Verselyzer: A Statistical Analysis Interface Enabling Non-Experts to Assess Metaverse Branding Effectiveness

Sotaro Yokoi*,‡
Cluster Metaverse Lab
The University of Tokyo

Satomi Tokida†,‡
Cluster Metaverse Lab
The University of Tokyo

Yuichi Hiroi§
Cluster Metaverse Lab

Takefumi Hiraki¶
University of Tsukuba
Cluster Metaverse Lab

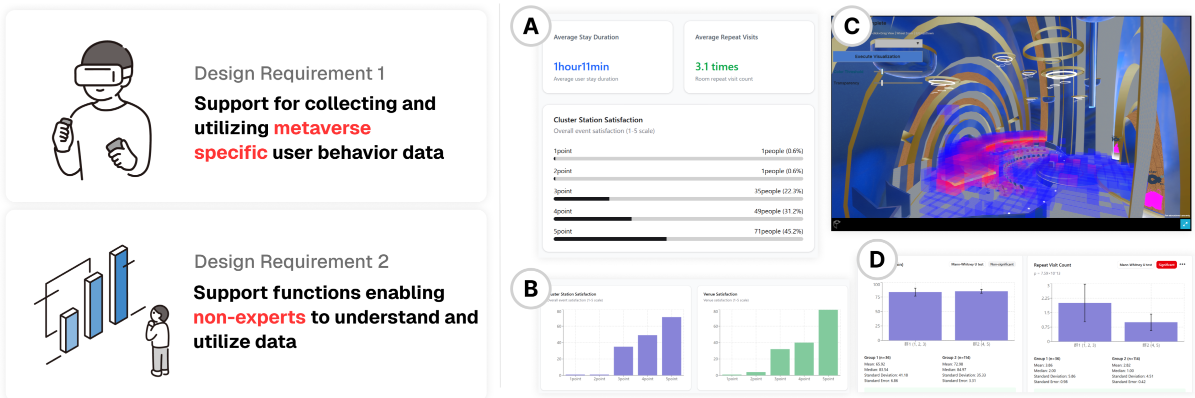

Figure 1: Concept of Verselyzer system. (Left) Required features for metaverse branding tools based on surveys with branding practitioners. (Right) Screenshots of the web interface we built. A: Overview tab showing survey results and user behavior metrics in a list. B: Data Visualization tab showing user behavior in graphs. C: The Spatial Visualization tab uses WebGL to show the user stay duration or visitor count by area. D: The Statistical Analysis tab enables users to run statistical tests by simply clicking buttons.

Abstract

Branding in metaverse spaces offers significant potential for immersive customer engagement through embodied, physical-world-like interactions. However, unlike traditional web-based marketing, metaverse branding involves users' active, spatial behaviors, requiring different KPI design and evaluation methods that remain underexplored. To address this gap, we propose "Verselyzer," a web interface that enables non-experts to design KPIs and evaluate branding effectiveness in metaverse environments. Based on a formative study with seven branding practitioners, we derived design requirements emphasizing (1) metaverse-specific behavioral data collection and (2) accessible analysis support for non-experts. The resulting system progressively supports hypothesis formulation, data visualization, comprehension of spatial behavior, and simplified statistical analysis. A user study with 11 practitioners suggested that Verselyzer can support hypothesis verification, while also indicating that statistical tools can uncover hidden insights and help formalize experience-based hypotheses. These findings provide practical implications for improving metaverse branding practices as commercial use of virtual spaces continues to grow.

Index Terms: Metaverse, branding, visualization interface, formative study.

