Large-scale, Longitudinal Field Study of AI-agent-User Interactions in Commercial Metaverse

Yuichi Hiroi
Cluster Metaverse Lab
Tokyo, Japan
y.hiroi@cluster.mu

Yuki Imai
Cluster, Inc.
Tokyo, Japan
y.imai@cluster.mu

Hikari Yanagawa
Cluster Metaverse Lab
Tokyo, Japan
h.yanagawa@cluster.mu

Takefumi Hiraki
University of Tsukuba
Ibaraki, Japan
Cluster Metaverse Lab
Tokyo, Japan
hiraki@slis.tsukuba.ac.jp

***Abstract*—Although commercial metaverse platforms have advanced social implementation, they struggle to retain users. Most new users abandon these platforms shortly after their initial experience, which hinders the realization of the metaverse's social and economic potential. Although AI-agents could promote user interaction and retention by acting as social catalysts, existing research has focused on laboratory-based short-term validation. This approach provides limited evidence of the long-term effectiveness of agents in commercial environments.**

This study examines the impact of Large Language Model (LLM)-based AI-agents on user continuation behavior by operating AI-agents for 31 days on the commercial metaverse platform Cluster and observing the natural usage behavior of 5,020 unique users. The analysis used two complementary approaches: (1) an aggregate effect analysis at the weekly habit formation level and (2) an within-user effect analysis at the daily decision-making level with individual difference controls. The results revealed that interaction with the AI-agent produces lasting changes in user behavior, especially among new users. Furthermore, the cumulative relationship-building process through continuous contact opportunities rather than single impressive experiences with AI-agents was found to decisively affect users' continued metaverse usage.

***Index Terms*—Metaverse, AI agents, User retention, Field study**

I. Introduction

The metaverse is defined as living experiences within computer-generated virtual worlds, representing a new form of online community where users experience social presence through avatars [1]. Commercial platforms such as VRChat[1], Resonite[2], and Cluster[3] have enabled social implementation in entertainment, education, and business domains [2]. Since the COVID-19 pandemic, these platforms have gained attention as an infrastructure for social activity that transcends physical constraints, and the metaverse market is predicted to experience rapid growth over the next decade [3], [4].

However, many metaverse platforms struggle to retain users. Most new users abandon the platforms shortly after their initial experience, and only a small percentage of users continue to use them. This low retention rate is the main barrier to achieving the social and economic potential of metaverse platforms [5]. The introduction of AI-agents in metaverse environments is expected to support the adaptation process to metaverse platforms by providing psychological safety, particularly for novice users, as well as exploration support through spatial guidance and continuous relationship building. These agents have the potential to function as social catalysts in metaverse spaces by facilitating interactions between users [6].

The development of large language models (LLMs) has allowed AI-agents to be created with real-time contextual understanding and natural dialogue capabilities [7]–[9]. However, most existing research on the implementation of AI-agents in VR and metaverse has been limited to short-term validation in laboratory settings or specific systems, with no long-term effect validation in actual commercial environments. From an ecological validity perspective, it remains unexplored whether the findings obtained in controlled experimental environments can be reproduced in real service environments with diverse user populations and complex social dynamics.

We address these challenges by conducting an empirical study to track Large Language Model (LLM)-based AI-agent-user interactions in metaverse spaces on a large scale and for extended periods. We observe the natural usage behavior of over 5,000 unique users for 31 days on the commercial metaverse platform Cluster, examining the impact of AI-agents on continued usage behavior. Our analysis employs two complementary approaches to capture effects at different temporal scales and decision-making levels: an aggregate weekly analysis of habit formation and a causal analysis controlling for individual differences at the daily decision-making level. This approach reveals the empirical effects of AI-agent introduction in metaverse platforms and provides the theoretical foundations and practical guidelines for future AI-agent designs.

The contributions of this research are: (1) conducting the first large-scale field study validating AI-agent effects through 31-day observation of 5,020 users in commercial metaverse

[1]https://hello.vrchat.com/
[2]https://store.steampowered.com/app/2519830/Resonite/
[3]https://cluster.mu/en

environments, (2) demonstrating causal mechanisms through two-tier analysis at habit formation and decision-making levels, and (3) establishing guidelines for AI-agent design in metaverse platforms from the validation.

II. Related Work

A. Social Agents in Virtual Environments

Research on social interactions between humans and computer agents is based on the Computers as Social Actors (CASA) paradigm [10], which showed that humans unconsciously apply social rules to computers and exhibit interpersonal-like responses.

Virtual navigation agents evolved from physical museum tour robots in the early 2000s [11], [12] to implementations on Second Life, the early metaverse platform [13]. Research on social interactions with VR dialogue agents demonstrated that agents can control attention and temporal information like humans through nonverbal communication capabilities, such as gaze direction and pointing gestures [14], [15]. Regarding the building of long-term relationships between humans and agents, Bickmore and Picard showed that relationship-building agents can establish trust over time, promoting behavioral change [6], [16].

While traditional virtual agents primarily use fixed route approaches [14], [17], [18], Bönsch et al. demonstrated that on-demand exploration agents facilitate social interactions [19]. Recent evolution of LLMs has enabled significant advances in VR agent functionality for on-demand exploration [7]. The integration of these agents with commercial platforms has also advanced. Recent implementations on commercial platforms include memory-integrated agents on VRChat [8] and LLM-based guidance agents on Cluster that increase dwell time [9].

Previous studies demonstrated technical implementation and short-term effects of LLM agents in VR, but none validated long-term behavioral effects in actual service environments. This study demonstrates sustained effects through cumulative relationship building over 31 days of continuous observation.

B. User Engagement in Metaverse and VR

Park and Kim [5] comprehensive review of metaverse taxonomy, components, and challenges, identifies user retention as a major challenge. Even in early virtual platforms on-line, such as Massively Multiplayer Online (MMO) RPGs, many users connect simultaneously but predominantly engage in solo play, suggesting a gap between superficial co-presence and substantial interaction relationship [20].

Seay et al. [21] demonstrated that participation itself enhances continued use in online game communities, emphasizing the importance of social participation. In quantitative modeling of continued use, Demediuk et al. [22] proposed a method for predicting when players leave the multiplayer online esports game League of Legends using mixed-effects regression analysis. More recently, Kang et al. [23] analyzed the relationship between match experiences and departure in multiplayer competitive games using logs of 6 million matches over 42 days.

Recent research has conducted community analyzes based on two weeks of user behavior logs on metaverse platforms, revealing the formation of small-scale, high-cohesion community and the presence of community mediators [24]. Although this research deepened our understanding of social dynamics in metaverses, no large-scale validation has been performed of the causal effects of AI-agents on the continued use of metaverses. This study addresses this important research gap by observing 5,020 participants for 31 days.

C. Onboarding and Social Support with AI-Agents

New-user onboarding is a critical factor in determining the sustainability of online communities. Butler [25] presented a resource model for community sustainability, theorizing that balancing the acquisition and retention of new members is important. Lampe and Johnston [26] demonstrated that the responses of existing members encourage new participants to remain involved. Viégas and Smith [27] showed that activity visualization promotes continued participation and provides design guidelines for feedback mechanisms.

