

Navigation Pixie: Implementation and Empirical Study Toward On-demand Navigation Agents in Commercial Metaverse

Hikari Yanagawa *
Cluster Metaverse Lab
University of Tsukuba

Yuichi Hiroi †
Cluster Metaverse Lab

Satomi Tokida ‡
The University of Tokyo
Cluster Metaverse Lab

Yuji Hatada §
The University of Tokyo

Takefumi Hiraki ¶
University of Tsukuba
Cluster Metaverse Lab



Figure 1: Our on-demand navigation agent implemented on a commercial metaverse platform. (a) Front and side views of the agent's 3D avatar design. The pixie-like design balances non-human, fictional characteristics with a non-verbal communication function for a tour guide. (b) The agent explaining the patio seating experience through voice and text, demonstrating spatial awareness and providing contextual information through natural language. (c) Actual interaction between a user and the agent on the commercial metaverse platform, showing the agent guiding the user to the bar and suggesting beverage options.

ABSTRACT

While commercial metaverse platforms offer diverse user-generated content, they lack effective navigation assistance that can dynamically adapt to users' interests and intentions. Although previous research has investigated on-demand agents in controlled environments, implementation in commercial settings with diverse world configurations and platform constraints remains challenging.

We present Navigation Pixie, an on-demand navigation agent employing a loosely coupled architecture that integrates structured spatial metadata with LLM-based natural language processing while minimizing platform dependencies, which enables experiments on the extensive user base of commercial metaverse platforms. Our cross-platform experiments on commercial metaverse platform Cluster with 99 PC client and 94 VR-HMD participants demonstrated that Navigation Pixie significantly increased dwell time and free exploration compared to fixed-route and no-agent conditions across both platforms. Subjective evaluations revealed consistent on-demand preferences in PC environments versus context-dependent social perception advantages in VR-HMD. This research contributes to advancing VR interaction design through conversational spatial navigation agents, establishes cross-platform evaluation methodologies revealing environment-

dependent effectiveness, and demonstrates empirical experimentation frameworks for commercial metaverse platforms.

Index Terms: On-demand Navigation Agent, Commercial Metaverse Platforms, User-generated Content

1 INTRODUCTION

The evolution of commercial metaverse platforms such as VR-Chat¹, Resonite², and Cluster³ has facilitated the formation of diverse virtual spaces centered on user-generated content (UGC) [31]. These UGC-based worlds encompass a wide range of purposes and structures, from art exhibitions to social events and gamified experiences, that encourage free-roaming and exploration within metaverse environments. However, this diversity presents challenges for visitors, who often struggle to comprehend destinations and construct meaningful exploration experiences.

In such environments, there is a growing demand for on-demand navigational assistance agents that can instantly respond to user interests and situations. Traditional agent functionality in commercial metaverse platforms has been implemented primarily through "non-intelligent agents," including script-based guides that follow predetermined routes and provide limited keyword responses [10, 17]. However, these approaches inadequately address users' ambiguous utterances or evolving interests, often limiting the exploration experience to superficial interactions.

The rapid development of Large Language Models (LLMs) has dramatically increased the feasibility of agents with real-time contextual understanding and natural dialog capabilities that were pre-

*e-mail: h.yanagawa@cluster.mu

†e-mail: y.hiroi@cluster.mu

‡e-mail: tokidasatomi@g.ecc.u-tokyo.ac.jp

§e-mail: hatada@nae-lab.org

¶e-mail: hiraki@slis.tsukuba.ac.jp

¹<https://hello.vrchat.com/>

²<https://store.steampowered.com/app/2519830/Resonite/>

³<https://cluster.mu/en>

viously difficult to achieve [29, 39, 41]. LLM-based agents that integrate speech, language, and spatial understanding can flexibly select actions based on user requests, effectively acting as “exploration companions” [5]. Such agents go beyond simply providing functionality and have the potential to transform the fundamental quality of the user experience. However, previous research on navigational assistance agents has mostly focused on specific researcher-designed systems or VR spaces. While these efforts have provided valuable insights into agent design and evaluation, most rely on specialized environment configurations that lack the portability required for direct implementation in commercial metaverse environments.

Furthermore, metaverse spaces centered on UGC present additional complexities: world configurations and interaction designs vary among creators, while platform-specific API specifications and development constraints impose further limitations. As a result, design insights for implementing agents that can be deployed by non-specialists remain limited, especially when considering these operational complexities and differences. This situation significantly impedes the enhancement of user experience in commercial metaverse platforms where UGC constitutes the core component.

This research addresses these challenges by proposing “Navigation Pixie,” an on-demand navigation agent with a loosely coupled architecture that users can implement themselves on commercial metaverse platforms. By integrating structured spatial information from metaverse environments with the natural language understanding capabilities of large language models, this agent provides flexible guidance that responds to user requests while accommodating the diverse configurations of different UGC worlds.

Beyond demonstrating technical feasibility, this research examines how Pixie’s intelligent flexibility impacts the actual user experience by comparing it to script-based fixed route agents (conventional “non-intelligent agents”), measuring changes in exploration behavior and immersive experiences across both PC client and VR-HMD platforms. Notably, these experiments used fully operational implementations on a commercial metaverse platform, Cluster, with real users. This approach allowed us to gather 99 PC client participants within 2 days and 94 VR-HMD participants within 5 days, thus achieving scalable empirical evaluation, and demonstrates the potential of using commercial metaverse platforms as socially implemented research infrastructure.

This paper discusses the potential qualitative transformation of metaverse experiences through LLM-based interactive agents and the methodological developments in agent research utilizing metaverse spaces, through the design architecture of Navigation Pixie, its specific implementation structure, and analysis based on user experiments.

2 RELATED WORK

2.1 Embodied Navigation Agents in VR

The development of embodied conversational agents that function as tour guides has been driven by spatial disorientation and wayfinding difficulties in large-scale 3D environments. Early implementations emerged in the form of physical museum tour robots around 2000 [8, 37], while the first virtual implementations appeared with Jan et al. interactive agents in Second Life [17]. The embodiment of these agents significantly improves the perception of companionship and immersion [25, 21]. In addition, their nonverbal communication capabilities, such as gaze direction and pointing gestures, enable them to control attention and time information similar to human museum curators [10, 18].

Traditional virtual guides have predominantly employed fixed-route approaches, ensuring comprehensive exhibit coverage while constraining user exploration freedom [10, 36, 7]. In contrast, companion-type agents that dynamically respond to user interests offer greater flexibility, despite implementation complexity [6].

While systems like English-learning NPCs on VRChat demonstrate conversational capabilities [29], they typically prioritize educational dialogue over spatial navigation functionality.

Recent advances in LLMs have accelerated the research of embodied AI agents that integrate language, vision, and spatial information [15, 26], with preliminary investigations conducted in a controlled virtual museum environment [42]. Our research builds on these approaches by implementing embodied agents that integrate LLMs with spatial information on commercial VR platforms.

2.2 On-demand Navigation

VR environments present unique challenges regarding self-localization and route formation due to their structural flexibility compared to physical spaces. To address these challenges while preserving user autonomy, on-demand approaches that provide guidance upon request have emerged. Early work introduced adaptive guides in web-based virtual museums, proposing various information delivery methods based on user models and browsing history [24]. More recent developments include a visual guide agent by Collins et al. for visually impaired users in social VR spaces, which provides environment descriptions and navigation support in response to user requests [11].

Ye et al. provide positioning optimization for guide agents that facilitate natural dialogue-based guidance in VR [43], while Hamady et al. developed guides that integrate real exhibition spaces with MR in HoloLens, improving visitor satisfaction through rich 3D displays and audio guidance [16]. However, architectural design and large-scale user evaluation methodologies for on-demand agents in diverse UGC environments within commercial metaverse platforms remain insufficiently explored.