*e-mail: s.yokoi@cluster.mu
†e-mail: s.tokida@cluster.mu
‡These authors contributed equally to this work.
§e-mail: y.hiroi@cluster.mu
¶e-mail: hiraki@slis.tsukuba.ac.jp

1 Introduction

Metaverse platforms such as VRChat [14], Cluster [4], and Resonite [16] are gaining popularity, and companies are increasingly using virtual spaces as new touchpoints for customer engagement and immersive promotion. The metaverse has attracted attention for its ability to deliver immersive brand experiences through interpersonal interactions that closely mirror those in the real world [6, 11]. Moreover, unlike in the real world, the metaverse enables the collection of fine-grained behavioral logs—such as user position, posture, speech, and gaze—allowing user experiences to be quantified more precisely than in conventional web-based analysis. Prior work such as LUIDA [7] and Ubiq-exp [13] has shown that metaverse platforms can be used as experimental infrastructures for collecting large-scale behavioral data.

To transform the rich behavioral data available in the metaverse into commercial success, it is necessary to define intermediate objectives aligned with the final goal and formalize them as measurable indicators, namely Key Performance Indicators (KPIs) [10]. Such measurement of user behavior and goal-oriented use of indicators has already been widely practiced in real-world marketing. For example, user behavior tracking has been used to visualize customer experience flows and to deliver advertisements based on users' location information, with the aim of achieving final commercial outcomes [2]. In addition, Philippopoulos et al. [12] evaluated museum performance using observable behavioral indicators such as visitor counts.

However, verifying branding effects using such rich metaverse data requires advanced analytical skills, and existing real-world knowledge on KPIs has not yet been fully explored in this context. For example, the existing "brand immersive time" metric [8] is a simple model that relies solely on dwell time and device coefficients and thus does not adequately capture users' diverse behaviors and intentions. In contrast, prior research has proposed analysis-support tools that enable UX (User Experience) designers to define tasks with minimal effort, allowing even non-experts to handle large-scale logs [5, 15]; similar support is considered effective in metaverse spaces.

To address this research gap, we interviewed planners and designers of commercial events and spaces in the metaverse platforms with large user bases. Based on the findings, we derived design requirements, developed a web-based prototype, and conducted a user study. Specifically, we propose "Verselyzer," a web interface that enables non-experts to design KPIs and evaluate intervention effects. The system incrementally supports the process from hypothesis generation to visualization and statistical analysis by integrating actual user behavioral data collected on Cluster with questionnaire results. Furthermore, through a user study, we examine the effectiveness of technical support for KPI-based evaluation in metaverse marketing.

2 Formative Study

2.1 Method

To clarify practical approaches to metaverse branding and identify functional requirements for support tools, we conducted a semi-structured questionnaire. All procedures in this research were conducted with the approval of the Cluster, Inc. Research Ethics Committee (Approval No. 2025-004). Seven practitioners involved in metaverse marketing participated in a Google Forms survey. We assigned anonymized identifiers (A1–A7) to respondents and attributed each quote to its source using these IDs. Responses were provided in Japanese and translated into English by the authors.

2.2 Analysis

We qualitatively analyzed open-ended responses to identify recurring practices and challenges, then synthesized them into Design Requirements (DRs) for the proposed support system (Fig. 1 left).

2.3 Findings and Design Requirements

We derived two design requirements from the findings.

DR1: Support for collecting and utilizing metaverse-specific user behavior data. Respondents described workflows that tightly couple client KPIs, hypothesis building, experience design, and instrumentation, rather than treating measurement as an afterthought. They emphasized that metaverse branding presupposes embodied, spatial, and interactive behaviors, which motivates tracking rich behavioral logs beyond conventional pageviews and clicks.

> "It would be helpful if it were easy to identify what was ultimately effective within the world (that is, what users viewed and through which navigation paths they moved)." *[A5]*

In addition, metaverse commercial campaigns were characterized as longitudinal and extensible, requiring planning for post-launch evolution and ongoing measurement.

> "Compared to conventional digital branding, metaverse initiatives are long-term and highly extensible. We propose how the post-launch user journey and in-metaverse experiences will be scaled or stepped up." *[A2]*

These views were echoed by other respondents, who emphasized tracking embodied and spatial behaviors such as avatar movement patterns, emote usage, and in-world navigation paths (A1, A3, A4, A6). These findings indicate that support tools must (i) guide data exploration (i.e., what data exists and where to find it), (ii) provide interpretable explanations of analyses, and (iii) help non-experts progressively perform hypothesis verification and reporting.