Regarding the implementation of social support in virtual environments, Baker et al. [28] demonstrated in elderly users of Social VR that deep social connections motivate revisits and promote health awareness. Research on virtual influencers, fictional characters created through CG technology that act as influencers in SNS [29], [30], has clarified the conditions for establishing pseudo-interpersonal relationships and demonstrated the possibility of forming emotional bonds with virtual agents.

Recently, metaverse platforms have gained attention as crowd-sourced VR experimental environments and ecological social experimental venues due to their diverse and continuously connected user bases. Ouvrai [31] and Ubiq-Exp [32] have developed remote and distributed VR experimental infrastructures that enable large-scale participant recruitment, although they have not yet achieved user bases comparable to commercial systems in terms of diversity and scale. There have been attempts to demonstrate the potential of crowd-sourced VR experiments by collecting user experience data and behavioral logs through commercial metaverse platforms, such as VRChat and Cluster [33]–[35].

Although previous studies have demonstrated the importance of onboarding and social support, providing the foundation for experimental methodology, no empirical studies have demonstrated the differential effects of AI-agents on onboarding of new users. This study clarifies the effectiveness of the social scaffolding function and the continuous relationships of AI-agents with new users.

III. Research Hypotheses

This study establishes four hypotheses to examine the impact of LLM-based AI-agents on continued usage behavior on metaverse platforms. The hypotheses are based on differences in temporal scale and decision-making level:

H1: Habit Formation. After contacting an AI-agent, users increase their weekly usage of metaverse platforms.

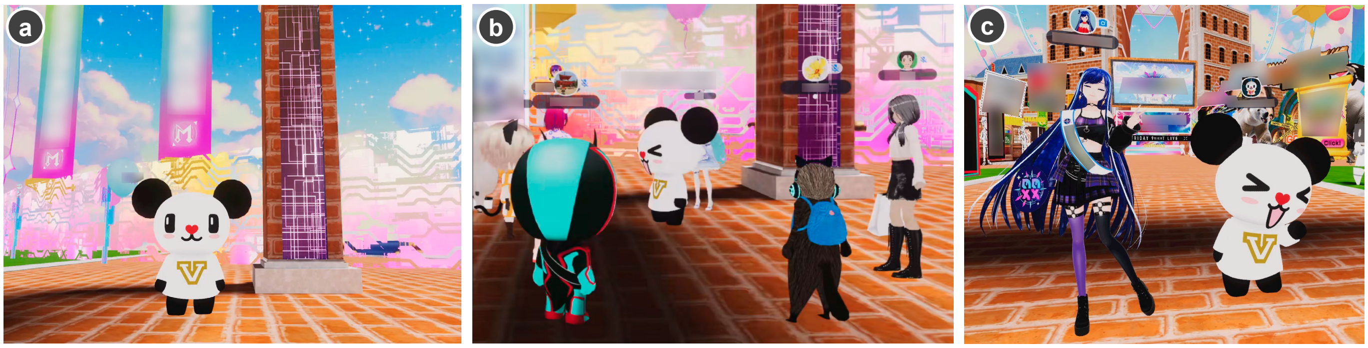

Fig. 1. Visual overview of the event space, the appearance of the AI agent, and the interactions between the agent and the user. (a) The AI agent avatar implemented on the Navigation Pixie architecture [9]. (b) AI-agent engaging in natural dialogue with users during exploration. The agent employs multimodal communication by orienting toward the most recent speaker, generating contextual responses through text chat and synthesized speech, and expressing engagement through facial expressions and gestural emotes. (c) Coordinated photography interaction demonstrating social responsiveness. When a user asks to take a photo together, the AI agent faces in the same direction as the user who made the request and performs cooperative poses, illustrating the system's capacity for joint attention and non-verbal social coordination. (Character Copyrights by TV Asahi Corp.)

H2: On-boarding. The impact of contacting AI-agents is greater for first-time metaverse users than for existing users.

H3: Immediate Continuation. Even when individual differences are controlled, the probability of logging on to the platform the next day increases after contacting AI-agents.

H4: Cumulative Relationship Building. Increases in cumulative contact days with AI-agents improve the probability of continuing to use metaverse platforms the next day.

H1 and H2 target the macroscopic and persistent effects of contact with an AI-agent on user habits. These capture stable behavioral changes, which are observed as modifications to the weekly average usage pattern. In contrast, H3 and H4 target the microscopic, immediate effects of AI-agents on daily continuation and departure decisions. They examine the impact on the specific decision-making process of "continued usage tomorrow" after individual usage sessions.

IV. Study Design

This study is an observational investigation of the effects of AI-agents on the commercial metaverse platform. Rather than conducting a controlled experiment with randomly assigned participants, we designed a large-scale field study focusing on natural behavioral patterns during actual service use. This approach enables us to assess the impact of incorporating AI-agents in practical settings while maintaining ecological validity.

This study is the first to observe user behavior on a metaverse platform during normal service operation without experimental intervention. The ethical framework for this emerging field is still developing. We will detail the IRB application, ethical compliances, and considerations for this study in Sec. VIII.

A. Study Environment

This study was conducted on the commercial metaverse platform Cluster. The platform allows users to create and share diverse virtual spaces centered on user-generated content (UGC) and supports access from multiple devices, including PCs, smartphones, and VR-HMDs. The platform provides real-time voice and text communication functions, as well as body expression (emote) functions through avatars. The platform supports free movement and social interaction within spaces.

B. Target Event Overview

This study examines user behavior in "MetaMeta Daisakusen 2025," a large-scale virtual space hosted by TV Asahi Corp., a major Japanese commercial broadcaster, on Cluster. The event took place from July 18 to August 17, 2025, for 31 days. With the theme "Summer Where Play Becomes Learning," the event was planned as a virtual summer festival targeting families, including those new to the metaverse, providing multifaceted entertainment and educational content.

The target space functioned as a central hub world within the event, designed to serve as the primary transit point through which users navigated to various content areas. AI-agents were placed on the main pathways that users traversed when moving between different attractions and content zones, maximizing natural encounter opportunities. During the investigation period, the metaverse space was accessible 24 hours a day, with AI-agents operating continuously, except during irregular maintenance periods. Both the event announcement page and the description on the platform clearly stated that AI agents would be deployed at the event and that these agents were powered by AI technology.

C. AI-agent System

In this study, we implemented "AI Agent Go-chan" as the AI agent system. "Go-chan" is the mascot character of TV Asahi Corp. This agent represents an evolution of the existing "AI Go-chan," originally created by TV Asahi Corp, into a digital space AI agent, developed using Cluster Inc.'s AI Agent Flex API[4], based on the Navigation Pixie architecture [9].