2.3 Experimentation in Metaverse Platforms

Recently, commercial metaverse platforms such as VRChat and Cluster have attracted attention as large-scale, remote VR experimental infrastructures due to their diverse and continuously connected user bases [30, 20]. Using commercial platforms effectively incorporates realistic diversity, which is difficult to replicate in laboratory environments [7, 46]. Saffo et al. [34, 33] and Kurai et al. [20, 19] demonstrated the potential of crowdsourced VR experiments by collecting user experience data and behavioral logs through commercial metaverse platforms. However, they also noted that reliance on the existing functionality of commercial metaverses creates limitations in experimental world design and data collection. Additionally, platform-specific APIs and version constraints pose challenges that necessitate purpose-appropriate architectural design.

UGC-centered environments, in particular, require technical innovation for external control and data acquisition from independently developed elements [45]. Although experimental platforms such as Ubiq [13, 35, 27, 14] and Ouvrai [9] have been developed to address these challenges, they have yet to achieve user bases comparable to those of commercial systems in terms of diversity and scale. Our research implements a loosely coupled architecture that reduces dependence on platforms, creating an agent evaluation environment that leverages commercial metaverse platforms with large user bases.

3 SYSTEM ARCHITECTURE

This section describes the Navigation Pixie architecture, an on-demand navigation agent designed for commercial metaverse platforms, which enables platform-independent operation through loose coupling.

3.1 Design Principles

Our Navigation Pixie architecture follows four core principles:

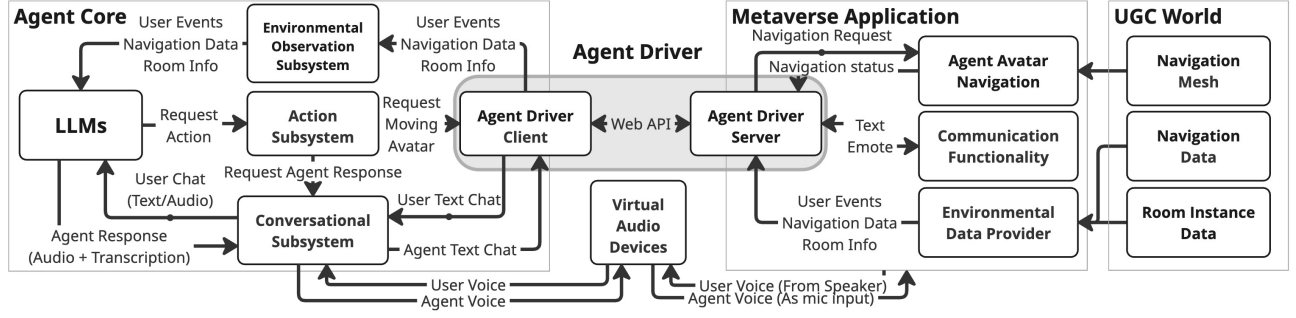


Figure 2: Architectural diagram of the Navigation Pixie system. The commercial metaverse platform and the external Agent Core, which serves as the cognitive center of the agent, are loosely coupled only through the Agent Driver, a compatibility layer that ensures platform independence. This design achieves effective navigation support by integrating spatial understanding and natural language processing while clearly separating platform-dependent processes.

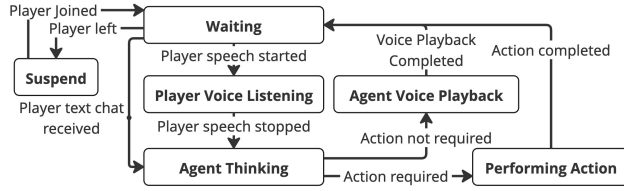


Figure 3: State transition diagram of the Navigation Pixie agent. Agent operation is managed through transitions between six primary states (suspend, waiting, player voice listening, agent thinking, agent voice playback, and performing action), with state changes governed by user input (voice and text), agent processing status, and action requirements.

- Platform Independence:** To enable deployment across multiple metaverse environment, we isolate platform-specific code in dedicated modules and establish standardized interfaces.
- UGC Compatibility:** To accommodate dynamically changing UGC, we provide platform-independent tools that allow creators to embed structured spatial information (e.g. locations and detailed descriptions) into their worlds.
- LLM Adaptability:** To adapt rapid model evolution of LLMs, we implement core functionality through rapid prompt engineering rather than custom model training.
- Operational Sustainability:** To ensure compatibility with frequently updated platforms, we define common interfaces to essential platform functionality.

3.2 Architectural Overview

As shown in Figure 2, Navigation Pixie consists of four primary components that communicate through the standardized agent driver interface. This architecture handles the complexities of LLM integration and cross-platform compatibility, allowing UGC creators to integrate navigation agents by embedding spatial metadata with less technical expertise.

Agent Driver forms the compatibility layer between Agent Core and metaverse platforms, serving as an abstraction layer that enables platform independence. This interface provides four core functions: (i) **Environmental Data Capture** retrieves and normalizes structured spatial information, room metadata, and user access information from the metaverse platform. (ii) **Event Propagation** manages real-time events such as user entry/exit notifications and text communications via publish-subscribe patterns. (iii) **Avatar Control** provides abstract operations for position control, orientation management, destination setting, and path status monitoring

while hiding platform-specific implementation details. (iv) **Communication Access** exposes standardized methods for text messaging and emote execution.

Agent Driver processes two types of structured spatial data essential for spatial understanding: (i) **Navigation Mesh** represents traversable areas for path computation, and (ii) **Navigation Data** provides structured information including spatial names, location identifiers, coordinates, and semantic descriptions for LLM context awareness.

Agent Core serves as the cognitive center of the navigation system, comprising five integrated subsystems: (i) **Agent Driver Client** communicates with the server through WebAPI, sending action commands and receiving environmental data. (ii) **Environmental Observation Subsystem** receives user events, navigation data, and room information, providing them to the LLM. (iii) **LLM** makes decisions based on environmental data and user input, issuing responses and action instructions. (iv) **Action Subsystem** generates specific action instructions based on LLM decisions. (v) **Conversational Subsystem** manages user input and agent dialog while generating responses.

Metaverse Application implements platform-specific functionality by encapsulating proprietary platform APIs within standardized agent driver interfaces. This component comprises four integrated subsystems: (i) **Agent Driver Server** processes requests from the agent driver client and handles conversion to platform-specific APIs. (ii) **Avatar Navigation** controls avatar movement to specified locations and provides navigation feasibility and remaining distance. (iii) **Communication Functionality** manages standard user communication features including text chat and emotes. (iv) **Environmental Data Provider** collects spatial information from UGC World and provides it to the Agent Driver Server.

UGC World contains spatial information created by users using the game engine or platform features, including navigation meshes that define traversable areas, navigation data with coordinates and semantic descriptions, and room instance metadata such as room names.

Virtual Audio Devices mediate voice communication between users and the agent, enabling natural conversational interaction.

4 IMPLEMENTATION

We implemented the architecture described in Sec. 3 on the commercial metaverse platform Cluster.

4.1 Component Implementation

Agent Core was implemented in Python and deployed on AWS EC2, using a two-tier LLM architecture: OpenAI GPT-4o Realtime

