Exploring the Impact of LLM-Powered Virtual Spaces on Privacy

Marina Weber
marina.weber@tum.de

Abstract—**Virtual reality (VR) systems increasingly integrate eye tracking as a standard feature, enabling both enhanced interaction and user analytics. However, the same data that supports these innovations may also compromise user privacy. This project investigates the extent to which personal attributes—such as age, gender, ethnicity, and body weight—can be inferred from eye tracking data alone in immersive VR environments. We developed a Unity-based XR application with four interactive task rooms designed to elicit natural gaze and as well as other behavioral patterns. In total, sixty-nine participants completed the experiment using the Varjo XR-3 headset. We found measurable correlations between gaze behavior, task performance (e.g., object ranking order and food selection), and participant demographics, supporting the hypothesis that eye tracking data can reveal sensitive personal information. These findings highlight significant privacy implications for future XR systems that integrate eye tracking as a default sensing modality.**

Index Terms—**extended reality, virtual reality, eye tracking, human behavior, unity, privacy, data protection, security**

I. Introduction

Virtual Reality (VR) technologies have rapidly evolved over the last decade and are now widely used in gaming, healthcare, education, and research. The immersive nature of VR makes it a powerful medium for studying human perception, attention, and behavior in controlled yet realistic environments. Recent advances in consumer-grade hardware have made it possible to integrate precise sensors, such as eye trackers, directly into VR headsets. As a result, large-scale eye tracking has become feasible beyond specialized laboratory settings.

While these technological developments open exciting opportunities, they also introduce new challenges regarding user privacy. Eye movements can reveal a wide range of personal information, including cognitive processes, emotional states, and even demographic traits. As eye tracking becomes an integral part of everyday devices, understanding what kind of personal data can be inferred—knowingly or unknowingly—from gaze behavior is becoming increasingly important.

This paper describes a Unity project and an accompanying user study conducted as part of an Interdisciplinary Project at the Technical University of Munich (TUM). The main goal of this project to test whether it is possible to predict personal information such as age, gender, or weight from a person's eye tracking data, behavioral patterns, and stated preferences. For this purpose, we created a simple Unity-based VR application consisting of several task-oriented virtual rooms. Each room presents a short task, and participants perform these tasks while their gaze data is recorded. The collected data is then analyzed to investigate which gaze patterns correlate with different personal attributes.

A. Motivation

With the increasing availability of consumer-grade eye tracking in devices such as VR headsets and AR glasses, the boundary between technological innovation and personal privacy is becoming blurred. Our motivation is to explore how much can be inferred about a user when their gaze data is available as well as how voice, facial, or motion data influence our results.

If even simple gaze features allow for the prediction of sensitive traits like gender, age, or weight, the implications for future digital environments are profound. Such findings emphasize the need for privacy-aware design principles in immersive systems. As VR and AR become ubiquitous, ensuring that users maintain control over the information implicitly conveyed through their gaze will be essential for building trustworthy and ethically sound virtual experiences.

B. Experimental Design

To test our hypothesis, we designed a series of virtual rooms inside a Unity application. Each room placed participants in a different everyday context with minimal distractions. Tasks included viewing food items, identifying objects, or observing spatial arrangements. These scenes allowed us to record gaze data and analyze various metrics. We specifically focused on four attributes to guess: age, gender, ethnic background, weight (overweight-healthy).

C. Privacy and Ethical Considerations

The results of this experiment are relevant far beyond academic curiosity. As eye tracking becomes an integral part of the metaverse, adaptive interfaces, and real-time advertising, the silent collection of gaze data may lead to user profiling without consent. If companies can infer sensitive attributes just from eye movements, it becomes crucial to define ethical boundaries, technical safeguards, and legal protections. Our work demonstrates how even minimal eye tracking data can become a vector for privacy invasion if not handled responsibly.