[4] https://www.biz.cluster.mu/ai-agent-flex

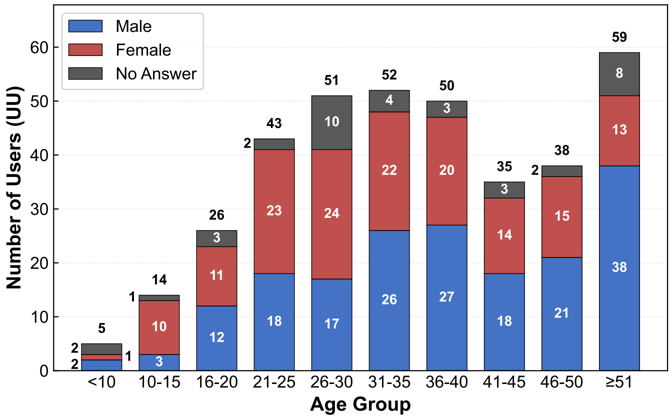

Fig. 2. Age and gender distribution of survey respondents (n=372, 7.4% response rate). Stacked bars show counts for male (blue), female (red), and no answer/other (gray) categories.

The system used the Gemini 2.5 Flash API controlled via system prompts, without fine-tuning or RAG (Retrieval-Augmented Generation). The design principles included the following: (1) using a friendly tone with the character's age-appropriate, plain language; (2) providing concise responses averaging 150 characters; (3) asking questions proactively to facilitate dialog; and (4) offering spatial guidance based on event understanding.

Functionally, the agent supported natural language dialog through text and voice, oriented itself toward speakers, performed emotes that match the emotional context (Fig. 1, b), and executed cooperative poses. When asked to take a photo together, they performed cooperative poses, enhancing the sense of social presence (Fig. 1, c).

To ensure safety, we applied all Gemini API safety filters (e.g., harassment, hate speech) at maximum strength, using `BLOCK_LOW_AND_ABOVE` option. We also deployed a secondary monitoring LLM to detect prompt injection attempts and implemented a mechanism to reset conversation memory every 2 hours. In addition, we conduct occasional manual verification of the inputs. Upon detection of inappropriate attempts, account suspension, and other penalties are applied.

D. Definition of "Contact" with the AI-agent

For the purpose of log extraction, "contact" between AI-agents and users in this study was defined as a bidirectional interaction in which users addressed the AI-agents via comments or voice input and the AI-agents recognized and responded to the input.

Currently, simple spatial proximity or passive listening to AI-agent utterances is not considered contact. However, a future study will compare passive listeners with active users.

E. Participants

1) Basic Participant Attributes: A total of 5,020 unique users entered the event space during the investigation period. Of these users, 328 (6.5%) contacted AI-agents.

Gender and age data were collected through optional online forms displayed when leaving the world. Valid responses were obtained from 372 users, which yielded a response rate of 7.4%. Of these, 58 reported having contact with an AI-agent. Figure 2 shows the age and gender distributions of the respondents. The age distribution peaked at 26–35 years old (103 users; 27.7%), and the gender composition was 182 males (48.9%), 153 females (41.1%), and 38 other or without response (10.2%).

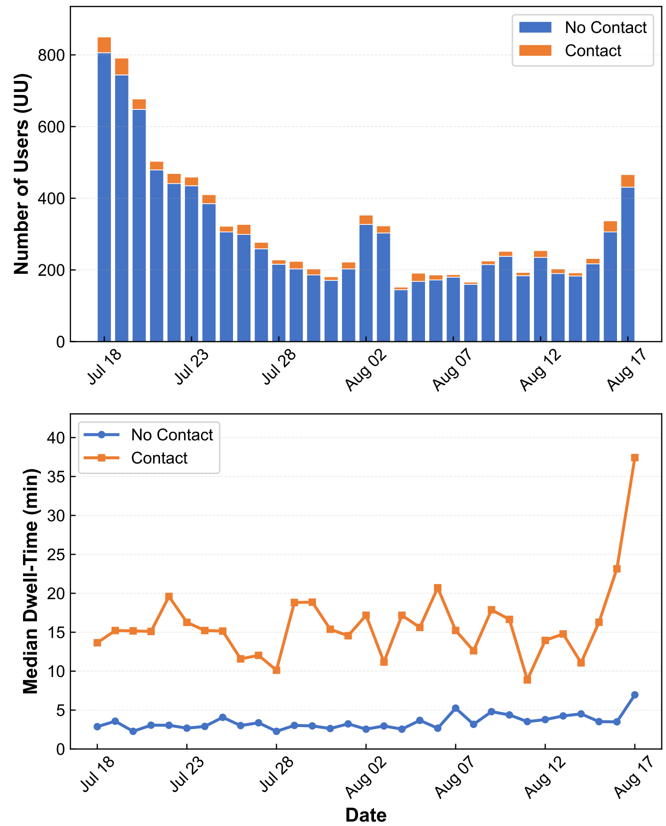

Fig. 3. Temporal dynamics of user engagement and AI agent contact during the field study period (Top) Daily unique users stratified by AI-agent contact status (blue: no contact, orange: contact). (Bottom) Daily median dwell-time by AI-agent contact status.

Figure 3 (top) presents daily unique user (UU) counts stratified by AI-agent contact status throughout the observation period. Substantial increases in user counts occurred on both the initial and final day, with additional localized peaks observed near the beginning of the month.

2) Participant Segmentation: To examine the effects of metaverse proficiency on the efficacy of AI-agent contact, participants were classified into three segments according to their patterns of use of the platform during the seven days prior to entering the initial room.

S1: New Users (1,653 users) – Users who registered for the platform on their first day in the target world. This segment represents initial metaverse experiences and is important for evaluating the impact of AI-agents on onboarding processes.

S2: Existing Low-Frequency Users (2,227 users) – Users with three or fewer days of activity in the seven days

TABLE I
BEHAVIORAL CHARACTERISTICS BY AI-AGENT CONTACT STATUS ACROSS ALL PARTICIPANTS (N=5,020) DURING THE 31-DAY STUDY PERIOD.

Metric	Overall	Contact	No Contact
Unique users (UU)	5,020	328 (6.5%)	4,692 (93.5%)
Mean entry days	2.0	5.4	1.8
Median dwell time (min)	4.7	36.2	4.2
Emote usage rate in UUs (%)	23.7	54.6	21.6
Mean emotes per user	63	187	41
Comment posting rate in UUs (%)	15.4	49.7	13.0
Mean comments per user	27	52	20

before entering the room for the first time. This segment represents users who have completed the platform registration but have not achieved sustained usage.

S3: Existing High-Frequency Users (1,140 users) – Users with four or more days of entries in the seven days prior to their initial room entry. This segment represents habitual users who are familiar with the platform.

V. RESULTS

A. Descriptive Statistics

1) Behavioral Characteristics: Table I presents the general behavioral characteristics of the participants and the basic differences by AI-agent contact status. On average, the participants spent 2.0 days in the study, with a median dwell time of 4.7 minutes. Social interaction metrics revealed emote usage rates of 23.7% and comment posting rates of 15.4% among all unique users. When comparing the contact group (C) and non-contact group (NC), C showed superior performance across all behavioral metrics.