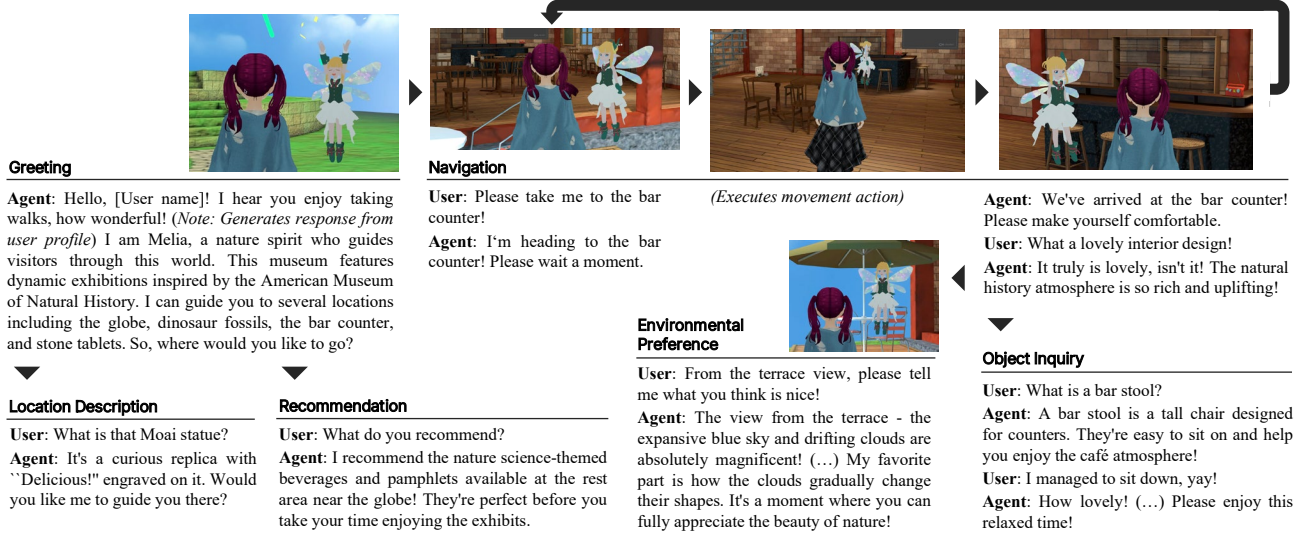


Figure 4: User experience flow and selected examples of conversations between Navigation Pixie (agent) and users. Through various dialogue patterns, the agent demonstrates capabilities such as using user profile information, performing spatial movements, providing detailed explanations about the environment, expressing personal views, and maintaining natural conversation.

(gpt-4o-realtime-preview-2024-12-17) for speech input/output processing as primary, and GPT-4o (gpt-4o-2024-11-20) for decision-making as secondary. These models integrate through OpenAI's function calling API, enabling both speech interaction and sophisticated decision processes.

Environmental Observation System formats spatial information in custom JSON structures, while Agent Driver Client communicates via REST API and WebSocket for action commands and environmental data exchange.

Metaverse Application was implemented by customizing the client source code provided with permission from Cluster, Inc., the developer of Cluster. We extended this Unity-based client with the agent driver server functionality, implementing text chat sending/receiving, avatar movement control, environmental data collection/transmission, and user event notification as Web API services.

Avatar Navigation utilizes Unity's NavMesh agent for movement control, while Communication functionality adapts Cluster's text chat and emote features with the addition of user attention-switching capabilities based on avatar state. Spatial information retrieval leverages Cluster's world loading mechanism, referencing Unity components and server data for room and user information.

Note that our implementation achieves high extensibility by encapsulating platform-specific processing within the agent driver. Porting to other metaverse platforms (e.g., VRChat) requires only implementing Agent Driver Server for the target platform. Detailed extension procedures for other platforms are provided in Supplementary Material.

4.2 Avatar Design

We selected a pixie-like avatar design to simplify agent driver implementation (Fig. 1). through gliding movement, eliminating walk/jump animation requirements. The design prioritizes approachability with intelligence signaling, visual differentiation from human avatars, high visibility, and non-verbal expression capabilities suitable for natural language interaction.

4.3 State Management and Agent Behavior Control

Figure 3 shows the state transition of the agent. Agent operation follows six state transitions: (i) **Suspend** waits for user entry, (ii) **Waiting** displays "(Please speak)" to encourage interaction,

(iii) **Player Voice Listening** detects and processes user's voice input, (iv) **Agent Thinking** processes utterances with LLMs while displaying "(Thinking)", (v) **Agent Voice Playback** plays LLM responses until completion, and (vi) **Action Execution** performs movements before returning to waiting state.

4.4 Agent Performance Evaluation

Response Time We measured the response time of the agent by measuring duration from user's voice utterance completion to response playback initiation. Based on 60 samples, the average response time was 6.40 ± 3.61 seconds, including speech-to-text transcription, response generation, and speech synthesis. This result is longer than turn-taking intervals in natural conversation (0.5-1 second) [32]. However, this latency reflects GPT-4o Realtime processing capabilities and can improve with future model evolution. To mitigate the perceived delay in the dialog, we implemented filler messages ("Thinking...") and head tilt animations [28].

Agent Conversation Examples Figure 4 demonstrates spatial guidance capabilities, environmental query responses, and the natural conversational flow of our agent through representative user interactions.

5 USER STUDY

This section details our user study of Navigation Pixie, addressing the following research questions:

RQ1 To what extent do on-demand navigation agents improve exploration behavior and dwell time in VR environments, including metaverse platforms?

RQ2 What design elements, such as interaction format, responsiveness, and navigation style, most effectively increase user interest and immersion in VR environments?

5.1 Experiment Setup

5.1.1 Experimental Design

We conducted two separate experiments to evaluate platform-specific effects: a PC client study and a VR-HMD study. Participants were randomly assigned to experimental conditions:

- (A) **On-Demand Agent:** Our Navigation Pixie implementation integrated natural language dialog with spatial navigation capabilities. Users freely conversed with the agent and requested guidance to specific destinations, with the agent responding using LLM contextual understanding and spatial structure information.
- (B) **Fixed-Route Agent:** Guided users along predetermined routes, moving to destinations upon user instructions (e.g., “OK”) and providing explanations at each location. While using the same spatially informed LLM as Group A, the route remained fixed. After completing planned exploration, users could explore freely.
- (C) **Control:** Users entered and explored the room without any agent or guidance. This condition was only included in the PC client study.

This design enables comparison of agent presence (A & B vs. C) and dialog capability (A vs. B) on user experience.

5.1.2 Experimental Environment

The experiment used two UGC worlds in Cluster (Fig. 5): **Museum** featured an exhibition gallery (36.1 m $\times$ 66.0 m) with 7 navigation points, while the world named **Ruina**⁴ offered a sky cafe (37.0 m $\times$ 51.7 m) with 5 navigation points. Both environments provided comparable explorable areas with contrasting contexts (indoor/cultural vs. outdoor/social) and structured spatial information for LLM initialization. Participants were able to communicate with agents using either text chat or voice.

5.1.3 Participants

Participants were recruited through the Cluster’s official events and social media channels. The PC client study (1000 JPY compensation) recruited 135 interested participants over 2 days, with 99 completing the four-day experimental period. The VR-HMD study provided higher compensation (2000 JPY) due to limited HMD availability, recruiting 132 participants over 5 days and yielding 94 valid datasets. Both studies applied identical exclusion criteria for no-shows, technical problems, and incomplete surveys.

Fig. 6 shows the demographics of both studies. The PC client participants (age: 36.29 ± 12.01 years, range 16-65; gender: 59 males, 36 females, 4 others) and VR-HMD participants (age: 32.6 ± 9.89 years, range 16-59; gender: 55 males, 23 females, 16 others) demonstrated similar demographic profiles. Participant distribution across conditions was: PC client study - Museum (A:18, B:17, C:17) and Ruina (A:15, B:15, C:17); VR-HMD study - Museum (A:26, B:26) and Ruina (A:21, B:21). Most participants were regular users of the metaverse platform and were familiar with virtual environments.

5.1.4 Experimental Protocol

The experiment followed three phases for both studies:

1. **Pre-participation Preparation:** Participants reviewed the experimental overview, completed a consent form on a reservation page, and reserved a time slot. Upon completion of the reservation process, participants were given URLs for randomly assigned world types and conditions.
2. **Experiment:** Participants entered worlds via distributed URLs at reserved times. Groups A and B encountered agents that provided condition-specific greetings and instructions upon entry. Although participants can leave the world at any time, all participants were asked to leave after 30 minutes via chat.

