, goals, intent, and trust levels [56]. However, the paper did not empirically investigate user traits such as personality and how they interplay with different system designs.

In general, three directions remain underexplored: (1) there has been lack of explorations on the impact of user factors; (2) existing work rarely explored the combined effects of multiple factors; (3) existing work rarely produced any implications to spatial wearable computing applications. To our knowledge, our work is one of the first that complemented all the three gaps. Of course, there are factors that our study could not cover but were explored in prior research, such as explainability, timing, AI's representation, and diversity/serendipity. We call for future research to further study these factors. Our goal for this paper is to cover the factors that are more familiar to users, more suitable to spatial HWD experiences, and more relevant to everyday decision-making scenarios.

3 Experiment

3.1 Hypothesis

In this research, our goal is to Explore and identify the effects of the system factors, as well as their interactions with user factors, on the user's perceived experience of using intelligent assistance during a spatial task wearing HWDs. Based on existing work in the literature [8, 56], we hypothesize that:

H1. User factors would impact the perceived HAI experiences, especially the user's conscientiousness and trust propensity levels.

H2. There would exist interactions between user factors and system factors on the perceived HAI experience in HWD uses. As such, an AI system design that is favorable/unfavorable to one type of users may not be for others.

3.2 System

The study used a VR simulation of performing a spatial task and interacting with AI assistance while wearing a HWD. This approach, also known as VR simulation, has been proven to be an effective strategy to gather user feedback in a wide variety of experiment scenarios [5, 18, 33]. We improved the visual fidelity of the simulation by applying scenes with baked lighting and high-resolution textures (see Fig. 2).

3.3 Scenario & Task

To design the task scenario, we had three main criteria: (1) it needs to represent a spatial task using HWDs and AI systems; (2) it needs to retain characteristics of typical human-AI teaming/partnership situations, in which humans and AI work collectively to achieve a certain goal; (3) the user should be able to gain experience and learn from their past interactions to make better decisions (either from themselves or the AI), similar to common decision-making tasks.

Taking these into consideration, our scenario simulates a scenario in which HWDs could augment the user's memory and offer assistance in object searching tasks at home [47]. It is a combination of spatial-navigation and visual-search task designs that are well-utilized in AI and AR/MR research [30, 61]. During the task, users need to find multiple objects (e.g., watch, coat, keys) scattered in a VR home, which is made of four spaces: living room, office, bedroom, and bathroom (see Fig. 2). All objects were represented by a white sphere with a text of its name to limit the effect of different visual features on the object-searching efficiency. In each experiment condition, participants had to find and collect eight out of a total of fourteen randomly-placed objects. We chose number fourteen because we wanted to make the task slightly more challenging than human memory capacity, so there would likely always be a need for AI assistance [43].

A loosely body-fixed instruction panel, which was located above participants' front view, displayed the name of the next object to be collected for each trial. During each condition, the positions of the objects remained unchanged. Therefore, we expected participants to gain knowledge about where the uncollected items were located during the search, just like what they would do in real-life tasks.

3.4 Interacting with the VE & the AI

To find objects, the user needed to make a decision about which room to check out next. Teleportation was used as the main mechanism to travel within/between the rooms (see Fig. 3). Users can either choose which room to teleport to by themselves, or follow an AI's suggestion. When AI delivered a suggestion, it was presented on a panel that was loosely head-fixed and always visible to the users (see Fig. 4 (c) and (d)). To proceed, the users had to choose either one of the two available options: (1) *Guide me there*, in which the user chose to adopt the AI's suggestion. They would then sense a vibration on the left controller, with the corresponding room button highlighted guiding them to the suggested room; (2) *Dismiss*, in which the user could choose to discard the AI's suggestion and make their own decisions.

The AI's suggestion could be right or wrong. The occurrences of correct/wrong suggestions were determined by a pre-defined confidence value (see section below). If the AI makes a wrong suggestion, the user had to locate the target item by choosing and finding the correct room to teleport to by themselves. Within each room space, users could also point their hand at the ground to teleport to different spots within the room (see Fig. 3 (c)). However, to visit a different room, users had to use the virtual buttons attached to the controller. The reason for this was to make error-recovery cost consistent when the AI makes a mistake (i.e., bringing up the menu for rooms and clicking a button (see Fig. 3 (a-b))), as well as to mitigate the learning effect that participants become familiar with where each room is and could travel faster in later trials.

To collect the virtual objects, participants point the ray at the sphere and click on the trigger button (see Fig. 3 (d)). Participants have to find and collect the right object to proceed to the next trial.

Table 2: Descriptive statistics of the seven user factors for the 80 participants.

User Factors	N	Min	Max	Mean	SD
Openness	80	3.5	7	6.12	.88
Conscientiousness	80	3	7	6.06	1.06
Extroversion	80	1	7	5.01	1.49
Agreeableness	80	2	7	5.38	1.21
Neuroticism	80	2.5	7	5.51	1.18
Trust Propensity	80	2.5	7	4.41	1.17
Prior Trust	80	1	7	4.92	.97

3.5 System Factors & Interface Conditions

- *AI's objective performance:* The AI has two performance levels. For *good* performance, the AI makes the correct suggestion in 6 out of the 8 trials (75% confidence). For *medium* performance, the AI makes the correct suggestion in 4 out of the 8 trials (50% confidence). Note that these percentages are higher than a blind guess without any prior knowledge, which is 25%.