Figure 3 (bottom) shows daily variations in the median dwell-time by contact status. Users who interacted with the AI-agent consistently exhibited higher median dwell-times throughout the observation period. However, these measurements include the time spent in dialog with the AI agent, making it difficult to interpret direct causal relationships. To control for this potential confounding factor, we examine behavioral changes for each user before and after their initial entry into the event world.

2) Defination of Log-Change Rate: To control for the confounding effects of users' prior metaverse usage frequency on the results of our analysis, we conducted a stratified analysis by usage experience level. To quantify behavioral changes at the individual level, we define individual-level log change rates, L_i, as follows:

$$L_i = \log_{10} \frac{Y_{i,post} + \epsilon}{Y_{i,pre} + \epsilon} \quad (1)$$

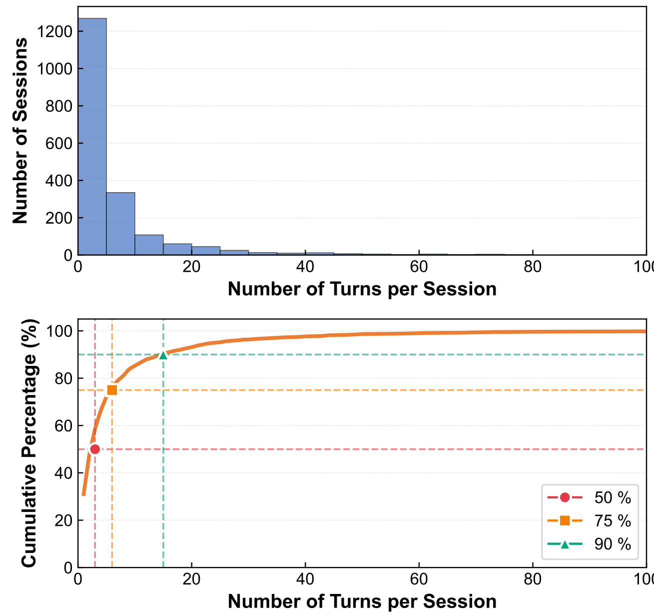

Fig. 4. Distribution of conversation turns per session with the AI agent. (Top) Histogram showing the frequency distribution of conversation turns per session. (Bottom) Cumulative distribution function with percentile markers (50th, 75th, and 90th percentile indicated by dashed lines and markers).

Where $Y_{i,\text{post}}$ represents the 7-day post-contact average, $Y_{i,\text{pre}}$ represents the 7-day pre-contact average, and $\epsilon = 10^{-4}$ serves as a small constant for zero-value handling. Log transformation enables symmetric treatment of $\times 2$ increases and $\times 0.5$ decreases, and enabling unified analysis including new users and existing users with zero-login records in the 7 days prior to contact.

3) Segment-Wise Effect Pattern Verification: Figures 5 show the log-change rates for the entry days and the dwell time in the segments. Mann-Whitney U tests were performed to statistically verify log-change rates between the C and NC groups within each segment. This test was selected for its robustness against extreme sample size imbalances that resulted from the small contact group. Cliff's delta (δ) was calculated as an effect size measure to evaluate the magnitude of the practical effect.

S1: New Users – Log-change rates showed $\times$ 2.1 $\sim$ 2.3 difference (C mean: 2.090 vs. NC mean: 0.998 for entry days, and C mean: 3.186 vs. NC mean: 1.427 for dwell time). Statistical testing confirmed highly significant effects for both metrics (entry days: $p < 0.001$, δ = 0.382 and dwell time: $p < 0.001$, δ = 0.392).

S2: Existing Low-Frequency Users – Log-change rates showed $\times$ 2.2 $\sim$ 2.4 difference for entry days (C mean: 0.346 vs. NC mean: 0.143 for entry days, and C mean: 0.495 vs. NC mean: 0.223 for dwell time), indicating improvement trends from AI-agent contact. However, the statistical results did not reach significance for both entry days (p = 0.059, δ = 0.116) and dwell time (p = 0.078, δ = 0.111).