DR2: Support functions enabling non-experts to understand and utilize data. Respondents noted that meaningful analysis is difficult without specialist staff, leading to superficial reporting that fails to leverage the rich data available. They also reported high overhead in determining which datasets to consult, suggesting that tool support should address data discovery and guidance, not only visualization.

> "Intuitive operation, explanations using easy-to-understand terminology, and the expectation that anyone can use it and anyone viewing the results can understand them." *[A7]*

Accordingly, respondents requested interpretability support: statistical comparisons accompanied by explanations that non-experts can understand, ideally accessible even to those with a limited mathematical background.

> "Given that many practitioners driving projects often lack statistical knowledge, it is necessary not only to show aggregates but also to provide explanations alongside statistical processing and comparisons . . . A considerate design is needed so that even middle-school-level math knowledge suffices to understand what the numbers represent." *[A1]*

They also articulated a need for interactive assistance that answers ad hoc questions about which data to use in a given situation.

> "Because the data needed changes depending on the situation, I want an AI concierge that tells me the answer when I ask, 'I want this kind of data; which should I look at?' " *[A3]*

Similar concerns about data discoverability and the need for guidance were raised by three other respondents (A5, A6). These findings indicate that support tools must (i) capture spatial and embodied behavioral logs and (ii) connect user experiences to KPIs for longitudinal evaluation.

3 Verselyzer

Based on the formative study, we developed "Verselyzer," a web interface designed to support branding practitioners. Verselyzer is designed to provide comprehensive support from hypothesis formulation through analysis and verification, using user behavior logs and survey results obtained in metaverse spaces (Fig. 1 right). The system primarily consists of the following four functional tabs:

- **Overview**: Displays fundamental data such as user count, duration of stay, and survey results in a comprehensive view, enabling users to grasp the overall project status (DR1, Fig. 1Right A).
- **Data Visualization**: Presents survey results and behavior logs in graphical formats, allowing users to visually understand trends (DR2, Fig. 1Right B).
- **Spatial Visualization**: Maps the number of visitors and duration of stay within the world, enabling comprehension of behavioral patterns in the metaverse space (DR1, Fig. 1Right C).
- **Statistical Analysis**: Allows grouping users by attributes or behaviors and enables easy inter-group comparisons and hypothesis testing (DR2, Fig. 1Right D).

4 User Study

4.1 Procedure

To evaluate the usefulness and usability of our system, we conducted a user study. Eleven practitioners engaged in metaverse branding and marketing initiatives participated in the study (experience: 0.5–6 years; M=2.65, SD=1.62). We used behavioral logs and survey results from initiatives conducted on the metaverse platform "Cluster." To approximate real-world practice, the study consisted of two steps: (1) *Hypothesis formulation*: participants cross-referenced survey items with behavioral data to formulate multiple marketing hypotheses; and (2) *Hypothesis verification*: participants configured analysis conditions in the proposed system and verified the hypotheses by analyzing and comparing groups. After completing the tasks, we administered the System Usability Scale (SUS) [3] and conducted short post-task interviews. Each session lasted approximately 30 minutes.

4.2 Results

4.2.1 SUS

The SUS results showed a mean score of 70.8 (min. 52.5, max. 82.5, SD 8.1), indicating sufficient acceptability for use by practitioners [1].

4.2.2 Interview Findings.

We organized findings into three themes aligned with our design goals: (1) hypothesis formulation process, (2) perceived system strengths, and (3) usability challenges and requested extensions. We cited representative statements from the interviews (e.g. *[B1]*).

Hypothesis Formulation Process. Participants reported that they formulated hypotheses by combining the system's Overview and Visualization tabs with their own domain-specific heuristics. At least five of eleven participants explicitly mentioned relying on their own experiential assumptions when interpreting the dashboard (B4, B5, B8, B9, B10). Some began with the Overview tab to grasp overall trends, while others sought patterns by triangulating multiple outputs. Notably, practitioners explicitly described injecting their experiential assumptions into their interpretation of the dashboard, such as skepticism about the reliability of certain demographic attributes.