- *Initiation:* In the *user-initiated* situation, the AI suggestion is only visible if the user initiates it by pressing on the trigger button on the controller (see Fig. 4 (a-b)). In the *system-initiated* situation, the AI assistance appears automatically shortly after the beginning of each trial (see Fig. 4 (c)).

- *Transparency:* In *non-transparent* cases, the confidence display is invisible (see Fig. 4 (c)). In *transparent* cases, the confidence display of the AI's suggestion is visible, at the end of the suggestion sentence (see Fig. 4 (d)).

Taking into account the three system factors, we studied eight variations of the interface with different combinations of AI's *objective performances*, *initiation strategies*, and *transparency* levels.

3.6 Participants

In the screening phase, we recruited 104 participants on dscout. In the end, 80 of them (34 female, 43 male, 3 non-binary) completed the study with valid data entries and a balanced order of conditions, ranging from 15 - 74 years old (Mean = 33, SD = 12.28). All participants were regular Meta Quest 2 users. They self-reported no color-deficiency and had access to a 5 by 5 feet space to safely move around in their physical area for the study. Participants' distributions in terms of the user factors are shown in Table 2.

3.7 Study Design

We utilized a mixed design for the study, with *AI Performance* being the between-subject factor, while *Initiation* and *Transparency* being the within-subjects factors. As such, each participant interacted with an AI with the same performance in four conditions. The reason we chose *AI Performance* as the between-subject factor instead of *Initiation strategy* or *Transparency level* was to prevent participants' experiences from being biased by the fluctuating performance of the AI across sessions.

Since the task involves interacting with a simulated AI in the same virtual environment for multiple rounds, we attempted to mitigate

Figure 2: The virtual home environment used in the study: (a) a living room; (b) a bedroom; (c) a home office room; and (d) a bathroom. Users needed to traverse through the rooms and collect items represented by spheres and texts.

Figure 3: (a-b) The user utilizes the virtual buttons located on the left-hand controller to move between multiple indoor spaces (e.g., the user moves from the living room (a) to the bedroom (b)); the user uses the hand ray to (c) teleport in the same space, or (d) collect virtual objects in the scene. The user had to collect the right object to proceed to the next trial.

the potential ordering and learning effect using two methods: (1) we applied Latin Square counterbalancing to the order of testing on both performance levels groups, with an equal number of participants experiencing each ordering sequence; (2) in the beginning of each condition, we randomized the placement of the objects to be found, so participants always had to search even if they become familiar with the virtual environment.

The user factors were not experimentally controlled since it was unfeasible to control participants' personality and background. Instead, to balance the user factors, we conducted stratified sampling with anti-clustering to maximize the within-group heterogeneity and between-group similarity [48,49]. As such, each of the two between-subject groups had users with a wide range of personalities, genders, ages, trust propensities, and prior trust levels, but the two groups shared similar distributions for more valid comparisons later.

3.8 Measures

AI Usage/Adoption Rate: In the 32 rounds of decision making (8 rounds$\times$2 Initiation$\times$2 Transparency) per participant, we recorded AI usage rate - how often they decided to use the AI (only in the user-initiated conditions); and AI adoption rate - how often they followed the path suggested by the AI. In the user-initiated conditions, participants' behaviors only counted as adoptions if they used the AI and then followed the AI after using it.

Trust & Intent to Use: In the post-study survey, we asked each participant about their perceived level of trust and intent to continue using the AI system using two 7-point Likert-scale questions: *"I trust the AI assistant's recommendations;"* and *"If I were to do the task again, I would like to continue working with the AI assistant."*

Agency: After the participant experienced each condition, we asked them to rate on four questions regarding controllability and agency adapted from work by Tapal et al. [53].

User Preference & Comments: As we applied within-subjects design for *Initiation* and *Transparency* factors, we were able to gather more data about participants' preferences and comments on comparing different system-factor combinations. We asked participants to rank the four interfaces they experienced at the end of the study, followed by a video interview response about why they ranked the way they did and what they liked/disliked about each interface.

3.9 Experiment Procedure

We used dscout[1] for conducting the study. The study was approved by an ethics committee. The study was completely remote and unsupervised, and was divided to five phases. (1) We sent out a screening survey to dscout users and only recruited users who had access to a Meta Quest 2 VR headset and a 5x5 feet space. We also collected users' demographic information, personality traits and trust levels during the screening phase. We then conducted stratified sampling on the qualified participants based on their personalities, genders, ages, trust propensities, and prior trust levels. The procedure was achieved by the flexible recruiting and multi-phase study design features on the dscout platform. (2) Qualified participants were instructed to install the prototype software on their own Quest devices. (3) The participants started with a tutorial about the environment, controls, and tasks. (4) Participants experienced the four interface conditions one by one. Before the formal testing session of each interface condition, a training session was provided to participants in VR to teach them how to use the interface. The usage data were logged to a server. (5) After they finished the eight trials per condition, they were instructed to take off the VR headset, go to their computer and complete one page of the questionnaire on dscout. The questionnaire asked about the levels of agency and trust when interacting with the AI system. We also asked them to comment on what they liked and disliked about each condition they experienced. At the end of the study, participants were asked to rank the four interface conditions based on their preference. The study took about 80 minutes in total. Participants were compensated with $70 USD.