S3: Existing High-Frequency Users – Log-change rates

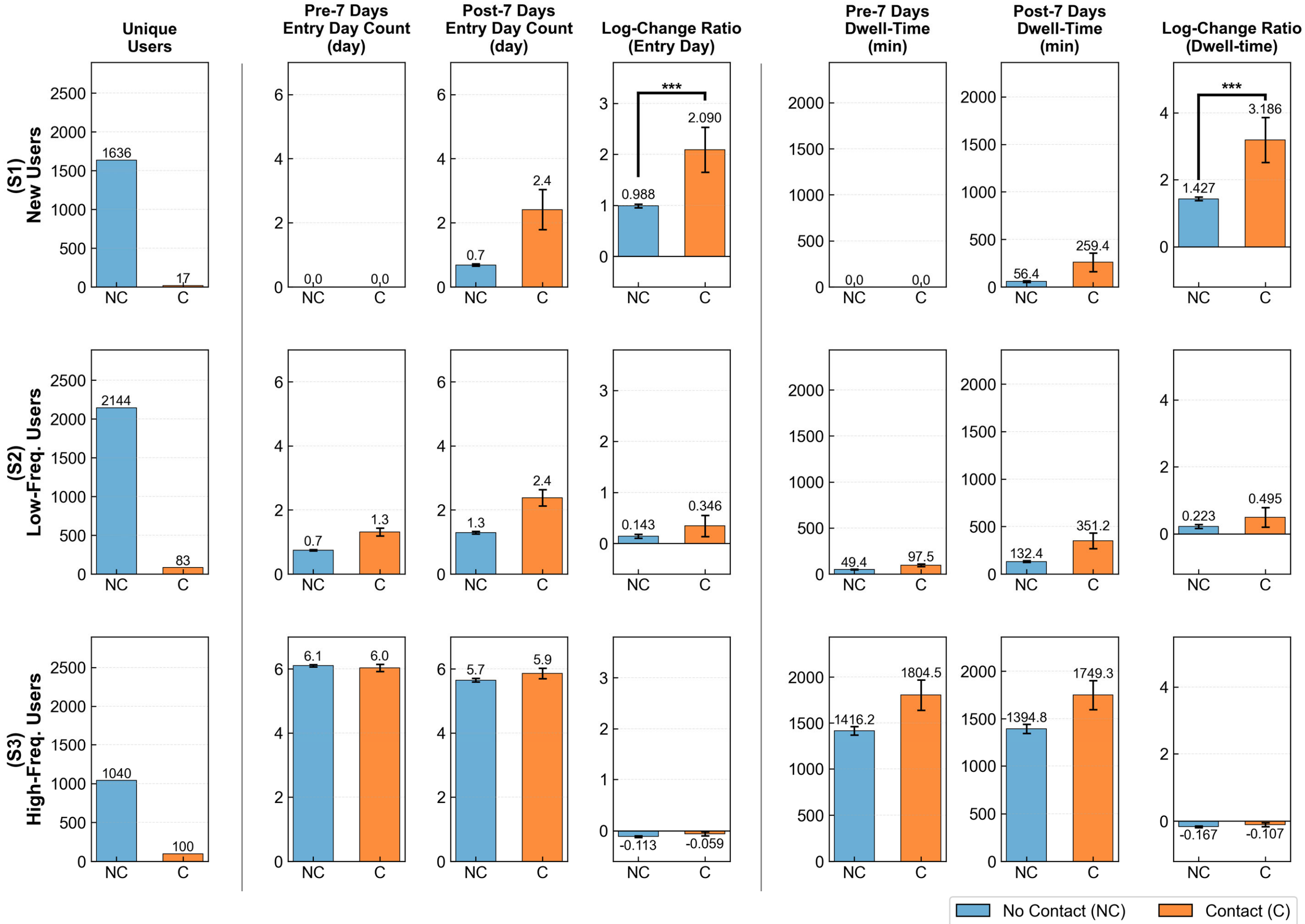

Fig. 5. Segment-wise comparison of platform usage behavior by AI-agent contact status. Rows represent (S1) new users (n=1,653), (S2) existing low-frequency users (n=2,227), and (S3) existing high-frequency users (n=1,140). Columns show: (1) unique user distribution, (2,3) pre- and post-7-day entry day counts, (4) entry day log-change ratios, (5,6) pre- and post-7-day dwell times (minutes), and (7) dwell-time log-change ratios, relative to initial event-world entry for each user. NC: no contact (blue); C: contact (orange). Error bars: standard error. ***$p < 0.001$.

showed slightly negative values (C mean: -0.059 vs. NC mean: -0.113 for entry days, C mean: -0.107 vs. NC mean: -0.167 for dwell time). This pattern indicates ceiling effects among users with already high activity levels. Statistical results did not show significance for entry days (p = 0.250, δ = 0.066) or dwell time (p = 0.837, δ = 0.012).

B. Conversation Log Analysis

To understand how users engaged with the AI agent, we analyzed the interaction patterns between users and the agent. The AI agent generated 20,669 responses, while users submitted 12,377 messages. The dialog sessions were segmented when comments ceased for more than 15 minutes, yielding 1,913 sessions for analysis.

1) Session Characteristics: Figure 4 shows the distribution of utterances per session. The average conversation length was 6.47 exchanges (median: 3; maximum: 178). Single-utterance sessions accounted for 31.1% (594 cases), while multi-turn conversations accounted for 68.9% (1,319 cases), indicating that approximately two-thirds of users engaged in dialogic interactions. Notably, 16.2% of users participated in extended conversations (10+ exchanges), demonstrating the LLM-based AI agent's capacity to sustain user engagement.

2) Conversation Pattern Classification: Table II categorizes multi-turn conversation patterns by type. The most prevalent category was *question-conversation type* (286 sessions, 15.0%), where users posed sequential questions that evolved into natural dialogue. The longest average conversations occurred in *game-play type* (mean: 25.8 exchanges), where users enjoy word games like the last-letter game.

Within the question-conversation category, we identified 6 dominant inquiry types: (1) location queries (108 instances, 7.8%), such as *"Where can I find delicious shaved ice in Roppongi?"*; (2) request-type questions (91 instances, 6.5%), such as *"May I rest here?"*; (3) event-related questions (42 instances, 3.0%), such as *"How long will this event run?"*; (4) questions about the agent itself (38 instances, 2.7%), such as *"What kind of animal are you?"*; (5) functionality questions (24 instances, 1.7%), such as *"What can you do?"*; and (6)

TABLE II
CONVERSATION SESSION PATTERN CLASSIFICATION

Category	n	%	Mean	Med.	Max
Single-utterance	441	23.1	1.0	1	1
Question-conversation	286	15.0	13.9	9	73
Very short utterance	238	12.4	1.9	1	19
Free conversation	185	9.7	8.1	7	38
Call-out	178	9.3	4.4	3	21
Other	139	7.3	4.2	3	17
Game-play	132	6.9	25.8	19	178
Short responses	123	6.4	4.4	3	26
Greeting/farewell	85	4.4	4.2	3	48
Gratitude	29	1.5	3.2	2	13
Request	25	1.3	3.0	2	7
Guidance	21	1.1	2.6	2	5
Emotional reaction	12	0.6	2.6	2	4
Conv. with others	10	0.5	2.9	2	7
Ask Photography	9	0.5	4.0	4	7
Total	**1,913**	**100.0**	**6.5**	**3**	**178**

language/translation questions (10 instances, 0.7%), such as *"How do you say 'good morning' in French?"*.

3) Representative Interaction Patterns: This section provides real-life examples of interactions between users and AI agents:

Deep Question Sequences. Users developed successive questions that evolved into extended dialogues, demonstrating sustained curiosity about the agent's characteristics and preferences:

> (U: User, A: Agent)
> U: "Go-chan, you like vegetables, but do you like fruit too?"
> A: "I love fruit! I especially like strawberries."
> U: "What fruit do you like?"
> A: "Strawberries are my favorite, but bananas are delicious too!"
> U: "Strawberry daifuku (mochi filled with red bean paste) is the best, right!"
> A: "Absolutely! The combination of strawberries and Japanese sweets is perfect."

Emotional Bonds. As the event approached its conclusion, users engaged in emotionally expressive dialogues expressing reluctance to part with the AI agent:

> U: "Including today, only 3 days left???"
> A: "That's right. But let's enjoy these last 3 days together!"
> U: "I'm so sad that I won't be able to see Go-chan in 3 days..."
> A: "I'm sad too, but let's make lots of fun memories!"

These patterns demonstrate that users treated the AI agent not merely as an information retrieval tool, but as a conversational partner with whom they could form emotional connections and engage in sustained dialogue.

C. Cumulative Effect Analysis

To distinguish between the immediate and cumulative effects of AI-agent contact and examine their detailed impact on next-day platform usage continuation intentions, we conducted an individual fixed-effects logistic regression analysis. Our objective was to determine whether AI-agent contact produces only short-term effects or maintains cumulative ones, while controlling for user-specific, time-invariant characteristics.

1) Variable Definitions and Model Specification: We index users as $i = 1, 2, \ldots, N$ and study period days as $t = 1, 2, \ldots, T$. Observation units are restricted to days when the user i entered the event world on day t, the dependent variable being the Cluster login status on day $t + 1$.

The dependent variable $Y_{i,t+1}$ indicates the next-day Cluster login behavior ($Y_{i,t+1} = 1$ if user i logs in on day $t + 1$, 0 otherwise). Independent variables include current-day AI-agent contact $A_{i,t}$ (1 if contact occurred, 0 otherwise), cumulative event world entry days $n_{i,t} = \sum_{s=1}^{t} L_{i,s}$, cumulative AI-agent contact days $k_{i,t} = \sum_{s=1}^{t} A_{i,s}$, previous-day event world entry $L_{i,t-1}$, previous-day AI-agent contact $A_{i,t-1}$, and linear time trend t. Complete variable definitions are provided in Table IV.