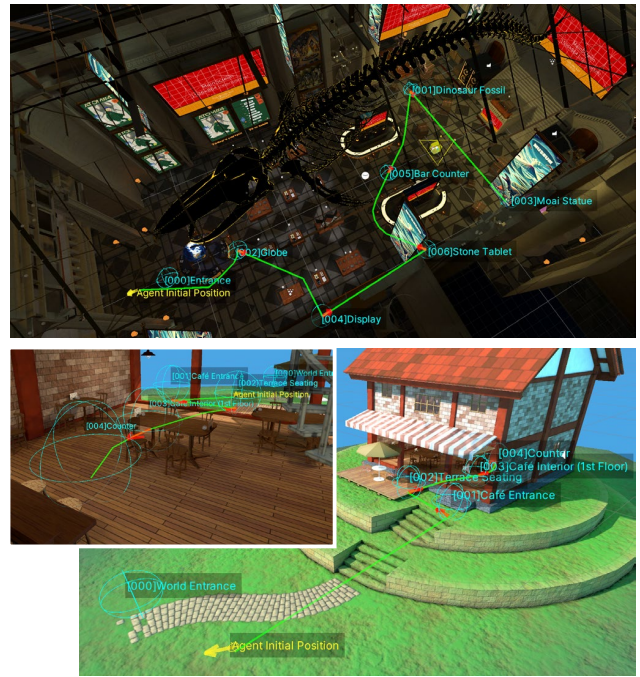


Figure 5: Two virtual environments and navigation points used in the experiment. (Top) The **Museum** environment has 7 navigation points centered around exhibits such as dinosaur fossils and globes. (Bottom) The **Ruina** sky cafe environment includes 5 navigation points across a social space with terrace seating and counters. Both environments provide structured spatial information and semantic descriptions to the LLM, enabling navigation from the agent's starting position to each point. Green lines indicate the predetermined route for group B (fixed route agent).

3. **Experiment Completion:** Upon exit, participants were redirected to a Google Form to provide demographic information and subjective ratings.

5.1.5 Evaluation Metrics

Metrics evaluated navigation effectiveness corresponding to each RQ:

RQ1 (Exploration Behavior) Metrics

- **Dwell Time:** Duration from user entry to exit.
- **Free Exploration Time:** Cumulative duration from when the agent completed movement to a destination and finished speaking until the user either requested another navigation instruction or exited. This metric was analyzed to eliminate the possibility that longer dwell times merely resulted from extended LLM agent speech duration.

RQ2 (Immersion and Agent Evaluation) Metrics

- **Immersion Rating:** Using items from engagement dimension of the Temple Presence Inventory (TPI) [22], which measures how deeply users engage and focus during media experiences. We evaluated immersion, involvement, sensory engagement, reality sensation, relaxation/stimulation, and story appeal using 7-point Likert scales (see Supplementary Material Sec. 2).
- **Agent Evaluation:** Using Godspeed Questionnaire Series (GQS) [2] to evaluate impressions of the agents across five dimensions: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety, assessed on 5-point Likert scales (see Supplementary Material Sec. 3).

⁴<https://cluster.mu/w/1aaa3e6e-4823-4339-ab51-ca37cd461902>

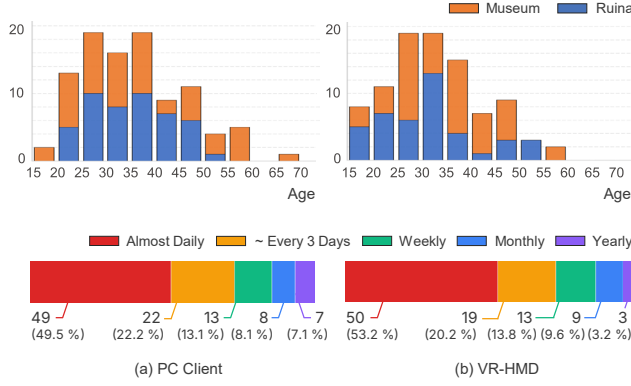


Figure 6: Participant demographics for (a) PC client study and (b) VR-HMD study. Top panels show age distribution by environment (Museum in orange, Ruina in blue). Bottom panels show frequency of logins to metaverse platforms with participant counts and percentages.

- **Free Response Answers:** Open-ended questions regarding exploration motivation (see Supplementary Material Sec. 4).

5.1.6 Analysis Methods

Statistical analyses employed appropriate tests for between-subjects comparisons with significance set at $p < .05$. We first assessed normality using Shapiro-Wilk tests and variance homogeneity using Bartlett's (three-group) or F-tests (two-group). For three-group test, when normality and variance homogeneity were confirmed, we used one-way ANOVA with Tukey HSD post-hoc tests; otherwise, Kruskal-Wallis tests with Holm-corrected Wilcoxon rank-sum post-hoc comparisons. For two-group test, we applied Student's t-test when both assumptions were met, Welch's t-test when only normality held, or Wilcoxon rank-sum test when normality was violated.

5.2 Results

5.2.1 Cross-Platform Behavioral and Perceptual Results

Dwell Time From Fig. 7 (a), both platforms demonstrated consistent agent effectiveness in extending user engagement. Group A significantly extended dwell time compared to Groups B and C in PC environments, with VR-HMD results confirming this pattern between Groups A and B.

Spatial Exploration Patterns Fig. 8 visualizes the distribution of user dwell times as heat maps in PC Client setup, revealing distinct exploration patterns: Group B concentrated along predetermined paths, Group C exhibited limited exploration in restricted areas, while Group A demonstrated extensive spatial coverage with prolonged engagement at key interest points.

To quantify spatial utilization patterns, we calculated spatial entropy of the dwell time distributions of the PC client users. Entropy was calculated using the formula $H = -\sum p_i \log p_i$, where p_i represents the proportion of dwell time spent in each spatial area. Higher entropy values indicate distributed spatial utilization, while lower values suggest concentrated exploration. Since agent speech constrains users to the same spatial region, we calculated spatial entropy based on pure movement and observation time, excluding agent speech duration. The results revealed the following entropy patterns of C (6.49) > B (5.82) > A (5.75) in Museum and C (5.57) > B (5.30) > A (4.90) in Ruina.

These results confirm that Group C exhibited dispersed, undirected exploration behavior as expected from the heatmap visualization. However, Group A recorded the lowest entropy values, despite demonstrating multiple circulation routes in the heatmaps.

This suggests that users engage in deeper observation at specific locations after talking to the on-demand agent, which indicates that conversational interaction promotes focused spatial engagement beyond simple navigation guidance.

Free Exploration Time From Fig. 7 (b), self-directed exploration behavior showed remarkable consistency across platforms. Group A participants engaged in substantially longer periods of autonomous exploration after agent guidance completion compared to Group B participants in both PC and VR-HMD conditions, with both environments demonstrating this effect across platforms.

Immersion and Agent Evaluation Fig. 9 and Fig. 10 shows the results of immersion and agent evaluations. PC results showed Group A trending toward higher immersion scores in Museum without reaching statistical significance, while Ruina's "Relaxation/Stimulation" dimension indicated that agent presence enhanced environmental stimulation value. Group A consistently scored higher on "Animacy" and "Likeability" dimensions across both PC environments.

VR-HMD results showed environment-dependent effects: Museum showed no significant differences between Groups A and B across all dimensions, suggesting exhibition contexts may constrain differential agent effects in immersive settings. Conversely, Ruina demonstrated Group A superiority across three dimensions: "Anthropomorphism," "Animacy," and "Perceived Intelligence", indicating that VR-HMD platforms amplify social perception of on-demand agents specifically within social contexts.

5.2.2 Experience Level Analysis

We analyzed participant behavior based on prior platform usage, categorizing users with less than 50 hours of total platform time as novices and others as veterans. While sample sizes in some novice subgroups preclude robust statistical testing, mean exploration times consistently favored on-demand agents across experience levels. For example, representative VR-HMD results demonstrate this pattern: in Ruina, novices achieved 10.24 ± 6.67 minutes of free exploration with Group A versus 1.44 ± 1.12 minutes with Group B ($n=6$ vs. $n=5$), while veterans showed comparable advantages (10.03 ± 9.62 vs. 1.23 ± 0.90 minutes, $n=15$ vs. $n=16$). PC client and Museum data also revealed similar directional effects. These consistent patterns suggest Navigation Pixie benefits extend regardless of prior metaverse familiarity. Detailed statistical analyses are provided in Supplementary Material.

5.2.3 Free-Text Responses

We categorized free-text responses from both PC client and VR-HMD studies and identified distinct thematic patterns for each experimental condition. Detailed comment examples are provided in Supplementary Material.