4 Results

To investigate both main effects and interactions of user and system factors, since we adopted a mixed-design in this study with repeated measures, we used generalized linear mixed-model (GLMM) to process our data in order to take into consideration both fixed and random effects, followed by sequential Bonferroni-corrected post-hoc tests for pairwise comparisons. We used a α value of 0.05 for all significance tests. For interactions between a pair of factors, we plotted the marginal effects of interaction terms. Following suggestions

[1] https://dscout.com/

Figure 4: A demonstration of the interfaces with different system factors: (a-b) in the user initiated situation, the AI assistance only appear when users click on the primary trigger button on the controller (i.e., do a fist gesture); (c) the non-transparent interface in which the AI did not display its confidence value; (d) the transparent interface with good performance, in which the AI displays its confidence value and has a good objective performance.

in previous work [44], we only interpreted main effects of factors that were not part of or affected by the presence of interactions.

4.1 Correlation between subjective measures and user factors

We conducted Pearson's correlation analysis on the dependent variables (i.e., *Trust, Intent to Use, Agency, Usage Rate, Adoption Rate*) to see how are they correlated with each other. Our results indicate a positive correlation between Trust and Intent to Use ($r = .65, p < .001$), Agency and Intent to Use ($r = .30, p < .001$) and Intent to Use and Adoption Rate ($r = .13, p = .03$). We also observed a positive correlation between Trust and Agency ($r = .30, p < .001$), between Trust and Conscientiousness trait ($r = .13, p < .025$), and between Conscientiousness and the intent to use the AI ($r = .14, p < .011$). In the user-initiated scenarios, the usage rate of the AI was positively correlated with the adoption rate of the AI ($r = .34, p < .001$) and the intent to use the AI ($r = .25, p < .001$). No other significant correlation was identified between a pair of variable.

4.2 AI Usage / Adoption Rate

We collected 80 (number of participants) $\times$ 8 (objects) $\times$ 2 (Initiation strategy) $\times$ 2 (Transparency) = 2,560 rounds of decision-making from all participants. Among all the rounds, participants decided to follow the AI in 1,503 rounds (58.71%). In all the 1,280 system-initiated rounds, users decided to follow the AI in 1,035 (80.86%) trials. In all the 1,280 rounds of user-initiated trials, users decided to use the AI in 526 (41.09%) trials. In the 526 used trials, participants decided to follow the AI's suggestion in 468 of the trials (88.97%).

GLMM revealed significant main effect of ***Initiation*** strategy on the adoption rate of AI suggestions ($\chi^2(1, N = 80) = 3.956, p = .047$). We also found borderline **interactions between *Initiation* and *AI Performance*** on the adoption rate ($\chi^2(1, N = 80) = 3.290, p = .070$). As shown in Fig. 6 (c), when the performance level was high, adoption rate was not influenced by initiation strategy. Users opted to adopt the AI's suggestion most of the time. However, when the performance level was low, the user-initiated condition led to a higher adoption rate than the system-initiated condition.

4.3 Trust and Intent to Use

Our results observed significant main effect of ***Conscientiousness*** ($\chi^2(1, N = 80) = 7.051, p = .007$) and ***Prior Trust*** ($\chi^2(1, N = 80) = 9.534, p = .002$) on the perceived Trust level on the AI system. We did not observe main effects of other system/user factors on the user's perceived Trust with the AI.

We found that ***Conscientiousness* interacted with *AI Performance*** on the trust level ($\chi^2(1, N = 80) = 4.714, p = .030$). As shown in Fig. 5 (a), people with a higher level of Conscientiousness (i.e., from impulsive to prudent) felt a decreased level of trust in the low-confidence AI condition. In contrast, for users who experienced the high-confidence AI condition, their trust level did not seem to be influenced much.

Prior Trust* was found to interact with *AI Performance on influencing the user's trust level and intent to use the AI system ($\chi^2(1, N = 80) = 7.129, p = .008$). Fig. 5 (b) and (c) show the interaction plots of marginal means. For the medium-performance condition, *Prior Trust* positively influenced the level of *Trust* with the AI system. However, for good-performance condition, the trust level was not affected by prior trust level. The trust level was more likely to be influenced by the performance of the AI system when the users hold low prior trust with AI systems.

Our results also revealed that ***Trust Propensity* interacted with *Initiation*** on the trust level ($\chi^2(1, N = 80) = 3.597, p = .030$). As indicated in Fig. 6 (a), users with lower trust propensity is more sensitive to initiation strategies. For users with high trust propensity, the difference between trust levels seems to be minimal between the two initiation strategies.

4.4 Agency

For main effects, GLMM found significant main effect of ***AI Performance*** ($\chi^2(1, N = 80) = 53.686, p < .001$) on the *Agency* level. Post-hoc analysis revealed that users perceived a lower sense of Agency while interacting with high performance AI as compared to low performance (p = .012).

The user's ***Conscientiousness*** ($\chi^2(1, N = 80) = 55.257, p < .001$) trait was found to influence the sense of Agency. The user's ***Conscientiousness* trait was also found to interact with *Initiation*** on the perceived agency level ($\chi^2(1, N = 80) = 5.072, p = .024$) (see Fig. 6 (b)). Participants with a higher *Conscientiousness* traits experienced a higher level of agency, and the user-initiated condition also led to a higher agency level than the system-initiated condition. However, the gap became smaller when the user's *Conscientiousness* level increased.

