Tracking Technologies in Virtual Reality

Wentao Li\textsuperscript{1}, John Feng\textsuperscript{1}, and Jianfeng Wang\textsuperscript{1}

\textsuperscript{1}University of Miami

March 11, 2024
Tracking Technologies in Virtual Reality

Wentao Li, John Feng, Jianfeng Wang

University of Miami

Abstract

In the realm of virtual reality (VR), immersive experiences are enhanced by sophisticated tracking and inference technologies that meticulously monitor user movements, interactions, and behaviors. While these advancements offer unprecedented levels of engagement and personalization, they also raise significant privacy concerns. This study explores the dual-edged nature of tracking and inference in VR, examining how these technologies influence user perceptions of privacy, trust, and comfort within virtual environments. Through a mixed-methods approach that combines quantitative surveys with qualitative interviews, this research delves into the attitudes of VR users towards the data collection practices embedded within these digital realms. The study investigates the awareness levels among users regarding the extent of tracking and inference, their reactions to such practices, and the potential impact on their continued use of VR technologies.

Findings reveal a nuanced landscape of user perceptions, where excitement for technological innovation often coexists with apprehension about privacy intrusions. The research identifies key factors that influence user attitudes, including the transparency of data collection practices, the perceived benefits of personalized experiences, and the availability of control mechanisms to manage privacy preferences. This study contributes to the ongoing discourse on digital privacy by highlighting the specific challenges and considerations unique to VR environments. It calls for a balanced approach to technological development, advocating for the implementation of ethical standards and user-centric privacy safeguards that ensure the responsible use of tracking and inference technologies. By navigating the complex interplay between innovation and privacy, this research underscores the importance of fostering trust and ensuring user protection in the rapidly evolving landscape of virtual reality.

1 Introduction

Virtual reality (VR) applications introduce users to potentially new forms of data collection, such as gaze tracking and extensive motion and position sensing. With VR’s growing popularity, it’s anticipated that companies will leverage this data to deduce information about users, similar to existing practices on the Internet. However, due to the relatively fresh nature of these tracking technologies, VR users might not be fully aware of or understand their implications. We developed and conducted a preliminary user study aimed at exploring two main questions:

1. How do individuals perceive tracking and inference within VR based on their personal experiences? Does the acceptability of these practices vary depending on the context?
2. How does the knowledge of potential tracking and inference influence individuals’ attitudes towards VR, including their comfort level with participating in VR activities?

In the remainder of this report, we will briefly review literature related to our topic; outline our study methodology, including how we designed a VR setting to introduce participants to various tracking techniques; present some intriguing preliminary findings from two initial studies; and describe our plans to refine our approach ahead of a comprehensive user study.

2 Background

We first present background information on tracking and inference in VR, before reviewing related studies on user perceptions of tracking and inference both in VR and in other contexts.
2.1 Tracking and inference in VR

The process of reasoning naturally hinges on representations of relationships between entities or concepts. Such relationships do not explicitly arise in the early-stage language models (BERT [Devlin et al., 2018], RoBERTa [Liu et al., 2019], etc.), but can be simply captured by graph-structured data, such as knowledge graphs. Thus, the early attempts in this field of tracking, VR is also immersive in a way that is from most other applications. While a fitness tracker may collect a user's physiological responses, such as heart rate, a VR application can know exactly what is in the user's environment that might prompt certain responses, and it can even alter the environment to test those responses. This raises additional concerns for user privacy in VR.

As a type of structured knowledge base, Knowledge Graphs (KGs) are models that store detailed factual information. They can improve LLMs with external knowledge for inference and clarity. However, it is often difficult to build and constantly update KGs to represent unseen knowledge. Thus, combining implicit knowledge from LLMs and explicit knowledge from KGs can better leverage their strengths in solving knowledge-intensive tasks. In this section, we review recent advances from two aspects: 1) LLM-aided KG Reasoning; and 2) KG-enhanced Language Generation.

2.2 User perceptions of tracking

Similar to prior work, limited previous research on user perceptions suggests that innovative methods of tracking and inference in virtual reality (VR) may not be foremost in users’ minds. In a study by Adams et al. in 2018, VR users primarily expressed concerns about data collection from microphones and infrared sensors, with some anticipating increased privacy issues in the future [1]. Similarly, in 2015, MoR et al. analyzed online comments and found minimal worries about privacy related to VR headsets [10]. Our study aims to make privacy concerns more concrete for current users to explore their reactions and perceptions. Outside the VR realm, other studies have employed a similar strategy of demonstrating tracking methods to participants to gauge their responses. For instance, Weinshel et al. developed a browser extension that simulated third-party trackers and visualized tracking data, which, in a longitudinal field study, helped participants comprehend tracking behaviors better and increased their intention to take privacy-preserving actions [18]. Wei et al. exhibited Twitter ad targeting data to participants, identifying practices that felt intrusive and suggesting improvements to ad targeting explanations [17]. Following a comparable approach, we simulate tracking in VR for our participants and pose inquiries to comprehend their perceptions.