(A) **On-demand Agent** Participants consistently provided positive feedback across four primary dimensions: **social presence**, **integration of speech understanding with spatial information**, **natural dialogue capabilities**, and **personalized experience**. VR-HMD users particularly emphasized enhanced character realism and social connection compared to PC users. Technical limitations emerged consistently across platforms, with participants reporting **voice synthesis quality issues** and **spatial recognition inaccuracies**.

(B) **Fixed-route Agent** Participants acknowledged benefits of **structured guidance** and **multimodal explanations** while expressing significant dissatisfaction with **constrained exploration freedom** and **limited interactivity**. VR-HMD participants particularly emphasized the impersonal, mechanistic nature of scripted responses within immersive environments.

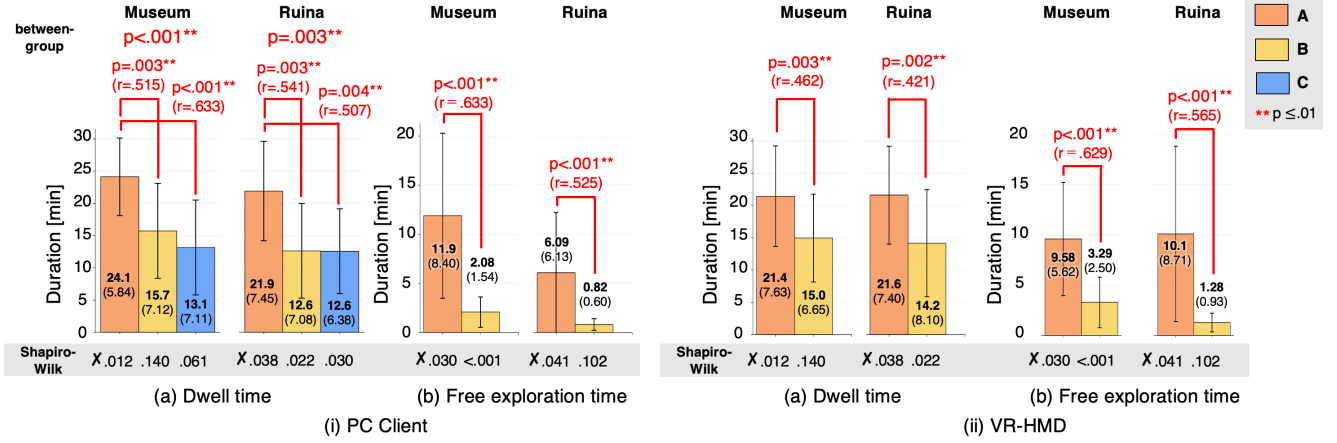


Figure 7: Behavioral metrics across experimental conditions and environments for (i) PC Client and (ii) VR-HMD studies: (a) dwell time and (b) free exploration time. Statistical significance and effect sizes shown above brackets. The legend for each statistic represents the group's name (A: On-Demand Agent, B: Fixed-Route Agent, C: Control)

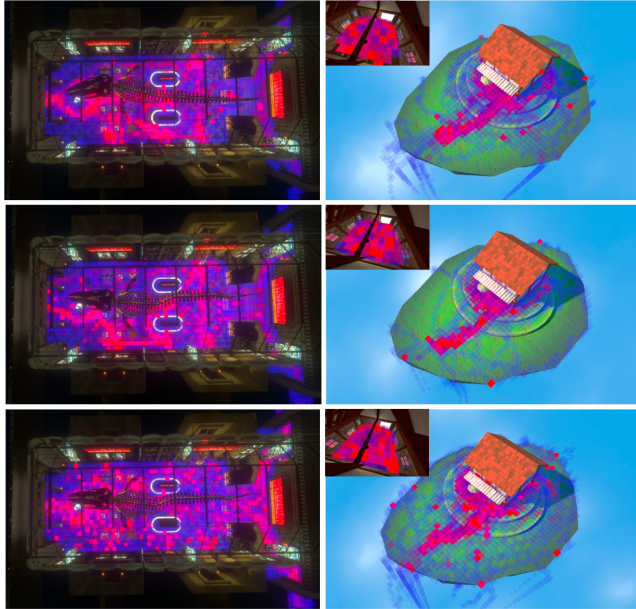


Figure 8: Heatmaps of user dwell distribution of user trajectories across experimental conditions in PC Client: (top) Group A, (middle), B, and (bottom) C, visualizing exploration patterns in (left) Museum and (right) Ruina. Red areas show long stay locations while blue areas indicate brief visits.

(C) Control PC client participants reported **spontaneous discovery** benefits but predominantly cited **absence of social interaction** and **insufficient contextual information**.

6 DISCUSSION

This study implemented and evaluated Navigation Pixie, an on-demand navigation agent, on the commercial metaverse platform Cluster, comparing it to fixed-route agent and no-agent conditions across both PC client and VR-HMD platforms. Results demonstrate that on-demand agents significantly increased dwell time and free exploration time across both platforms, though subjective evaluation patterns revealed distinct platform-specific and context-dependent effects.

This section discusses the analysis for each RQ, the differences between platforms, future improvement directions for on-demand

navigation AI agents in commercial metaverse platforms, and current limitations of our study with directions for future research.

6.1 RQ1: Enhancement of Exploration Behavior and Dwell Time

The experimental results showed that Group A had 1.5-1.7 times longer dwell times and 3-5 times longer free exploration periods compared to Group B and Group C under both PC client and VR-HMD conditions. This is consistent with previous findings that users who can actively inquire, rather than passively receive explanations, exhibit more active behavior [12].

Free-text responses and behavioral observations suggest that interactive exploration support with immediate responses increases curiosity and exploration motivation, not only in controlled VR settings, but also in UGC environments. VR-HMD participants expressed stronger emotional engagement with phrases like “I wanted to see much more” and “it felt like exploring together,” indicating heightened motivational effects in immersive contexts. In addition, the agent’s detailed environmental descriptions extend the established finding that “additional information provided by the guide increases exploration volume” [1] to UGC contexts.

6.2 RQ2: Interaction Format, Navigation Design, and User Experience

The on-demand agent demonstrated the following key advantages and challenges with distinct manifestations across PC client and VR-HMD platforms:

- 1. Flexible Natural Language Dialogue and Personalized Experience:** Incorporating user-specific information into conversations was found to be critical to improving likeability. Referencing users’ names, preferences, and past interactions is known to increase trust and attachment [5]. The on-demand agent changed perceptions from a mere guide to a “virtual partner” by demonstrating interest and empathy toward users.
- 2. Integration of Spatial Information with Detailed Explanations:** Group A users appreciated the high resolution of the spatial responses, indicating effective integration between the LLM’s flexible responses and pre-configured navigation data. However, spatial recognition errors were also noted, highlighting the need for improved accuracy.
- 3. Immersion Reduction Due to Technical Limitations:** Free responses indicated that unnatural speech, slow response and hallucinations in GPT-4o real-time speech responses increased