4.5 User Preferences

Fig. 7 shows the rankings of the four interface conditions for the medium and good performance levels. Out of the four conditions, the **user-initiated strategy with transparently displayed confidence** was liked the most under both AI performance levels (ranked 1st/2nd by 31/30 users for medium/good AI performance).

Friedman test revealed a significant main effect of interface condition on the mean ranking for both accuracy levels ($\chi^2(3) = 79.08, p < .001$ for 50% confidence and $\chi^2(3) = 77.92, p < .001$ for 75% confidence). For AI with 50% confidence, Bonferroni post-hoc comparisons showed that user-initiated transparent AI was ranked significantly higher than the other three conditions (all $p < .001$). System-initiated transparent AI was also ranked significantly higher than both user-initiated and system-initiated non-transparent conditions ($p < .001$). For AI with 75% confidence, Bonferroni post-hoc comparisons revealed that user-initiated transparent AI was significantly more preferred than both system-initiated conditions (both $p < .001$). No difference was observed between user-initiated transparent and user-initiated non-transparent conditions ($p = .192$).

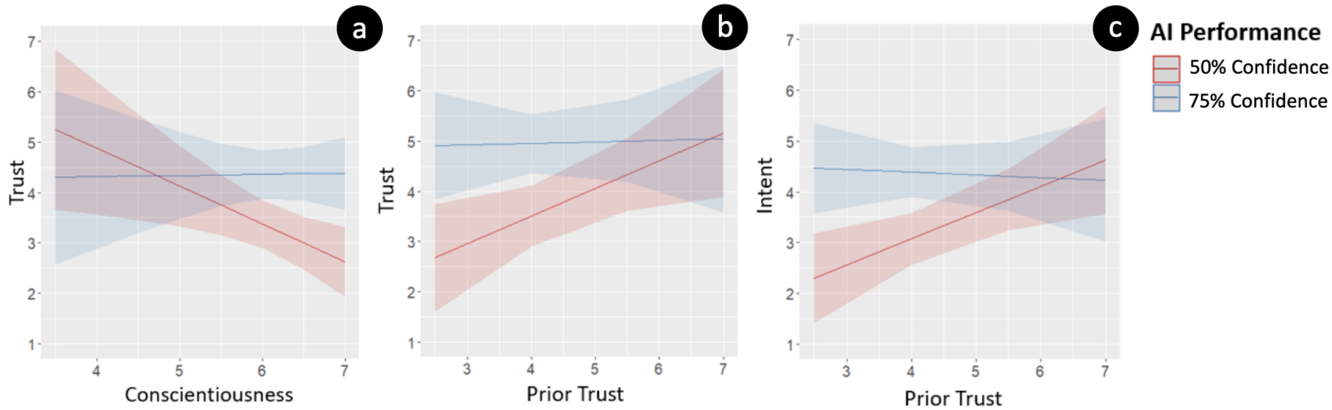

Figure 5: The interaction between *AI performance* and (a) *Conscientiousness*, (b,c) *Prior Trust* on the *Trust* and *Intent to Use* levels (areas indicate 95% confidence intervals).

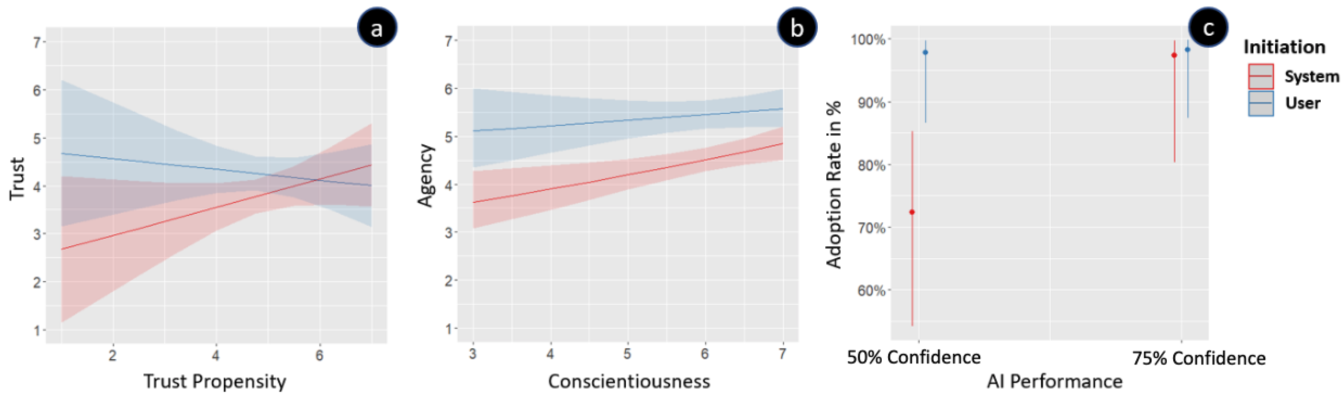

Figure 6: The interaction between *Initiation* and *Trust Propensity*, *Conscientiousness*, and *AI Performance* on the (a) *Trust*; (b) *Agency*; and (c) *Adoption Rate* levels (areas/bars indicate 95% confidence intervals).