We employ a conditional fixed-effect logistic regression model specified as:

$$\operatorname{logit} P(Y_{i,t+1} = 1) = \alpha_i + \beta_1 A_{i,t} + \beta_2 n_{i,t} + \beta_3 k_{i,t} + \beta_4 L_{i,t-1} + \beta_5 A_{i,t-1} + \beta_6 t \qquad (2)$$

where α_i represents individual fixed effects, capturing time-invariant user characteristics including baseline propensity for platform engagement and continuation behavior.

The conditional likelihood approach eliminates α_i, allowing for within-individual comparisons that control for unobserved heterogeneity. This specification requires users to exhibit variation in the next-day login behavior to contribute to parameter identification. Users with invariant outcomes (always logging in or never logging in following event world visits) provide no information for coefficient estimation, regardless of AI-agent contact variation. This constraint reduces the analysis sample from 5,020 to 771 users but ensures that estimated effects reflect within-individual behavioral changes attributable to AI-agent contact rather than between-individual differences.

2) Regression Results: The fixed-effect model analysis uses 3,683 observations from 771 users. Table III presents coefficient estimates, standard errors, test statistics, and odds ratios. Statistical significance is evaluated using Wald tests, where test statistics $z = \hat{\beta}/SE(\hat{\beta})$ follow standard normal distributions under the null hypothesis $H_0 : \beta = 0$. We report two-tailed p-values.

The current-day AI-agent contact effect ($\beta_1 = 0.170$, $z = 0.880$, $p = 0.379$) is not statistically significant, indicating no detectable immediate effect on next-day login probability within individuals. In contrast, cumulative AI-agent contact demonstrates a significant positive effect ($\beta_3 = 0.289$, $z = 3.461$, $p < 0.001$, odds ratio = 1.335). Each additional cumulative contact day increases next-day login odds by 33.5% within individuals, suggesting that sustained exposure to AI-agents produces substantial facilitative effects on continuation behavior.

Cumulative event-world entries exhibit a significant negative effect ($\beta_2 = -0.215$, $z = -7.605$, $p < 0.001$, odds ratio = 0.807), reducing next-day login odds by 19.3% per additional entry. This pattern reflects usage fatigue or satiation effects that emerge with repeated platform engagement. Similarly, previous-day event world entry similarly shows a strong negative effect ($\beta_4 = -0.820$, $z = -9.331$, $p < 0.001$, odds ratio = 0.440), decreasing next-day login odds by 56.0%, consistent with attenuation following consecutive usage. Previous-day AI-agent contact ($\beta_5 = -0.072$, $z = -0.279$, $p = 0.781$) yields no significant effect.

The time trend coefficient ($\beta_6 = 0.030$, $z = 3.631$, $p < 0.001$, odds ratio = 1.031) indicates a 3.1% increase in next-day login odds per day within the study period, suggesting modest increases in engagement propensity over time within individuals.

VI. Discussion

A. Hypothesis Verification

The verification results for the four hypotheses established in this study are shown below:

H1: Partial Support – A segment-based analysis revealed that the effects of AI-agent contact differ markedly depending on the level of user experience. Statistically significant effects were confirmed only among new users (S1), while existing users did not reach significance. These results demonstrate that the effects of the AI-agent contact do not manifest uniformly across all users but rather concentrate within specific populations. However, despite the absence of significant differences, existing low-frequency users (S2) showed increasing usage trends.

H2: Strong Support – The statistically significant effects observed only among new users strongly support this hypothesis. This finding shows that AI-agents act as social scaffolding, supporting the adaptation process to metaverse environments. The limited effects observed among existing users indicate that AI-agents' primary value lies in onboarding support. This is an important discovery for providing empirical solutions to the fundamental challenge of retaining new users faced by metaverse platforms.

H3: Reject – The effects of AI-agent contact were not statistically significant with regard to the next-day metaverse login, which led to the rejection of hypothesis H3. This suggests that the AI-agent does not influence immediate emotional responses, such as thinking "Today was enjoyable, so I will log in again tomorrow.", and clarifies that the essence of the effects of the AI-agent lies in the accumulation of continuous interactions rather than isolated contacts.

H4: Support – The most important finding of this study is the demonstration of significant positive effects of cumulative contact days with AI-agents on the probability of continuing the next day under individual difference controls. This effect was confirmed as the only positive factor that persisted against the attenuation effects of fatigue from general usage. The qualitative analysis of conversation logs also show that users engaged in extended conversations, which demonstrates the AI agent's capacity to sustain user attention and facilitate relationship building. These results suggest that building relationships with AI-agents functions as a genuine mechanism of behavioral change.

B. Practical Implications

The verification results of the hypotheses provide strategic insights for designing and operating AI-agents on the metaverse platform.

First, the observation of notable effects with a contact rate of 6.5% demonstrates that significant results can be achieved by providing opportunities for interaction during natural exploration. Although this study examined a single AI-agent, improvements in platform-wide user retention are anticipated through the deployment of multiple AI-agents and proactive engagement, which will increase contact rates.

Second, the pronounced effects on new users demonstrate the effectiveness of AI-agent-based onboarding strategies. From a resource allocation perspective, improving AI-agent functionality for new users is likely the most efficient investment.

Third, the discovery of cumulative effects highlights the importance of designing opportunities for continuous contact. The progression from functional to emotional engagement observed across multiple sessions underscore the value of persistent agent presence rather than episodic deployment. Integrating natural contact opportunities within routine usage processes is more effective for improving long-term engagements. Specifically, implementing automatic contact mechanisms upon user entry and relationship continuation functionality through personalized dialogue based on past contact history (e.g., recognition expressions such as "We've met before!") is expected to further promote cumulative effects.

VII. Limitations and Future Work

Our naturalistic approach provided essential ecological validity by capturing authentic user behavior in actual commercial service environments, and demonstrating measurable impacts, particularly among new users, establishes important proof of concept for AI-agent effectiveness in real-world metaverse platforms. In this section, we discuss the remaining challenges and future research directions.

A. Causal Inference and Individual Differences

Although this study attempted to control for individual differences, there is ambiguity regarding how well the intercept α_i reflecting individual behavioral characteristics captures true personal characteristics. Model construction incorporating metaverse platform login rates before the event through difference-in-differences methods, or comparative analysis between users with similar behavioral characteristics through propensity score matching, could make causal interpretations of observed effect patterns more robust. Also, elucidating

TABLE III
FIXED-EFFECT LOGISTIC REGRESSION RESULTS

Coefficient	Variable	Estimate	SE	z	p	Odds Ratio	Interpretation
β_1	$A_{i,t}$	0.170	0.193	0.880	0.379	1.186	No immediate effect
β_2	$n_{i,t}$	$\mathbf{-0.215}^{***}$	0.028	-7.605	$<\mathbf{0.001}$	0.807	19.3% decrease
β_3	$k_{i,t}$	$\mathbf{0.289}^{***}$	0.084	3.461	$<\mathbf{0.001}$	1.335	**33.5% increase**
β_4	$L_{i,t-1}$	$\mathbf{-0.820}^{***}$	0.088	-9.331	$<\mathbf{0.001}$	0.440	56.0% decrease
β_5	$A_{i,t-1}$	-0.072	0.258	-0.279	0.781	0.931	No lagged effect
β_6	t	$\mathbf{0.030}^{***}$	0.008	3.631	$<\mathbf{0.001}$	1.031	3.1% increase per day

Notes: Standard errors clustered at the individual level. Wald tests used for significance testing: $z = \hat{\beta}/SE(\hat{\beta})$. $^{***}p < 0.001$. Odds ratios calculated as $\exp(\hat{\beta})$. Interpretation column shows percentage change in odds for interpretable effects.