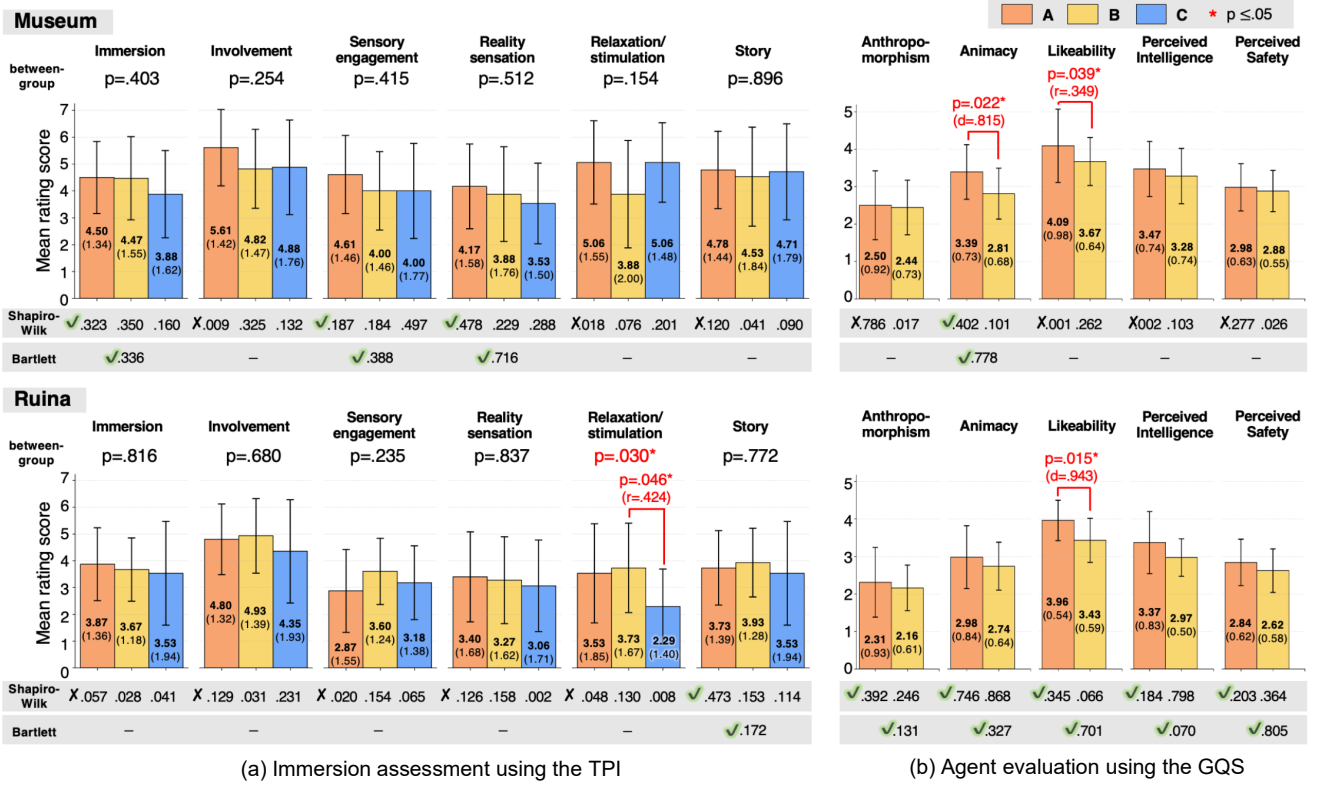


Figure 9: Comparison of immersion experience and agent ratings in Museum (top) and Ruina environments (bottom) in PC client study. Bar graphs show means with standard deviations in parentheses and corresponding error bars. Statistical test selection was based on Shapiro-Wilk normality tests and Bartlett's tests for variance homogeneity, with results displayed at the bottom of each panel. Significant differences between groups are indicated by lines with red asterisks (*) based on $p \leq .05$. (a) Immersion assessment using the Temple Presence Inventory (TPI), with p-values below each dimension representing between-group difference test results. (b) Agent evaluation using the Godspeed Questionnaire Series (GQS).

distrust of the agent. Previous research suggests that the naturalness of speech processing directly affects agent trustworthiness [3, 23], which may explain the lack of significant differences in the GQS dimensions of anthropomorphism, intelligence, and perceived safety.

6.3 Embodiment Potential for On-Demand Navigation Agents

We observed Group A users treating agents as human-like entities - saying "see you later" at the end of the experiment, or following agent accounts in anticipation of future encounters. These behaviors confirm that agent embodiment in VR spaces enhances social presence. This phenomenon is consistent with the "Computers Are Social Actors" paradigm [25], where minimal social cues (voice or humanoid avatar) lead people to perceive agents as "present companions" and apply social rules to interactions [21].

Eliciting interactions similar to human-to-human exchanges requires a coherent integration of natural voice, facial expressions, and gestures that further enhance trust and familiarity [3]. Previous research has suggested that NPCs can create a social atmosphere even in uninhabited spaces [6], and our study clearly demonstrates that agents can provide users with "community-like warmth" beyond mere automated guidance.

6.4 Platform-Specific Agent Perception and Environmental Context Dependencies

PC environments showed consistent directional preferences for on-demand agents, regardless of the context. In contrast, VR-HMD environments exhibited strong environmental dependency. Exhibition

contexts (museum) showed no differential agent effects, whereas social contexts (ruins) demonstrated the superiority of on-demand agents across the dimensions of anthropomorphism, animacy, and intelligence.

We attribute this phenomenon to context-specific evaluation mechanisms. In museum environments, users primarily evaluate agents as information delivery tools, and immersion in the experience reinforces functional assessment criteria. In contrast, VR-HMD immersion in Ruina environments enhances co-presence perception. This also occurs when fixed-route agents function appropriately in exhibition contexts but generate dissatisfaction in social spaces, where conversational flexibility is paramount. These platform-specific differences suggest that immersive devices function as cognitive amplifiers for agent presence perception.

6.5 Approaches for Improving Agent Performance

Several technical approaches could further improve the effectiveness of on-demand agents:

1. **Dynamic Navigation Data Generation and Updates:** Integrating algorithms that automatically update NavMesh during exploration would allow flexible adaptation to newly added elements in changing UGC environments [38].
2. **Personalization Through Long-term Memory Modules:** Building on relational agent research [5], accumulating and referencing user-specific conversation histories would allow more intimate communication based on previous visit experiences.
3. **Integration of Multimodal Recognition and Non-verbal Expression:** Combining reinforcement learning-based agent [40]

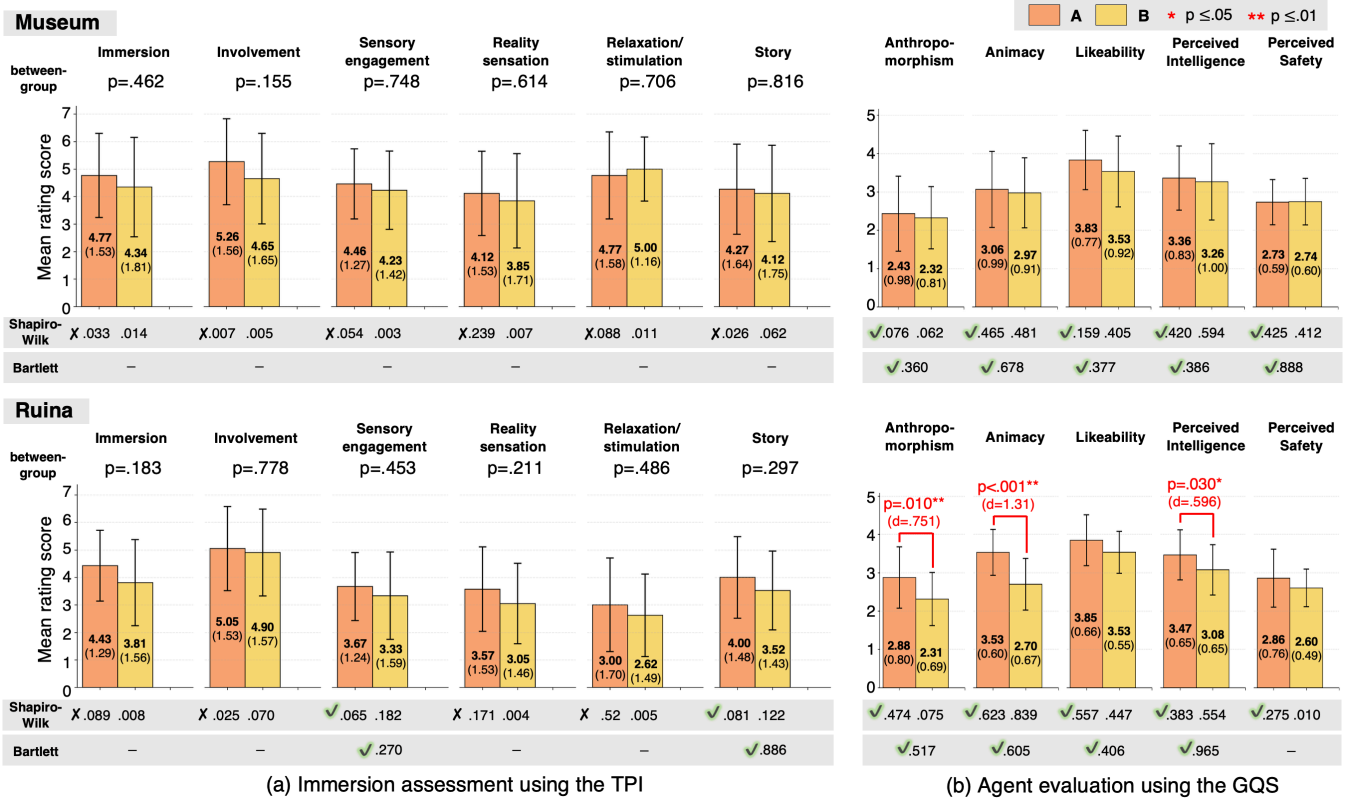


Figure 10: Immersion experience and agent ratings across Museum and Ruina environments in VR-HMD study. Data presentation format and statistical analysis procedures follow those described in Fig. 9. (a) TPI immersion assessment. (b) GQS agent evaluation.