> "I referred to the Overview and Visualization tabs, and also my own heuristics." *[B5]*

> "I first checked the satisfaction data, then looked at behavioral logs; I also relied on my heuristics (e.g., age/gender may be less reliable)." *[B8]*

These findings indicate that analysis tools should augment, rather than replace, practitioners' existing expertise and judgment.

Perceived System Strengths Two key strengths emerged from the interviews. First, participants valued at-a-glance comprehension of complex data. Three participants (B1, B3, B8) positively evaluated the Overview dashboard when engagement-related metrics were made salient and visually readable, enabling rapid understanding of overall project status. This is particularly important for busy practitioners who cannot dedicate extensive time to detailed analysis.

> "The visualization dashboard is good because you can see at a glance what is effective." *[B3]*

> "When engagement was visualized, it was easy to understand." *[B1]*

Second, the Statistical Analysis tab appeared to facilitate the discovery of hidden relationships that practitioners had not anticipated through intuition alone. Two participants explicitly reported such discoveries (B3, B4), while others noted that objective, data-based outputs were useful for correcting prior assumptions.

> "I noticed things only after seeing the data, such as the relationship between comment counts and gender." *[B3]*

> "When you look at the data, you can see that results differ by age group." *[B4]*

Usability Challenges and Requested Extensions Despite positive evaluations, several usability barriers remained. Five of the 11 participants highlighted difficulties stemming from statistical terminology (e.g., p-value, normality) and configuration complexity (B1, B5, B7, B8, B9). While some expected the interface to become easier with familiarity, these barriers pose challenges for initial adoption by non-experts.

> "The terminology is too academic; it is hard to understand without statistical knowledge." *[B5]*

> "What does $n = 5$ mean? What is a p-value?" *[B9]*

Participants also requested that the system provide interpretations alongside results, not merely statistical outputs. Two participants (B7, B8) emphasized that, for time-constrained practitioners, concise trend summaries were preferred over detailed analyses that require extensive interpretation effort.

> "For busy planners, it is important to present concise trend results rather than requiring time-consuming analysis." *[B7]*

Finally, at least four participants suggested extensions (B3, B4, B5, B6): finer-grained behavioral segmentation by demographics, integration of qualitative survey comments for richer client reporting, and metaverse-specific benchmark metrics for cross-project comparisons.

> "I want a feature to obtain detailed behaviors by age and gender (e.g., what resonated with men in their 40s vs women in their 20s)." *[B5]*

> "If qualitative comments can be linked, I can communicate the findings to the client." *[B3]*

5 Discussion

5.1 Implications for Metaverse Branding Analytics

Unlike physical-world marketing, where behavioral data collection is often constrained by sensor placement and privacy concerns, metaverse environments enable the continuous capture of spatial and embodied user behaviors over extended periods. This capability requires new KPI frameworks that reflect the unique spatiality and embodiment of metaverse experiences, moving beyond conventional metrics such as dwell time and visit counts.

Our findings also underscore the importance of at-a-glance comprehension for busy practitioners. In particular, clearly presenting three-dimensional spatial information—such as visitor density and movement patterns within virtual worlds—is essential for enabling rapid interpretation of user behavior.

Furthermore, the user study indicated that practitioners rely on their own heuristics when formulating hypotheses. This suggests that analysis tools should not aim to fully automate or replace expert judgment. Rather, their value lies in surfacing hidden relationships and enabling hypothesis verification, thereby augmenting practitioners' existing expertise.

5.2 Limitations and Future Work

Several limitations remain. First, statistical terminology (e.g., *p*-value, normality) posed comprehension barriers. We created a prototype to simplify tasks related to statistics, but the process and terminology may have been difficult for beginners to understand. Future work should incorporate plain-language explanations and support for guided interpretation.

Second, participants expressed a need to link quantitative metrics to qualitative data, such as open-ended survey comments. Integrating AI-based sensemaking features—potentially using large language models—could help practitioners extract insights from large-scale, heterogeneous datasets.