5 Discussion

In our study, both of our hypothesis were supported. In the context of conducting an AI-assisted spatial task in HWDs, we not only identified factors that significantly impacted the adoption rate, trust or agency with AI systems, but also revealed the interactions between user-related factors and system-related factors.

First, our results showed that understanding users' traits, such as *Conscientiousness* level and *Prior Trust in AI*, will be critical when designing or evaluating HAI experiences, because:

(1) *Conscientiousness* and *Prior Trust in AI* significantly affected user's trust in using a new HAI system; and *Conscientiousness* had a significantly effect on the sense of agency. Therefore, when evaluating user's trust towards HAI systems, we need to measure and balance these two traits for different user groups to avoid confounding factors, since they could potentially affect the result significantly.

(2) *Conscientiousness* and *Prior Trust in AI* interact with the AI's performance level. If people have lower prior trust in AI or higher level of conscientiousness, their trust toward AI is more affected by the AI's performance (i.e. confidence). Previous research suggests that people with a higher conscientiousness level have higher trust in automated systems when conducting decision-making tasks [8]. In contrast, our study proved that this may not hold when the AI performance is not ideal. The reason could be that highly-conscientious users tend to be more responsible and strive to achieve better performance on their tasks [25], and they may get frustrated more with a lower-performance AI.

Secondly, our data shows that the *Initiation Strategy* (whether it is the user or the system that initiated the AI suggestions) is important for the following reasons:

(1) Users may not always initiate the AI suggestions, but when they do, they tend to adopt the suggestion more. In contrast, we surprisingly did not find AI performance having a significant main effect on the adoption rate.

(2) *Initiation Strategy* interacted with multiple other factors, including *AI Performance Level* and *Trust Propensity*. When the AI's performance level is lower, user-initiated suggestions had higher adoption rates than system-initiated suggestions. Users with lower trust propensity tended to trust the suggestions less when they were initiated by the system.

(3) When AI performance is high, users tend to prefer the user-initiated strategy. This could be explained by some of the users' comments about losing their sense of making a difference when the AI performs well and always pops up. As put by one of the participants, *"I felt like I didn't really do anything. I was kind of just going along with whatever the AI suggested."* This also aligns well with our finding that participants experienced a significantly lower sense of agency when the AI performance was high. Interestingly, the user group who experienced lower AI performance did not share a similar kind of frustration.

(4) We observed a somewhat low AI usage rate during user-initiated conditions (41.09%). From the participants' comments, the main reason could be that participants gained more task-related knowledge as they went through the tasks, which led to their in-

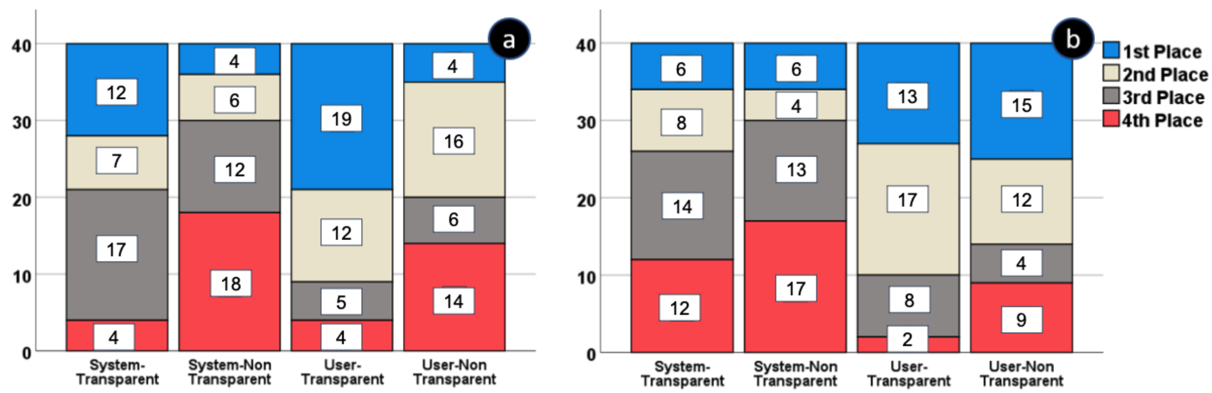

Figure 7: Ranking of the four interface conditions for (a) medium AI performance (50%); and (b) good AI performance (75%).

creased certainty and preference in making decisions by themselves rather than always relying on the AI suggestions. This result resonates with recent work in recommendation systems, which showed that users with high domain knowledge preferred to explore by themselves in AI-mediated context exploration scenarios [8].

Thirdly, although we did not find significant main or interaction effects regarding transparency, we observed that users preferred transparent, user-initiated conditions the most despite the AI performance being higher or lower. We believe that transparency is still an important factor, especially when combined with other factors such as initiation. Previous research has found that displaying the confidence level impacted user behaviors of using the AI system [3, 52]. Future research is needed to understand how transparency may impact the HAI experience with the variance of other factors.

