Poster: PeekXR: Understanding Privacy Leakages from Eye Gaze in Extended Reality

Chuyang Peng
University of California, Irvine
Irvine, CA, USA
chuyangp@uci.edu

Mutahar Ali
University of California, Irvine
Irvine, CA, USA
mutahara@uci.edu

Habiba Farrukh
University of California, Irvine
Irvine, CA, USA
habibaf@uci.edu

ABSTRACT

Extended Reality (XR) headsets are increasingly integrating eye tracking for enhanced user experience, adaptive interfaces, and foveated rendering. However, this rich biometric signal introduces new privacy risks. In this work, we demonstrate that eye-tracking data collected by commercial XR devices can be exploited to infer sensitive user activity. We leverage users' gaze sequences captured while interacting with a VR app to classify the type of content a user is watching. Our results reveal that eye movements alone, without any video or audio context, carry enough information to accurately predict content categories. We discuss the implications of this threat and outline how eye tracking can potentially be used to fingerprint applications and user behavior. This work is a step towards exposing and mitigating emerging privacy threats in immersive systems.

KEYWORDS

Extended Reality, Eye Tracking, Privacy, Gaze Inference

ACM Reference Format:
Chuyang Peng, Mutahar Ali, and Habiba Farrukh. 2025. Poster: PeekXR: Understanding Privacy Leakages from Eye Gaze in Extended Reality. In *The 23rd Annual International Conference on Mobile Systems, Applications and Services (MobiSys '25), June 23–27, 2025, Anaheim, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3711875.3734567

1 INTRODUCTION

Extended Reality (XR) headsets, including augmented (AR), virtual (VR), and mixed reality (MR) headsets, are becoming increasingly immersive, and eye tracking is quickly emerging as a core sensing modality in next-generation devices. Modern XR headsets, such as the Meta Quest Pro and Apple Vision Pro, incorporate high-frequency eye trackers capable of capturing precise gaze direction, fixation points, and pupil dilation [5]. This technology powers a wide range of compelling features, including foveated rendering, which reduces GPU load by rendering only where the user is looking in high detail, gaze-based interfaces, which enable intuitive control, and attention analytics, which can inform adaptive content or training applications. As eye tracking becomes a standard part of the XR experience, so does the quantity and granularity of gaze data collected about users [1].

MobiSys '25, June 23–27, 2025, Anaheim, CA, USA

ACM ISBN 979-8-4007-1453-5/2025/06
https://doi.org/10.1145/3711875.3734567

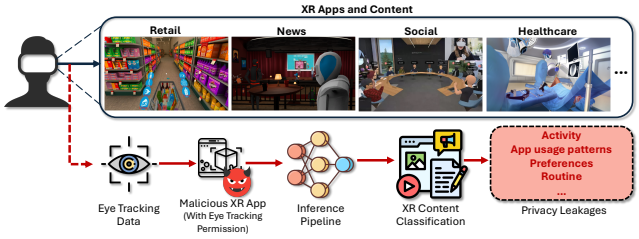

Figure 1: Overview of PeekXR: A benign-looking XR application with access to eye tracking data (e.g., for avatar animation) can repurpose this data for inferring sensitive user information. The collected gaze signals are fed into the inference pipeline to classify XR content (e.g., video type, app usage), enabling adversaries to infer sensitive user attributes such as preferences, routines, or even identities.

While eye tracking is designed to enhance user experience, it also introduces profound and underexplored privacy risks. Unlike input methods like touch or voice, gaze is often subconscious and continuous, revealing not just what users are looking at but also how they think, feel, and behave. Prior research has linked gaze patterns to users' keystrokes [3], emotional states, cognitive load, personality traits [2], and even medical conditions [7]. Recent work has also shown that gaze can be leveraged to infer PINs and passwords on virtual keyboards [6], further emphasizing the privacy risks posed by gaze-based interfaces. However, in this work, we ask a more system-centric question: Can eye gaze data alone be used to infer what a user is doing in an immersive environment, even when content is hidden from the adversary? This leads us to investigate eye movements as an attack surface for inferring user activity in immersive environments.

2 PEEKXR APPROACH

Figure 1 presents an overview of our *PeekXR*. The goal of *PeekXR* is to assess the feasibility of inferring sensitive user activity in XR environments using only eye-tracking data. As an initial step, we focus on a specific task: classifying the type of video content a user is watching in XR, without any access to screen content, audio, user interaction (e.g., touch/controllor input), and application metadata.

To achieve this, we build an inference pipeline that collects gaze signals from a commercial XR headset (i.e., Meta Quest Pro), processes these signals into structured temporal sequences, and trains deep learning models to infer user activity. The pipeline simulates a realistic adversarial scenario, assuming access to only the gaze stream and nothing else about users' activity and interactions.

Threat Model. We assume an adversary with access to real-time or logged eye gaze data from an XR headset. This adversary could be a seemingly benign application or plugin that requests gaze data for a legitimate purpose, such as enabling avatar eye movement in a social XR setting.

Modern XR platforms (e.g., Meta Quest, Apple Vision Pro) allow third-party apps to request user permissions to access eye-tracking data. Users often grant this, especially in contexts where gaze is used to animate avatars or support natural interactions. However, once permission is granted, the app may gain continuous access to high-frequency gaze data streams with minimal system-level restrictions on processing or storing them. In most XR platforms, an app running in the background cannot access any information about users' activity (e.g., what content they are watching, what actions they are taking) in the foreground app.