with visual and sensor data integration into LLMs could enable finer-grained guidance through reduced object misidentification rates and real-time explanation of items users are focusing on [5].

6.6 Limitations and Future Directions

Our research has several limitations that suggest directions for future work:

Temporal Constraints and Long-term Effects This experiment measured effects within single sessions (approximately 30 minutes), leaving long-term behavioral changes and learning effects unexamined [4]. The observed “following” behavior toward agents suggests the need for longitudinal studies that examine ongoing relationship formation and its effects.

Participant Characteristics and Device Constraints Our recruitment approach, which drew participants from the platform’s existing user base, resulted in predominantly experienced users (> 70% weekly users). Preliminary behavioral metrics revealed directionally consistent agent effectiveness across experience levels, but limited novice sample sizes preclude definitive statistical conclusions. Furthermore, our evaluation excluded smartphone-based interactions. Mobile platforms introduce distinct constraints through touch-based navigation and interface limitations that may fundamentally alter agent effectiveness patterns, which represent an important future investigation.

Influence of Diverse Design Elements While this study used a fairy-like avatar for the agent, the effect of different avatar designs (humanoid, animal, etc.) on user responses remains unexplored [44]. Given the heightened anthropomorphic attribution observed in VR-HMD environments, systematic investigation of

avatar design effects across platforms represents a particularly important research direction. Systematic comparative studies of avatar designs would help to establish optimal design guidelines for specific purposes.

Extending to Multi-Agent, Multi-User Environments Real-world metaverse use involves the coexistence of multiple agents and users, necessitating research into navigation support in such environments. Agent designs that serve as social catalysts facilitating user interactions and supporting collaborative exploration are important directions for development.

User Co-Creation in Agent Development Approaches where “User A creates navigation data through dialogue that User B experiences” demonstrate the potential of co-creative development using UGC. Investigating how such co-creation processes affect agent attachment and user satisfaction is an interesting research topic.

7 CONCLUSION

This paper presents Navigation Pixie, an LLM-based conversational navigation agent for commercial metaverse platforms. Our experiments in Cluster show that on-demand navigation agents significantly extend user engagement, increasing dwell time by 1.5-1.7 times and free exploration time by 3-5 times compared to fixed-route agents or no-guidance conditions.

Our key contributions include a loosely coupled architecture that separates platform-dependent components for improved portability, effective integration of LLM capabilities with spatial information to enable contextual guidance in UGC spaces, and a large-scale study demonstrating the effectiveness of on-demand agents in different virtual environments. We found that the social presence of agents consistently increased exploration motivation; users naturally developed human-like interactions with embodied agents, consistent

with the finding that minimal social cues trigger companion-like perceptions.

Our Navigation Pixie establishes a foundation for on-demand guidance frameworks that leverage the diversity of UGC. We anticipate that this work will establish new directions for metaverse agent design and advance human-centered interaction research in VR environments.

ACKNOWLEDGMENTS

This work was partially supported by JST Moonshot Research & Development Program Grant Number JPMJMS2013 and JST ASPIRE Grant Number JPMJAP2327.

We would like to express our sincere gratitude to the creators who provided virtual worlds for our experiment: Taki, the creator of Ruina: A Small Cafe In The Sky, and Zawa, the creator of Museum.

REFERENCES

- [1] V. Barneche-Naya and L. Hernandez-Ibañez. Assisted navigation for digital architectural walkthroughs in natural user interface-based installations. *Universal Access in the Information Society*, 23, 08 2023. doi: 10.1007/s10209-023-01034-6 7
- [2] C. Bartneck. Godspeed questionnaire series: Translations and usage. In *International handbook of behavioral health assessment*, pp. 1–35. Springer, 2023. 5
- [3] D. Becker, L. Braach, L. Clasmeier, T. Kaufmann, O. Ong, K. Ahrens, C. Gäde, E. Strahl, D. Fu, and S. Wermter. Influence of robots' voice naturalness on trust and compliance. *J. Hum.-Robot Interact.*, 14(2), Jan. 2025. doi: 10.1145/3706066 8
- [4] T. Bickmore. Relational agents for chronic disease self-management. In *Health Informatics: A Patient-Centered Approach to Diabetes*. The MIT Press, 08 2010. doi: 10.7551/mitpress/8336.003.0011 9
- [5] T. W. Bickmore and R. W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.*, 12(2):293–327, June 2005. doi: 10.1145/1067860.1067867 2, 7, 8, 9
- [6] A. Bönsch, J. Ehret, D. Rupp, and T. W. Kuhlen. Wayfinding in immersive virtual environments as social activity supported by virtual agents. *Frontiers in Virtual Reality*, 4, 2024. doi: 10.3389/frvir.2023.1334795 2, 8
- [7] A. Bönsch, D. Hashem, J. Ehret, and T. W. Kuhlen. Being guided or having exploratory freedom: User preferences of a virtual agent's behavior in a museum. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, IVA '21*, p. 33–40. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472306.3478339 2
- [8] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1):3–55, 1999. doi: 10.1016/S0004-3702(99)00070-3 2
- [9] E. Cesanek, S. Shivkumar, and D. Wolpert. Ouvrai opens access to remote virtual reality studies of human behavioural neuroscience. *Nature Human Behaviour*, 8:1–16, 04 2024. doi: 10.1038/s41562-024-01834-7 2
- [10] L. Chittaro, L. Ieronutti, and R. Ranon. Navigating 3d virtual environments by following embodied agents: a proposal and its informal evaluation on a virtual museum application. *PsychNology Journal*, 2:24–42, 01 2004. 1, 2
- [11] J. Collins, C. Jung, Y. Jang, D. Montour, A. S. Won, and S. Azenkot. "the guide has your back": Exploring how sighted guides can enhance accessibility in social virtual reality for blind and low vision people. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '23*. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3597638.3608386 2
- [12] S. Freitag, B. Weyers, and T. W. Kuhlen. Interactive exploration assistance for immersive virtual environments based on object visibility and viewpoint quality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 355–362, 2018. doi: 10.1109/VR.2018.8447553 7
- [13] S. J. Friston, B. J. Congdon, D. Swapp, L. Izzouzi, K. Brandstätter, D. Archer, O. Olkkonen, F. J. Thiel, and A. Steed. Ubiq: A system to build flexible social virtual reality experiences. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology, VRST '21*, Dec. 2021. 2
- [14] D. Giunchi, N. Numan, E. Gatti, and A. Steed. DreamCodeVR: Towards democratizing behavior design in virtual reality with Speech-Driven programming. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 579–589. IEEE, Mar. 2024. 2
- [15] J. Gu, E. Stefani, Q. Wu, J. Thomason, and X. Wang. Vision-and-language navigation: A survey of tasks, methods, and future directions. In S. Muresan, P. Nakov, and A. Villavicencio, eds., *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 7606–7623. Association for Computational Linguistics, Dublin, Ireland, May 2022. doi: 10.18653/v1/2022.acl-long.524 2
- [16] R. Hammady, M. Ma, C. Strathearn, and M. Mohamad. Design and development of a spatial mixed reality touring guide to the egyptian museum. *Multimedia Tools and Applications*, 79, 02 2020. doi: 10.1007/s11042-019-08026-w 2
- [17] D. Jan, A. Roque, A. Leuski, J. Morie, and D. Traum. A virtual tour guide for virtual worlds. In *Intelligent Virtual Agents*, Lecture notes in computer science, pp. 372–378. Springer Berlin Heidelberg, 2009. doi: 10.1007/978-3-642-04380-2.40 1, 2
- [18] R. Kadobayashi and K. Mase. Seamless guidance by personal agent in virtual space based on user interaction in real world. In *Proceedings of the third International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology (PAAM '98)*, pp. 191–200, Mar. 1998. 2
- [19] R. Kurai, T. Hiraki, Y. Hiroi, Y. Hirao, M. Perusquia-Hernandez, H. Uchiyama, and K. Kiyokawa. An implementation of MagicCraft: Generating interactive 3D objects and their behaviors from text for commercial metaverse platforms. In *Abstracts and Workshops of the 32nd IEEE Conference on Virtual Reality and 3D User Interfaces*, pp. 1–2, 2025. 2
- [20] R. Kurai, T. Hiraki, Y. Hiroi, Y. Hirao, M. Perusquia-Hernández, H. Uchiyama, and K. Kiyokawa. Magicitem: Dynamic behavior design of virtual objects with large language models in a commercial metaverse platform. *IEEE Access*, 13:19132–19143, 2025. doi: 10.1109/ACCESS.2025.3530439 2
- [21] C. Kyrilitsias and D. Michael-Grigoriou. Social interaction with agents and avatars in immersive virtual environments: A survey. *Frontiers in Virtual Reality*, Volume 2 - 2021, 2022. doi: 10.3389/frvir.2021.786665 2, 8
- [22] M. Lombard, T. B. Ditton, and L. Weinstein. Measuring presence: the temple presence inventory. In *Proceedings of the 12th annual international workshop on presence*, pp. 1–15. University College London, 2009. 5
- [23] C. Maltezou-Papastilianou, R. Scherer, and S. Paulmann. How do voice acoustics affect the perceived trustworthiness of a speaker? a systematic review. *Frontiers in Psychology*, Volume 16 - 2025, 2025. doi: 10.3389/fpsyg.2025.1495456 8
- [24] L. Marucci and F. Paternò. Design and evaluation of an adaptive virtual guide for web applications. *Universal Access in the Information Society*, 1:163–176, 06 2002. doi: 10.1007/s102090100016 2
- [25] C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103, 2000. doi: 10.1111/0022-4537.00153 2, 8
- [26] H. Naveed, A. U. Khan, S. Qiu, et al. A comprehensive overview of large language models. arXiv preprint arXiv:2307.06435, 2023. Accessed from <https://arxiv.org>. 2
- [27] N. Numan, D. Giunchi, B. Congdon, and A. Steed. Ubiq-genie: Leveraging external frameworks for enhanced social vr experiences. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 497–501, 2023. doi: 10.1109/VRW58643.2023.00108 2
- [28] N. Ohshima, K. Kimijima, J. Yamato, and N. Mukawa. A conversational robot with vocal and bodily fillers for recovering from awkward