3 Method

In this section, we outline our process of creating a VR application to gather data from participants. Subsequently, we detail our study methodology, wherein participants engaged within this VR environment and subsequently provided feedback on their experiences and perceptions. It's important to highlight that the main objective of the VR activity wasn’t to precisely infer participant characteristics but rather to realistically simulate a company’s inference-making process. The aim was to immerse participants sufficiently in the experience to encourage exploration of new perspectives.

3.1 Tracking participants in a VR application

As the backbone of our user study, we created a VR:

1. For each object category on which we tracked a user's gaze, we deduced their product interest by analyzing the duration of their focus.
2. Some of the tracked objects were linked to either liberal or conservative political inclinations, as indicated by prior research [3]. By assessing the average time users spent observing objects associated with each side, we made a simplistic inference regarding their political alignment—a trait relevant for advertising purposes, distinct from product interest.
3. By comparing a user's reaction time to a hypothetical baseline, we inferred whether their cognitive processing speed—a trait relevant to health—was below, average, or above average.

The integration of Large Language Models (LLMs) in drug discovery offers a promising path to accelerate various aspects of the process. One critical challenge is representing molecular graphs, which lack the inherent sequential structure of text data. Converting these structures into a format amenable to LLMs is a significant challenge. LLMs can facilitate drug discovery by providing insights into potential side effects, predicting drug interactions, understanding molecular structure-activity relationships, guiding
optimization of lead compounds, supporting drug repurposing, and streamlining clinical trials.

The paper introduces a prototype system called DrugChat, inspired by ChatGPT, tailored for drug molecular graphs. It consists of a graph neural network (GNN), an LLM, and an adaptor. The GNN processes molecular graphs, and the adaptor transforms the graph embeddings into prompts for the LLM. This system is trained on thousands of molecular graphs and question-answer pairs, enabling users to interactively query about compounds. Data is sourced from Chambly and PubChem, with prompts following a specific template format. These approaches collectively represent the dynamic landscape of incorporating LLMs in drug discovery, addressing challenges in different ways and offering valuable insights and applications.

3.2 User study design

Our user study comprises three main parts, summarized in Table 1. Firstly, participants engaged in our VR environment for approximately 10 minutes, during which their activities were automatically recorded in the background. We instructed participants to envision themselves entering a vast virtual world created for leisure and social interaction by a company named Total VR. Clarifying that the specific space within our VR application was crafted by Total VR to facilitate users’ acclimation to VR navigation, we directed them to engage in the cat minigame while also acquainting themselves with the virtual environment and its controls.

Subsequently, upon exiting the VR environment, participants reviewed up to five instances of tracking and inference with us, providing feedback on their level of surprise and perceptions of fairness. We inquired about their stance on Total VR retaining these data and inferences for personalizing advertisements and content suggestions, as well as whether they found it acceptable for Total VR to sell this information to other companies, potentially integrating it with other data to construct more comprehensive marketing profiles.

Finally, participants underwent a semi-structured interview concerning their broader perceptions of VR. This included inquiries about their comfort level in utilizing VR social media platforms as spaces for uninhibited expression and authenticity, as well as their preferences regarding features or settings akin to private browsing mode in VR, if any.

At the study’s commencement, participants completed a questionnaire detailing demographic information and their VR and social media experiences. At the study’s conclusion, they responded to an 11-item true-or-false questionnaire concerning online tracking and inference in the U.S., drawn from a survey conducted by Turow et al. [16]. Participation was voluntary, with participants assured of future baked goods as a token of appreciation.

4 Findings

We conducted two pilot studies, which we present here as the process of reasoning naturally hinges on representations of relationships between entities or concepts. Such relationships do not explicitly arise in the early-stage language models but can be simply captured by graph-structured data, such as knowledge graphs. Thus the early attempts in this line of research mainly focus on combining the structural information in knowledge graphs and the semantic representation in language models for language tasks involving reasoning, such as question answering. With the advent of LLMs, recent progress has shown that LLMs are powered with reasoning capabilities and still yield improvements. Therefore, in addition to coalescing LLMs with graph-based approaches, attention has also been gathered on using the reasoning ability of LLMs for prompt augmentation to support downstream tasks. We give more details of related work as below.

To address the challenge of reasoning in question answering tasks to proposes QA-GNN where the QA context is extracted into a knowledge graph (KG), and the language model is integrated to estimate the relevance score of each entity, which later becomes an additional feature of the KG nodes. To perform joint reasoning, it connects an extra node representing the QA context to each entity in the KG and employ common GNN architectures () for inference or learning. Later extends the approach for the same task with more involved architecture design to better fuse the GNN and LM representations.