D. Contribution

This project contributes to the ongoing discourse on privacy and the ethical use of biometric data in human–computer interaction by demonstrating, through a working VR prototype, that eye tracking data alone can reveal sensitive personal traits. Our contributions are both technical and empirical:

- **Technical Contribution:** We developed a modular Unity-based VR application integrated with the Varjo XR-3 headset and its official eye tracking SDK. The system records real-time gaze data, including fixation points, gaze targets, and object interactions, and links these to contextual task metadata. A speech interface connects to OpenAI and Amazon Web Service (AWS) Polly APIs, enabling conversational interactions with AI-driven NPCs. This setup represents a reproducible framework for gaze-based behavioral experiments in extended reality.
- **Experimental Contribution:** We designed four task-oriented VR rooms—each probing a different personal attribute (gender, age, weight, and ethnicity)—to isolate the predictive power of the collected data. The tasks were specifically structured to elicit natural eye movements under controlled conditions, allowing statistical inference between behavior and participant attributes.

II. Related Work

Eye tracking has evolved from a niche research tool into a powerful data source used in consumer electronics, medicine, gaming, and marketing. As sensors become cheaper and more accurate, the privacy implications of eye tracking are becoming increasingly urgent. Several studies have demonstrated that gaze data—though seemingly innocuous—can reveal a wide array of personal information about users, including biometric identity, mental state, demographic group, and even subconscious preferences [1]–[3]. Empirical research further shows that gaze features can predict specific user attributes such as gender [4], age [5], and body weight or BMI [6]. Additionally, gaze-based indicators have been applied in the detection of mental health conditions such as anxiety, depression, and ADHD [1]. These findings underline the dual nature of gaze data: while useful for interaction and adaptation, it simultaneously exposes highly personal information that users may not intend to share.

A. Gaze as a Source of Biometric and Demographic Information

Kröger et al. [1] provide an extensive overview of personal information that can be inferred from eye tracking data. Their review categorizes gaze-based inferences into areas such as biometric identity, age, gender, ethnicity, health conditions, personality traits, and emotional states. For example, specific saccade patterns, fixation durations, and pupil dynamics can uniquely identify individuals or distinguish demographic groups. Age influences oculomotor behavior, with older individuals exhibiting slower saccades and altered fixation distributions, while gender differences have been reported in how participants attend to emotional or visually complex stimuli [4].

B. The Uniqueness and Sensitivity of Gaze Data

Liebling and Preibusch [2] argue that gaze data is fundamentally different from other behavioral inputs such as mouse or touch interactions. Because gaze is only partially under voluntary control, it can inadvertently reveal personal traits such as emotional arousal, sexual preference, or cognitive workload. This involuntary nature makes gaze data particularly sensitive and difficult to anonymize, as even micro-saccades or pupil dilation patterns can act as unique behavioral signatures.

C. Gaze-Based Inference of Preferences and Health Conditions

Liebling and Preibusch [2] and Kröger et al. [1] both describe how gaze data can be used to infer preferences, especially through the duration and frequency of fixations. In particular, Kröger et al. cite studies showing that overweight individuals display longer fixation times on high-calorie food images. These types of attentional biases are not only predictive of preferences but can also reveal underlying conditions such as binge-eating disorder or substance dependency.

Furthermore, eye tracking has been successfully applied to detect a range of mental health conditions, including anxiety, depression, and ADHD. Abnormal gaze patterns—such as reduced fixations on faces or delayed response to stimuli—can serve as diagnostic markers. In many cases, these findings have already been implemented in patented technologies or commercial products.

D. Privacy Implications and Ethical Concerns

The collection and analysis of gaze data raise serious ethical and legal concerns. Liebling and Preibusch [2] emphasize that users are often unaware of the extent of information they disclose through their gaze. Similarly, Kröger et al. [1] argue that this lack of awareness is particularly problematic in systems without transparent feedback—such as head-mounted displays or public kiosk-based trackers. Liebling and Preibusch [2] further stress the importance of policy frameworks and technical safeguards, including data minimization, noise injection, and opt-out mechanisms, to protect users from unintended inferences and misuse of gaze data.