TABLE IV
VARIABLE DEFINITIONS

Variable	Definition
Dependent Variable	
$Y_{i,t+1}$	Next-day Cluster login status (1 = logged in, 0 = not logged in)
Independent Variables	
$A_{i,t}$	AI-agent contact on day t (1 = contact, 0 = no contact)
$n_{i,t}$	Cumulative event world entry days through day t ($n_{i,t} = \sum_{s=1}^{t} L_{i,s}$)
$k_{i,t}$	Cumulative AI-agent contact days through day t ($k_{i,t} = \sum_{s=1}^{t} A_{i,s}$)
$L_{i,t-1}$	Previous day event world entry (1 = entry, 0 = no entry)
$A_{i,t-1}$	Previous day AI-agent contact (1 = contact, 0 = no contact)
t	Day within study period (linear time trend)
α_i	Individual fixed effects (time-invariant user characteristics)

Notes: Observation units restricted to days when user i entered the event world on day t. Individual fixed effects α_i control for time-invariant characteristics through conditional likelihood estimation.

the moderating effects of individual personality, cognitive styles, and social skills would enable the development of more effective personalization strategies.

B. Measurement Metrics and Evaluation Methods

Our definition of ”contact” was limited to active interaction, failing to capture the potential effects on users who passively listened or merely co-existed in the space (vicarious experience). Furthermore, while this study focused on quantitative behavioral metrics, qualitative evaluations such as subjective satisfaction and the quality of social connection were not assessed. Future work requires a comprehensive measurement framework that integrates detailed sentiment analysis of conversations and survey-based qualitative data.

C. Platform Specificity and Generalizability

Our findings derive from a single 31-day event on Cluster, potentially influenced by device differences (VR-HMD vs. flat-screen) and the Japanese cultural context regarding anthropomorphic agents. With VR-HMD usage estimated at 10-15 %, interaction effects with device immersion levels remain unverified. Future research should verify generalizability through replication studies across different platforms and cultural regions, as well as systematic comparisons of agent design parameters, such as appearance and personality.

VIII. ETHICAL CONSIDERATIONS

A. Ethical Challenges in Fieald Research within the Metaverse

This study is a retrospective field study observing users' natural behavior during normal service operations without experimental intervention. Similar large-scale observational research has been widely conducted in SNS [36], [37] and online game studies [20], [22], [23], establishing ethical foundations through comprehensive consent via platform terms of service and privacy policies.

However, metaverse environments differ qualitatively from conventional online platforms in several dimensions: immersive presence, embodied spatial co-presence, and the depth of pseudo-interpersonal relationships with AI agents. This study involves natural language interactions with LLM-based AI agents, introducing distinct ethical considerations including potential user confusion between AI and human interlocutors and the formation of emotional attachments. Rather than directly applying existing Internet research ethics frameworks, careful safety design must account for metaverse-specific characteristics.

B. Ethical Compliance and Safety Measures

Based on the above considerations, we implemented the following ethical protocols:

1) Informed Consent and Transparency: This research was conducted in cooperation with Cluster Inc. according to its privacy policy. Specifically, Items 7, 8, and 14 of Article 2 (Purpose of Personal Information Use) state: ”to investigate and analyze Cluster usage status,” ”to improve our services and develop new services" and "to create statistical data processed into formats that cannot identify individuals." All participants agreed to this privacy policy upon beginning the use of the platform, providing prior consent to the research use of their data. Additionally, AI agent deployment was announced in advance, with "AI" displayed on agent nameplates, surrounding signage, and self-disclosure at dialog initiation, explicitly conveying the agent's AI nature. This prevented the formation of unintended relationships based on misidentification as human.

2) Safety and Privacy Protection: Taking into account the context of family-oriented events, we implemented multi-layered content moderation as detailed in Sec. IV-C. All data collected underwent irreversible anonymization, with analysis conducted exclusively on data with completely removed personally identifiable information.

3) Ethics Review: Although this retrospective observational study was initially outside the scope of the pre-study IRB review, concerns raised during peer review prompted consultation with the Cluster Inc. Research Ethics Committee (No. 2025-011) regarding the research protocol and validity of the data protection measure. The committee issued a conditional approval determination. As conditions for approval, we conducted analysis with personally identifiable information completely excluded and adopted presentation formats making individual user identification impossible. This paper fully complies with these conditions.

C. Toward Establishing Ethical Standards for Metaverse Research

Metaverse research represents an emerging field integrating VR technology, AI technology, and social media research, with ethical frameworks still under development. This study contributes to establishing ethical standards for metaverse research by transparently sharing the concerns raised during the peer review, the retrospective ethics review process, and specific protective measures implemented.

Researchers, platform operators, regulatory authorities, and users themselves must collaborate to construct ethical frameworks appropriate for this emerging field. We hope this study serves as a foundation for such ongoing discussion.

IX. Conclusion

We conducted the first empirical investigation to examine how LLM-based AI-agents influence user continuation behavior in metaverse environments through large-scale, longitudinal tracking in operational settings. While previous research was limited to short-term validation in laboratory conditions, we observed the natural usage patterns of 5,020 unique users over 31 days on the commercial metaverse platform Cluster. This large-scale field study demonstrated that contact with an AI-agent produces persistent changes to users' weekly usage patterns. New users showed particularly significant improvements in stay duration and time metrics. Through causal inference that controlled for individual differences, we determined that providing continuous opportunities for contact with AI-agents is more important than offering isolated experiential encounters. We hope that the insights obtained from this large-scale validation in operational settings will guide the design of metaverse experiences utilizing AI-agents and contribute to collaborative human-AI relationships in digital societies.

Acknowledgment

We thank Takafumi Funabiki and Masaru Yokoi from TV Asahi Corporation for their collaboration in this research.

This work was partially supported by JST Moonshot Research & Development Program Grant Number JPMJMS2013, JST ASPIRE Grant Number JPMJAP2327 and JST PRESTO Grant Number JPMJPR2529, Japan.

References

[1] G. D. Ritterbusch and M. R. Teichmann, "Defining the metaverse: A systematic literature review," *IEEE Access*, vol. 11, pp. 12 368–12 377, 2023.