6 Design Recommendations

Overall, our results suggest that instead of an *one-size-fits-all* approach, Human-AI systems in future HWD applications need to adapt to different user traits:

- Intelligent systems in HWDs need to have a higher performance threshold while providing suggestions to users with high *Conscientiousness* or low *Prior Trust* levels. Wrong or unfit suggestions may easily lead to lower trust for these users.
- We recommend HAI systems to give users control of whether and when to initiate the inquiry to AI for many reasons. It can increase users' agency especially when the *AI performance* is high or user *Conscientiousness* is low (e.g. more impulsive/disorganized). It can also increase the user's trust level compared to system initiated suggestions, when users *tend to trust others* less, or when the *AI performance* is lower. The caveat is that it costs an additional step for users to trigger the AI suggestion.
- When a system has no idea about user's traits like *Conscientiousness* or *Prior trust in AI*, we recommend having users initiate the AI suggestions and be transparent about the AI's performance, despite what the performance level is.

7 Limitations & Future Work

We identify some limitations to our work. First, we utilized teleportation as a fixed cost for better experimental control. This design choice makes the simulation less real-life like. Second, we studied performance level as a fixed value during participants' interactions with the AI. While we took this approach to simplify the experimental protocol which allowed us to answer our research questions more easily, we recognize that this presents some issues with generalizability. In real-world human-AI interactions, the AI's performance level will fluctuate depending on the scenario and task environment. Or, if the AI takes user-in-the-loop data as continuous input overtime, its performance could increase as the model is trained and customized with more data. Future studies could look into how different fluctuation patterns of AI performance levels could influence HAI experiences. Third, our study simulated short interactions with an AI in limited rounds of decision making. Future studies could look into longitudinal interactions in HWD-empowered decision-making scenarios, and how to build trust and partnership with users of different characteristics in the long run. Fourth, our participants had a slightly skewed distribution on the user factors. This could be attributed to the population distribution of the crowd-sourcing platform that we recruited participants from. Future studies could investigate a larger sample of participants with more even distribution on the user factors. Fifth, our scenario simulates a common everyday decision-making scenario powered by HWD/AR/MR devices, in which users have domain knowledge and gain knowledge along the way. Future studies could explore scenarios when the user's and AI's knowledge on the subject are at various levels. Last, the tasks simulated in our study is futuristic and may be challenging to implement in current HWDs. Future research could investigate applications scenarios that are more practical in current settings.

8 Conclusion

Our research was motivated to enhance the human-AI experience when users receive contextualized everyday-task suggestions in spatial tasks wearing head-worn displays. This work identified the user and system factors that impact how users trust, adopt, and feel agency with these suggestions, and how these factors interplay with each other. We not only found the individual factors that made a significant difference, such as *AI Performance*, *Initiation*, user's *Conscientiousness* and *Prior Trust in AI*, but also identified several significant interactions between factors, such as the interplay between the user's *Conscientiousness* and *Prior Trust in AI* with *AI Performance*; and the *Initiation Strategy* with *AI performance* and *Trust Propensity*. Our findings fill the gap of the prior research in which the factors were mostly studied separately and not in spatial wearable computing scenarios. Based on the findings, we provide concrete design recommendations about how to adapt the system design with different users' traits for tailored and favorable human-AI experiences in spatial head-worn display experiences.

Acknowledgments

The authors would like to thank Mohamamd Shahidul, Mary Hood, Missie Smith, Nahal Norouzi, and Sophie Kim from Reality Labs Research for their support on this work. The authors wish to thank the anonymous reviewers for their valuable comments and all participants for their time to take part in the study.

REFERENCES

[1] J. Allen, C. Guinn, and E. Horvtz. Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications*, 14(5):14–23, 1999. doi: 10.1109/5254.796083

[2] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz. Guidelines for human-ai interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300233

[3] S. Antifakos, N. Kern, B. Schiele, and A. Schwaninger. 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*, MobileHCI '05, p. 9–14. Association for Computing Machinery, New York, NY, USA, 2005. doi: 10.1145/1085777.1085780

[4] G. Bansal, T. Wu, J. Zhou, R. Fok, B. Nushi, E. Kamar, M. T. Ribeiro, and D. Weld. Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445717

[5] D. A. Bowman, C. Stinson, E. D. Ragan, S. Scerbo, T. Höllerer, C. Lee, R. P. McMahan, and R. Kopper. Evaluating effectiveness in virtual environments with mr simulation. In *Interservice/Industry Training, Simulation, and Education Conference*, vol. 4, p. 44, 2012.

[6] T. Bruzzese, I. Gao, G. Dietz, C. Ding, and A. Romanos. Effect of confidence indicators on trust in ai-generated profiles. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, p. 1–8. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3334480.3382842

[7] J. Budzik and K. J. Hammond. User interactions with everyday applications as context for just-in-time information access. In *Proceedings of the 5th International Conference on Intelligent User Interfaces*, IUI '00, p. 44–51. Association for Computing Machinery, New York, NY, USA, 2000. doi: 10.1145/325737.325776

[8] W. Cai, Y. Jin, and L. Chen. Impacts of personal characteristics on user trust in conversational recommender systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517471

[9] W.-H. Cheng and D. Gotz. Context-based page unit recommendation for web-based sensemaking tasks. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, IUI '09, p. 107–116. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1502650.1502668

[10] Y. Cheng, Y. Yan, X. Yi, Y. Shi, and D. Lindlbauer. Semanticadapt: Optimization-based adaptation of mixed reality layouts leveraging virtual-physical semantic connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, p. 282–297. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472749.3474750