Therefore, we assume the adversary cannot access video content, user input, screen buffers, or audio. Instead, their goal is to infer private aspects of the user's behavior—such as what type of content they are consuming or which application is in use and what users' actions are purely from gaze dynamics. This represents a realistic, low-permission attack surface in today's XR ecosystems.

Feature Extraction. We first normalize and synchronize gaze data streams across users and sessions. Each session is segmented into fixed-length windows (e.g., 10-second intervals) to allow temporal analysis. From each window, we extract 3D gaze vectors, angular velocities, and fixation durations. These segments are encoded into a fixed-size representation and fed as sequences to the learning models. To reduce overfitting and promote generalization, we augment data with Gaussian noise and apply random dropout during model training.

Inference Pipeline. We collect and analyze gaze data from users watching videos within a XR environment and evaluate whether it's possible to infer the type of content being viewed using only eye movement patterns. To model the temporal dynamics of gaze, we employ a variety of deep learning models, including convolutional neural networks (CNN), attention-based Long-Short-Term Memory (LSTM) networks, and transformer-based architectures. These models are trained to classify video categories (e.g., gaming, documentary, adult content) based purely on sequences of gaze vectors (position and orientation), fixations, and saccadic behavior over time. Importantly, the models operate without any visual or audio data, emulating a restricted threat model where only eye-tracking streams are available to the adversary.

3 FEASIBILITY STUDY

We evaluate our approach on a dataset consisting of eye gaze data collected using a Meta Quest Pro headset equipped with integrated eye tracking, with a sampling rate of 72Hz, while watching VR video clips across seven genres: fashion, gaming, music, news, podcasts, movies and sports. For each genre, we sample 20 different immersive videos from Youtube VR and split them into training and test sets.

Effectiveness. We report classification accuracy on held-out gaze data from unseen video samples (20% of the dataset) using both LSTM and transformer models. Our inference pipeline achieves an average classification accuracy of 75%, indicating that eye gaze patterns contain meaningful signals that can be leveraged for content inference. Both models performed well, even with short (4s) sequences of gaze data. Attention maps suggest that the models leverage gaze shift patterns and fixation density differences across genres. We also evaluate different eye gaze data sequences lengths and observe that our inference pipeline works well even with small sample lengths.

4 FUTURE WORK AND CONCLUSIONS

Our immediate next step is to expand beyond video classification and explore app fingerprinting and user activity inference for data collected from a larger number of XR users. We hypothesize that different applications induce distinct gaze signatures due to UI layout, task types, and interactivity levels. We will also explore how different users have unique gaze patterns, even for the same app, and adapt our inference pipeline for arbitrary users. Furthermore, by incorporating additional sensor data (e.g., head movement, hand tracking), we aim to explore multimodal inference attacks and determine how privacy-sensitive information can be extracted.

We also plan to explore defenses to protect eye gaze data in XR. Prior works [4] have explored privacy-preserving methods (e.g., differential privacy) to add controlled noise to gaze data before making it accessible to applications. However, such noise addition also impacts the utility of eye gaze data for XR applications (e.g., cursor movement and content selection). Future work will explore eye gaze obfuscation techniques to balance utility and privacy tradeoffs for XR applications.

This work exposes a novel and practical privacy risk in XR systems: inferring sensitive user activity from eye-tracking data alone. We show that even without contextual content, gaze patterns leak rich semantic information. As eye tracking becomes ubiquitous in immersive computing, there is an urgent need for system-level safeguards. Our findings urge the mobile and immersive systems community to treat gaze data as a first-class privacy concern, similar to camera, GPS and microphone data.

REFERENCES

[1] Samantha Aziz and et al. 2024. Evaluation of Eye Tracking Signal Quality for Virtual Reality Applications: A Case Study in the Meta Quest Pro. In *Proceedings of the Symposium on Eye Tracking Research and Applications*.
[2] Shlomo Berkovsky, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. 2019. Detecting personality traits using eye-tracking data. In *Proceedings of the conference on human factors in computing systems (CHI)*.
[3] Yimin Chen, Tao Li, Rui Zhang, Yanchao Zhang, and Terri Hedgpeth. 2018. Eyetell: Video-assisted touchscreen keystroke inference from eye movements. In *IEEE Symposium on Security and Privacy (SP)*.
[4] Jingjie Li and et al. 2021. kaleido: Real-Time Privacy Control for Eye-Tracking Systems. In *USENIX Security Symposium*.
[5] Meta. 2025. *Eye Tracking on Meta Quest Pro.* Retrieved February 24, 2025 from https://www.meta.com/help/quest/8107387169303764/
[6] Hanqiu Wang, Zihao Zhan, Haoqi Shan, Siqi Dai, Maximilian Panoff, and Shuo Wang. 2024. GAZEploit: Remote Keystroke Inference Attack by Gaze Estimation from Avatar Views in VR/MR Devices. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*.
[7] Sheng Wang, Xi Ouyang, Tianming Liu, Qian Wang, and Dinggang Shen. 2022. Follow my eye: Using gaze to supervise computer-aided diagnosis. *IEEE Transactions on Medical Imaging* (2022).