Third, our system collects detailed behavioral logs, raising concerns about user privacy and the ethical use of data. Future development must incorporate transparent consent mechanisms and data minimization principles [9].

Finally, our evaluation was limited to a single platform (Cluster) with a small sample size (N=11). Validation across other metaverse platforms and with larger, more diverse practitioner populations is needed. Additionally, longitudinal studies are required to assess how practitioners' analysis practices and tool usage evolve over time.

6 Conclusion

We proposed "Verselyzer," an accessible interface for non-experts to visualize and support the analysis of branding effectiveness in the metaverse. Based on design requirements derived from a formative study, we developed a system that comprehensively supports hypothesis formulation, visualization, and statistical analysis. Through user studies, we found preliminary evidence of the system's effectiveness for hypothesis verification and its operability, while also identifying remaining challenges in understanding statistical terminology and analysis settings. Future work will focus on further improving clarity and enhancing automatic interpretation and explanation features to support a broader range of practitioners.

Acknowledgments

This research was supported by the JST Moonshot R&D Program Grant Number JPMJMS2013, JST ASPIRE Program Grant Number JPMJAP2327, and JST PRESTO Grant Number JPMJPR2529, Japan.

References

[1] A. Bangor, P. Kortum, and J. Miller. Determining what individual sus scores mean: adding an adjective rating scale. *J. Usability Studies*, 4(3):114–123, May 2009. 3

[2] C. Bauer and C. Strauss. Location-based advertising on mobile devices: A literature review and analysis. *Manag. Rev. Q.*, 66(3):159–194, June 2016. 1

[3] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996. 3

[4] cluster, Inc. cluster. 1

[5] P. Ebel, K. J. Gülle, C. Lingenfelder, and A. Vogelsang. Iceboat: An interactive user behavior analysis tool for automotive user interfaces. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22 Adjunct. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3526114.3558739 2

[6] 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*, CHI '25. Association for Computing Machinery, New York, NY, USA, 2025. doi: 10.1145/3706598.3714170 1

[7] Y.-H. Hu, S. Yokoi, Y. Hatada, Y. Hiroi, T. Narumi, and T. Hiraki. Luida: A large-scale unified infrastructure for digital assessments in social virtual reality. *arXiv preprint arXiv:2504.17705*, 2025. 1

[8] D. G. Inc. Dentsu group inc. proposes "brand immersive time," an index for measuring the marketing effectiveness of 3d spatial media, 2024. 1

[9] K. D. Martin and P. E. Murphy. The role of data privacy in marketing. *J. Acad. Mark. Sci.*, 45(2):135–155, Mar. 2017. 4

[10] D. Parmenter. *Key Performance Indicators: Developing, Implementing, and Using Winning KPIs*. John Wiley & Sons, Hoboken, NJ, USA, 4 ed., 2019. doi: 10.1002/9781119620785 1

[11] R. Payal, N. Sharma, and Y. K. Dwivedi. Unlocking the impact of brand engagement in the metaverse on real-world purchase intentions: Analyzing pre-adoption behavior in a futuristic technology platform. *Electronic Commerce Research and Applications*, 65:101381, 2024. doi: 10.1016/j.elerap.2024.101381 1

[12] P. I. Philippopoulos, I. C. Drivas, N. D. Tselikas, K. N. Koutrakis, E. Melidi, and D. Kouis. A holistic approach for enhancing museum performance and visitor experience. *Sensors (Basel)*, 24(3):966, Feb. 2024. 1

[13] 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*, Volume 3 - 2022, 2022. 1

[14] VRChat, Inc. Vrchat. 1

[15] J. Wang, M. AlKadi, and B. Bach. Show me my users: A dashboard visualizing user interaction logs. In *2023 IEEE Visualization and Visual Analytics (VIS)*, p. 156–160. IEEE, Oct. 2023. 2

[16] Yellow Dog Man Studios s.r.o. Resonite. 1