[11] J. A. Colquitt, B. A. Scott, and J. A. LePine. Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance. *Journal of applied psychology*, 92(4):909, 2007. doi: 10.1037/0021-9010.92.4.909

[12] H. Cramer, V. Evers, S. Ramlal, M. Van Someren, L. Rutledge, N. Stash, L. Aroyo, and B. Wielinga. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-adapted interaction*, 18(5):455–496, 2008. doi: 10.1007/s11257-008-9051-3

[13] S. Davari, F. Lu, and D. A. Bowman. Occlusion management techniques for everyday glanceable ar interfaces. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 324–330, 2020. doi: 10.1109/VRW50115.2020.00072

[14] S. Davari, F. Lu, and D. A. Bowman. Validating the benefits of glanceable and context-aware augmented reality for everyday information access tasks. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 436–444, 2022. doi: 10.1109/VR51125.2022.00063

[15] D. Dewez, R. Fribourg, F. Argelaguet, L. Hoyet, D. Mestre, M. Slater, and A. Lécuyer. Influence of personality traits and body awareness on the sense of embodiment in virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 123–134, 2019. doi: 10.1109/ISMAR.2019.00-12

[16] E. Diener and R. E. Lucas. Personality traits. *General psychology: Required reading*, 278, 2019.

[17] L. Findlater and J. McGrenere. A comparison of static, adaptive, and adaptable menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, p. 89–96. Association for Computing Machinery, New York, NY, USA, 2004. doi: 10.1145/985692.985704

[18] J. L. Gabbard, J. E. Swan, and D. Hix. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence*, 15(1):16–32, 2006. doi: 10.1162/pres.2006.15.1.16

[19] S. D. Gosling, P. J. Rentfrow, and W. B. Swann. A very brief measure of the big-five personality domains. *Journal of Research in Personality*, 37(6):504–528, 2003. doi: 10.1016/S0092-6566(03)00046-1

[20] D. M. Greenberg, D. Müllensiefen, M. E. Lamb, and P. J. Rentfrow. Personality predicts musical sophistication. *Journal of Research in Personality*, 58:154–158, 2015. doi: 10.1016/j.jrp.2015.06.002

[21] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1706–1724, 2017. doi: 10.1109/TVCG.2016.2543720

[22] Y. Hayashi and K. Wakabayashi. Can ai become reliable source to support human decision making in a court scene? In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17 Companion, p. 195–198. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3022198.3026338

[23] K. A. Hoff and M. Bashir. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3):407–434, 2015. PMID: 25875432. doi: 10.1177/0018720814547570

[24] E. Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, p. 159–166. Association for Computing Machinery, New York, NY, USA, 1999. doi: 10.1145/302979.303030

[25] O. P. John and S. Srivastava. The big-five trait taxonomy: History, measurement, and theoretical perspectives. 1999.

[26] C. A. Johns, J. a. M. Evangelista Belo, C. N. Klokmose, and K. Pfeuffer. Pareto optimal layouts for adaptive mixed reality. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544549.3585732

[27] A. Katifori, C. Lougiakis, and M. Roussou. Exploring the effect of personality traits in vr interaction: The emergent role of perspective-taking in task performance. *Frontiers in Virtual Reality*, 3, 2022. doi: 10.3389/frvir.2022.860916

[28] A. Kumar, K. Srivastava, K. Yadav, and O. Deshmukh. Multi-faceted index driven navigation for educational videos in mobile phones. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, IUI '17, p. 357–361. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025171.3025221

[29] W. S. Lages and D. A. Bowman. Walking with adaptive augmented reality workspaces: Design and usage patterns. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, p. 356–366. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3301275.3302278

[30] A. Lakehal, S. Lepreux, C. Efstratiou, C. Kolski, and P. Nicolaou. Investigating smartphones and ar glasses for pedestrian navigation and their effects in spatial knowledge acquisition. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '20. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3406324.3410722

[31] M. H. Lee, D. P. Siewiorek, A. Smailagic, A. Bernardino, and S. Bermúdez i Badia. A human-ai collaborative approach for clinical decision making on rehabilitation assessment. In *Proceedings of*

the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445472

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

[33] Y. Li, F. Lu, W. S. Lages, and D. Bowman. Gaze direction visualization techniques for collaborative wide-area model-free augmented reality. In *Symposium on Spatial User Interaction*, SUI '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3357251.3357583

[34] D. Lindlbauer, A. M. Feit, and O. Hilliges. Context-aware online adaptation of mixed reality interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, p. 147–160. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3332165.3347945

[35] F. Lu and D. A. Bowman. Evaluating the potential of glanceable ar interfaces for authentic everyday uses. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 768–777, 2021. doi: 10.1109/VR50410.2021.00104

[36] F. Lu, S. Davari, L. Lisle, Y. Li, and D. A. Bowman. Glanceable ar: Evaluating information access methods for head-worn augmented reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 930–939, 2020. doi: 10.1109/VR46266.2020.00113

[37] F. Lu and Y. Xu. Exploring spatial ui transition mechanisms with head-worn augmented reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517723

[38] G. Matthews, I. J. Deary, and M. C. Whiteman. *Personality traits*. Cambridge University Press, 2003.