4.1 RQ1: perceptions of tracking and inference

Either participant expressed surprise regarding the extent of tracking and inference demonstrated in the study. Interestingly, while Participant 1 (P1) acknowledged the possibility of companies gathering similar data from website
usage, they found the VR tracking to be more invasive, noting that browsing is a more unconscious activity compared to intentionally navigating a webpage. They remarked, "It was really interesting to just see that data collected and charted in front of me. That definitely felt different from just conceptually knowing that this is a thing that could be done.

4.2 RQ2: effect on antecedents towards VR

Both participants indicated their use of privacy-enhancing tools beyond VR, such as private browsing mode, and expressed a desire for similar tools within the VR environment. Participant 2 (P2) remarked, "I don’t feel like it makes much sense to draw much information from [what I was absent-mindedly gazing at on the TV], but hey, maybe psychologists would disagree." However, they also recognized the validity of certain inferences, noting, "I remember I did distinctly pause to read those, and I kind of did glaze over the [other cookbook]... But I did make a point to read the title of the spicy food one." P2 generally perceived most inferences as inaccurate due to the limited data collected during the study.

Regarding the acceptability of tracking and inference practices, Participant 1 (P1) regarded such actions as "quite invasive" when companies utilize or sell this data. Conversely, P2 expressed a nuanced perspective, indicating that using data could be acceptable if disclosed in a privacy policy and important that such a tool doesn’t interfere with the “cool” aspects of VR that allow users to see each other’s body language and make eye contact. It introduces a framework for recommendation system, where a personalized reasoning graphs is obtained from an adaptive reasoning procedure powered by LLMs. The personalized reasoning graph is combined with a sequential recommendation model to predict candidate items for users.

5 Discussion

5.1 Changes to VR application

Gaze tracking stands out as one of the most intriguing “new” tracking methods in VR for us. While our initial implementation in the first pilot study was somewhat convincing, we aim to enhance accuracy by utilizing a headset equipped with eye tracking technology.

Participant 2 (P2) demonstrated that participants can easily avoid tracking while in the VR environment, posing a challenge for reflecting on plausible inferences in the subsequent phase of the user study. To address this, we plan to increase tracking elements, such as incorporating more tracked objects and diverse tracking methods into the environment. Additionally, we may consider redesigning the activity; for instance, rather than having participants search for a cat, they could engage with pre-programmed avatars representing other users in a simulated social VR world. This approach would closely mimic real social VR interactions and potentially yield more meaningful data. We also acknowledge P2’s skepticism regarding the fairness of comparing users’ reaction times when sound cues were played. To address this concern, we may explore alternative
health-related inferences, such as height or posture. Furthermore, both participants experienced motion sickness, partly due to joystick-based controls. Introducing different locomotion methods could help alleviate discomfort. To address the challenge of reasoning in question answering tasks proposes QA-GNN where the QA context is extracted into a knowledge graph (KG), and the language model is integrated to estimate the relevance score of each entity, which later becomes an additional feature of the KG nodes. To perform joint reasoning, it connects an extra node representing the QA context to each entity in the KG and employ common GNN architectures for inference or learning. Later, extends the approach for the same task with more involved architecture design to better fuse the GNN and LM representations, improving upon a single pooling layer in the architecture of regarding changes to the user study protocol, we observed during pilot studies that some questions require refinement. For instance, the term "fair" needs clarification in questions like "How fair do you think this inference is, based on your activity?" Additionally, the true-or-false questions at the end of the study were originally designed for a survey targeted at Americans and contain U.S.-specific statements. P2, who is not from the U.S., pointed out the difficulty in answering some questions due to this bias. Hence, we will select questions that are less influenced by nationality to ensure a fair assessment of knowledge across all participants.

5.2 Changes to user study protocol
During our pilot studies, we identified areas where certain questions require refinement. For instance, the term "fair" in the question "How fair do you think this inference is, based on your activity?" lacks clear definition and may lead to ambiguity in responses.

Additionally, the true-or-false questions administered at the end of the study were originally tailored for assessing American knowledge. Some questions include U.S.-specific statements, such as "The Health Insurance Portability and Accountability Act (HIPAA) prevents apps that provide information about health from selling data collected about app users to marketers." This renders the assessment less suitable for non-American participants, as noted by P2, who faced difficulty answering some questions due to this bias. Moving forward, we will select questions that are less influenced by nationality to ensure a fair assessment of knowledge across all participants.

References


[8] Liu, Fuxiao, Tianrui Guan, Zongxia Li, Lichang Chen, Yaser Yacoob, Dinesh Manocha, and Tianyi Zhou. "Hallusionbench: You see what you think? or you think what you see? an image-context reasoning benchmark challenging for gpt-4v (ision), llava-1.5,