- silence at turn-takings. *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 325–330, 08 2015. doi: 10.1109/roman.2015.7333677 4
- [29] M. Prpa and M. Pan. Ellma-t: Llm-powered english language learning in social vr. In *Proceedings of CHI 2023*, 2023. 2
- [30] R. Radiah, V. Makela, S. Prange, S. D. Rodriguez, R. Piening, Y. Zhou, K. Kohle, K. Pfeuffer, Y. Abdelrahman, M. Hoppe, A. Schmidt, and F. Alt. Remote vr studies: A framework for running virtual reality studies remotely via participant-owned hmds. *ACM Trans. Comput.-Hum. Interact.*, 28(6), Nov. 2021. doi: 10.1145/3472617 2
- [31] G. D. Ritterbusch and M. R. Teichmann. Defining the metaverse: A systematic literature review. *IEEE Access*, 11:12368–12377, 2023. 1
- [32] H. Sacks, E. Schegloff, and G. Jefferson. A simple systematic for the organisation of turn taking in conversation. *Language*, 50:696–735, 12 1974. doi: 10.2307/412243 4
- [33] D. Saffo, S. Di Bartolomeo, L. Panavas, C. Yildirim, and C. Dunne. Two paths towards the future of remote studies using social VR. In *Proc. 1st XR Remote Research Workshop @ CHI 2021*, XR-CHI, 2021. Preprint & supplemental material: <https://osf.io/4we6g>. doi: 10.31219/osf.io/4we6g 2
- [34] D. Saffo, S. Di Bartolomeo, C. Yildirim, and C. Dunne. Crowdsourcing virtual reality experiments using VRChat. In *Proc. CHI Conference on Human Factors in Computing Systems Late-Breaking Work*, CHI LBW, pp. 1–8, 2020. Preprint & supplemental material: <https://osf.io/w569f>. doi: 10.1145/3334480.3382829 2
- [35] A. Steed, L. Izzouzi, K. Brandstätter, S. Friston, B. Congdon, O. Olkkonen, D. Giunchi, N. Numan, and D. Swapp. Ubiqu-exp: A toolkit to build and run remote and distributed mixed reality experiments. *Frontiers in Virtual Reality*, 3, 2022. doi: 10.3389/frvir.2022.912078 2
- [36] N. Techasartikul, P. Ratsamee, J. Orlosky, T. Mashita, Y. Uranishi, K. Kiyokawa, and H. Takemura. Evaluation of Embodied Agent Positioning and Moving Interfaces for an AR Virtual Guide. In Y. Kakehi and A. Hiyama, eds., *ICAT-EGVE 2019 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2019. doi: 10.2312/egve.20191276 2
- [37] P. Trahanias, A. Argyros, D. Tsakiris, A. Cremers, D. Schulz, W. Burgard, D. Haehnel, V. Savvaides, G. Giannoulis, and M. Coliou. Tourbot-interactive museum tele-presence through robotic avatars. *Proc. of the 9th International World Wide Web Conference*, 01 2000. 2
- [38] W. G. van Toll, A. F. Cook, and R. Geraerts. A navigation mesh for dynamic environments. *Comput. Animat. Virtual Worlds*, 23(6):535–546, Nov. 2012. doi: 10.1002/cav.1468 8
- [39] H. Wan, J. Zhang, A. A. Suria, B. Yao, D. Wang, Y. Coady, and M. Prpa. Building llm-based ai agents in social virtual reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613905.3651026 2
- [40] G. Wang, Y. Xie, Y. Jiang, A. Mandlekar, C. Xiao, Y. Zhu, L. Fan, and A. Anandkumar. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv: Arxiv-2305.16291*, 2023. 8
- [41] H. Wang. Enhancing art museum experience with a chatbot tour guide, 2024. 2
- [42] Z. Wang, L.-P. Yuan, L. Wang, B. Jiang, and W. Zeng. Virtuwander: Enhancing multi-modal interaction for virtual tour guidance through large language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642235 2
- [43] Z.-M. Ye, J.-L. Chen, M. Wang, and Y.-L. Yang. Paval: Position-aware virtual agent locomotion for assisted virtual reality navigation. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 239–247, 2021. doi: 10.1109/ISMAR52148.2021.00039 2
- [44] N. Yee and J. Bailenson. The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, 33(3):271–290, 2007. doi: 10.1111/j.1468-2958.2007.00299.x 9
- [45] Y. Zhang and W. Hu. Ai-driven creativity unleashed: Exploring the synergistic effects of ugc and aigc in metaverse gaming from a user perspective. In *Proceedings of the 29th International ACM Conference on 3D Web Technology*, Web3D '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3665318.3678230 2
- [46] D. Zielasko and T. Weissker. Stay vigilant: The threat of a replication crisis in vr locomotion research. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*, VRST '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3611659.3615697 2