[2] M. Hide, Y. Hatada, H. Kuzuoka, and T. Narumi, ""closer than real": How social vr platform features influence friendship dynamics," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25. New York, NY, USA: Association for Computing Machinery, 2025. [Online]. Available: https://doi-org.utokyo.idm.oclc.org/10.1145/3706598.3714170

[3] M. Barreda-Ángeles and T. Hartmann, "Psychological benefits of using social virtual reality platforms during the covid-19 pandemic: The role of social and spatial presence," *Computers in Human Behavior*, vol. 127, p. 107047, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563221003708

[4] Statista, ""Number of Virtual Reality (VR) and Augmented Reality (AR) Users in the United States From 2017 to 2023.," https://www.statista.com/statistics/1017008/united-states-vr-ar-users/, (accessed Jan. 30, 2024)."

[5] S.-M. Park and Y.-G. Kim, "A metaverse: Taxonomy, components, applications, and open challenges," *IEEE access*, vol. 10, pp. 4209–4251, 2022.

[6] T. Bickmore, "Relational agents for chronic disease self-management," in *Health Informatics: A Patient-Centered Approach to Diabetes*. The MIT Press, 08 2010. [Online]. Available: https://doi.org/10.7551/mitpress/8336.003.0011

[7] 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*, ser. CHI '24. New York, NY, USA: Association for Computing Machinery, 2024. [Online]. Available: https://doi.org/10.1145/3613904.3642235

[8] 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*. New York, NY, USA: Association for Computing Machinery, 2024. [Online]. Available: https://doi.org/10.1145/3613905.3651026

[9] H. Yanagawa, Y. Hiroi, S. Tokida, Y. Hatada, and T. Hiraki, "Navigation pixie: Implementation and empirical study toward on-demand navigation agents in commercial metaverse," 2025. [Online]. Available: https://arxiv.org/abs/2508.03216

[10] C. Nass and Y. Moon, "Machines and mindlessness: Social responses to computers," *Journal of Social Issues*, vol. 56, no. 1, pp. 81–103, 2000. [Online]. Available: https://spssi.onlinelibrary.wiley.com/doi/abs/10.1111/0022-4537.00153

[11] 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*, vol. 114, no. 1, pp. 3–55, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0004370299000703

[12] 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.

[13] D. Jan, A. Roque, A. Leuski, J. Morie, and D. Traum, "A virtual tour guide for virtual worlds," in *Intelligent Virtual Agents*, ser. Lecture notes in computer science. Springer Berlin Heidelberg, 2009, pp. 372–378.

[14] 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*, vol. 2, pp. 24–42, 01 2004.

[15] 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)*, Mar. 1998, pp. 191–200.

[16] T. W. Bickmore and R. W. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Trans. Comput.-Hum. Interact.*, vol. 12, no. 2, p. 293–327, Jun. 2005. [Online]. Available: https://doi.org/10.1145/1067860.1067867

[17] N. Techasarntikul, 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 *ICAT-EGVE 2019 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, Y. Kakehi and A. Hiyama, Eds. The Eurographics Association, 2019.

[18] 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*, ser. IVA '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 33–40. [Online]. Available: https://doi.org/10.1145/3472306.3478339

[19] 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*, vol. 4, 2024. [Online]. Available: https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2023.1334795

[20] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore, ""alone together?" exploring the social dynamics of massively multiplayer online games," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006, pp. 407–416.

[21] A. F. Seay, W. J. Jerome, K. S. Lee, and R. E. Kraut, "Project massive: A study of online gaming communities," in *CHI'04 extended abstracts on Human factors in computing systems*, 2004, pp. 1421–1424.

[22] S. Demediuk, A. Murrin, D. Bulger, M. Hitchens, A. Drachen, W. L. Raffe, and M. Tamassia, "Player retention in league of legends: a study using survival analysis," in *Proceedings of the Australasian Computer Science Week Multiconference*, ser. ACSW '18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: https://doi.org/10.1145/3167918.3167937

[23] H. Kang, C. Suh, and H. K. Kim, "Match experiences affect interest: Impacts of matchmaking and performance on churn in a competitive game," *Heliyon*, vol. 10, no. 3, p. e24891, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405844024009228

[24] H. Tsutsui, T. Hiraki, Y. Hiroi, and S. Hasegawa, "Community analysis of social virtual reality based on large-scale log data of a commercial metaverse platform," 2025. [Online]. Available: https://arxiv.org/abs/2509.23654

[25] B. S. Butler, "Membership size, communication activity, and sustainability: A resource-based model of online social structures," *Information systems research*, vol. 12, no. 4, pp. 346–362, 2001.

[26] C. Lampe and E. Johnston, "Follow the (slash) dot: effects of feedback on new members in an online community," in *Proceedings of the 2005 ACM International Conference on Supporting Group Work*, 2005, pp. 11–20.

[27] F. B. Viégas and M. Smith, "Newsgroup crowds and authorlines: Visualizing the activity of individuals in conversational cyberspaces," in *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*. IEEE, 2004, pp. 10–pp.

[28] S. Baker, R. M. Kelly, J. Waycott, R. Carrasco, T. Hoang, F. Batchelor, E. Ozanne, B. Dow, J. Warburton, and F. Vetere, "Interrogating social virtual reality as a communication medium for older adults," *Proceedings of the ACM on human-computer interaction*, vol. 3, no. CSCW, pp. 1–24, 2019.

[29] K. J. Byun and S. J. g. Ahn, "A systematic review of virtual influencers: Similarities and differences between human and virtual influencers in interactive advertising," *J. Interact. Advert.*, vol. 23, no. 4, pp. 293–306, Oct. 2023.

[30] A. Laszkiewicz and M. Kalinska-Kula, "Virtual influencers as an emerging marketing theory: A systematic literature review," *Int. J. Consum. Stud.*, vol. 47, no. 6, pp. 2479–2494, Nov. 2023.

[31] E. Cesanek, S. Shivkumar, and D. Wolpert, "Ouvrai opens access to remote virtual reality studies of human behavioural neuroscience," *Nature Human Behaviour*, vol. 8, pp. 1–16, 04 2024.

[32] A. Steed, L. Izzouzi, K. Brandstätter, S. Friston, B. Congdon, O. Olkkonen, D. Giunchi, N. Numan, and D. Swapp, "Ubiq-exp: A toolkit to build and run remote and distributed mixed reality experiments," *Frontiers in Virtual Reality*, vol. 3, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frvir.2022.912078

[33] 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*, ser. CHI LBW, 2020, pp. 1–8, preprint & supplemental material: https://osf.io/w569f.

[34] R. Kurai, T. Hiraki, Y. Hiroi, Y. Hirao, M. Perusquía-Hernández, H. Uchiyama, and K. Kiyokawa, "Magicitem: Dynamic behavior design of virtual objects with large language models in a commercial metaverse platform," *IEEE Access*, vol. 13, pp. 19 132–19 143, 2025.

[35] ——, "Magiccraft: Natural language-driven generation of dynamic and interactive 3d objects for commercial metaverse platforms," *IEEE Access*, vol. 13, pp. 132 459–132 474, 2025.

[36] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 2007, pp. 29–42.

[37] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao, "User interactions in social networks and their implications," in *Proceedings of the 4th ACM European conference on Computer systems*, 2009, pp. 205–218.