[39] M. Maybury. Intelligent user interfaces: An introduction. In *Proceedings of the 4th International Conference on Intelligent User Interfaces*, IUI '99, p. 3–4. Association for Computing Machinery, New York, NY, USA, 1998. doi: 10.1145/291080.291081

[40] R. R. McCrae and O. P. John. An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2):175–215, 1992. doi: 10.1111/j.1467-6494.1992.tb00970.x

[41] D. H. Mcknight, M. Carter, J. B. Thatcher, and P. F. Clay. Trust in a specific technology: An investigation of its components and measures. *ACM Trans. Manage. Inf. Syst.*, 2(2), jul 2011. doi: 10.1145/1985347.1985353

[42] S. M. McNee, J. Riedl, and J. A. Konstan. Being accurate is not enough: How accuracy metrics have hurt recommender systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, p. 1097–1101. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1125451.1125659

[43] G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956. doi: 10.1037/h0043158

[44] J. Neter, M. H. Kutner, C. J. Nachtsheim, W. Wasserman, et al. Applied linear statistical models. 1996.

[45] A. Niyazov, B. Ens, K. A. Satriadi, N. Mellado, L. Barthe, T. Dwyer, and M. Serrano. User-driven constraints for layout optimisation in augmented reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3580873

[46] P. Nurmi, A. Salovaara, S. Bhattacharya, T. Pulkkinen, and G. Kahl. Influence of landmark-based navigation instructions on user attention in indoor smart spaces. In *Proceedings of the 16th International Conference on Intelligent User Interfaces*, IUI '11, p. 33–42. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1943403.1943410

[47] J. Orlosky, T. Toyama, D. Sonntag, and K. Kiyokawa. Using eye-gaze and visualization to augment memory. In N. Streitz and P. Markopoulos, eds., *Distributed, Ambient, and Pervasive Interactions*, pp. 282–291. Springer International Publishing, Cham, 2014.

[48] M. Papenberg and G. W. Klau. Using anticlustering to partition data sets into equivalent parts. *Psychological Methods*, 26(2):161, 2021. doi: 10.1037/met0000301

[49] V. L. Parsons. *Stratified Sampling*, pp. 1–11. John Wiley & Sons, Ltd, 2017. doi: 10.1002/9781118445112.stat05999.pub2

[50] T. C. Peck and J. J. Good. Measuring embodiment: Movement complexity and the impact of personal characteristics. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–13, 2023. doi: 10.1109/TVCG.2023.3270725

[51] Q. Roy, F. Zhang, and D. Vogel. Automation accuracy is good, but high controllability may be better. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–8. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300750

[52] S. Sankaran, C. Zhang, H. Aarts, and P. Markopoulos. Exploring peoples' perception of autonomy and reactance in everyday ai interactions. *Frontiers in Psychology*, 12, 2021. doi: 10.3389/fpsyg.2021.713074

[53] A. Tapal, E. Oren, R. Dar, and B. Eitam. The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Frontiers in Psychology*, 8, 2017. doi: 10.3389/fpsyg.2017.01552

[54] N. Tintarev and J. Masthoff. Effective explanations of recommendations: User-centered design. In *Proceedings of the 2007 ACM Conference on Recommender Systems*, RecSys '07, p. 153–156. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1297231.1297259

[55] M. P. White, S. Pahl, M. Buehner, and A. Haye. Trust in risky messages: The role of prior attitudes. *Risk Analysis*, 23(4):717–726, 2003. doi: 10.1111/1539-6924.00350

[56] X. Xu, A. Yu, T. R. Jonker, K. Todi, F. Lu, X. Qian, J. a. M. Evangelista Belo, T. Wang, M. Li, A. Mun, T.-Y. Wu, J. Shen, T. Zhang, N. Kokhlikyan, F. Wang, P. Sorenson, S. Kim, and H. Benko. Xair: A framework of explainable ai in augmented reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581500

[57] K. Yadav, K. Shrivastava, S. Mohana Prasad, H. Arsikere, S. Patil, R. Kumar, and O. Deshmukh. Content-driven multi-modal techniques for non-linear video navigation. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, IUI '15, p. 333–344. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2678025.2701408

[58] D. Yu, R. Desai, T. Zhang, H. Benko, T. R. Jonker, and A. Gupta. Optimizing the timing of intelligent suggestion in virtual reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3526113.3545632

[59] M. R. Zhang, H. Wen, W. Cui, S. Zhu, H. Andrew Schwartz, X. Bi, and J. O. Wobbrock. *AI-Driven Intelligent Text Correction Techniques for Mobile Text Entry*, pp. 131–168. Springer International Publishing, Cham, 2021. doi: 10.1007/978-3-030-82681-9_5

[60] R. Zhang, N. J. McNeese, G. Freeman, and G. Musick. "an ideal human": Expectations of ai teammates in human-ai teaming. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW3), jan 2021. doi: 10.1145/3432945

[61] Y. Zhao, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. Evaluating augmented reality landmark cues and frame of reference displays with virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2710–2720, 2023. doi: 10.1109/TVCG.2023.3247078

[62] J. Zhou, S. Luo, and F. Chen. Effects of personality traits on user trust in human–machine collaborations. *Journal on Multimodal User Interfaces*, 14(4):387–400, 2020. doi: 10.1007/s12193-020-00329-9

[63] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, p. 22–32. Association for Computing Machinery, New York, NY, USA, 2005. doi: 10.1145/1060745.1060754