Our project builds upon this body of work by demonstrating, through a practical implementation, how gaze data collected in minimal task environments can be used to predict personal attributes such as weight and gender. While prior studies have shown these inferences are theoretically possible [1], our contribution is a concrete and reproducible system that explores these implications within modern interactive environments.

E. Predicting Demographic Traits from Eye Tracking

Beyond biometric identification, multiple studies have shown that demographic traits such as gender, age, and body weight can be inferred with high accuracy from eye movement patterns. For instance, Wang et al. [4] developed a deep learning model that achieved over 80% accuracy in classifying gender from eye-tracking data recorded during reading tasks. Their findings suggest that subtle differences in saccadic amplitudes, fixation durations, and reading dynamics reflect underlying cognitive and perceptual differences between male and female participants.

Age prediction has also been explored through features like saccade velocity, blink rate, and fixation stability. Sugano and Bulling [5] proposed a convolutional neural network model that predicts user age from gaze patterns collected during screen-based viewing tasks. Their results reveal a continuous, rather than categorical, relationship between gaze behavior and age—highlighting the potential of regression-based approaches for age estimation.

With respect to body weight or body mass index (BMI), Graham et al. [6] found that participants with higher BMI scores fixated longer and more frequently on high-calorie food images in a forced-choice paradigm. These attentional biases were linked not only to preference but also to impulsivity and emotional eating tendencies. Such correlations demonstrate how gaze behavior can be leveraged to develop predictive models that infer health-related traits from visual attention alone.

F. Applications and Privacy in VR and the Metaverse

As eye tracking becomes a standard component in consumer-grade VR and AR systems, its applications extend into immersive platforms. Modern headsets such as Meta's Quest Pro and Apple's Vision Pro integrate gaze tracking for foveated rendering, user interface selection, and emotion-responsive avatars [3]. While these features enhance user experience, they simultaneously introduce substantial privacy risks. Steil et al. [3] demonstrated that even short segments of gaze data recorded in VR environments can accurately identify users and reveal personal traits. Their work underscores the importance of implementing transparent consent mechanisms and privacy-by-design principles in future XR systems.

G. Toward Ethical Gaze-Based Systems

Taken together, the existing literature highlights the dual nature of eye tracking: it provides powerful insights into human cognition and behavior, yet it also creates new risks to user privacy and autonomy. Steil et al. [3] and Kröger et al. [1] call for the integration of privacy-preserving measures, such as local data processing, real-time anonymization, and user-controlled calibration procedures that limit what inferences can be drawn. Moreover, several authors advocate treating gaze data as biometric information under privacy regulations such as the GDPR and the California Consumer Privacy Act (CCPA) [1], [2].

Our project contributes to this discussion by providing an empirical demonstration of these privacy risks. By designing an interactive experiment where only gaze data are recorded yet still enabling inference of traits such as gender and weight, we show that eye-tracking data can act as a soft biometric identifier. These findings reinforce the growing need for privacy-by-design strategies in the development of gaze-enabled technologies.

III. Room Description

The Unity project used to conduct the experiment consists of one tutorial room and four task rooms. Each of the four rooms was designed to gather information about specific personal details (such as weight or ethnicity) by analyzing the user's eye tracking as well as other, room specific data. This section covers the rooms and their main purposes within the experiment.

A. Basic Setup of the Rooms

Each room is a plain room with minimal decoration to not distract from the main task. In each room, there is an info wall located and a robot. The info wall, depicted in figure 1, displays the current instructions to the user. The robot, depicted in figure 2, serves as an assistant and further explains what the user needs to do in each room. Furthermore, the user can talk to the robot at any given point to ask questions. The questions can be of any nature, whether the participants need help with a certain task or are just curious about certain aspects of the room. The robot replies using AI-generated answers. A detailed description can be found in section V-E. All the conversations between robot and participant are protocolled for analysis purposes.

Figure 1. The info wall that displays instructions in every room

Furthermore, every room has a blue button in the room. The color blue indicates that the button is currently inactive and has no effect if pressed. Once the tasks in a given room are completed, the button turns green, indicating that it is now active. By touching the button in the application with the controller in VR, the user can advance to the next room.

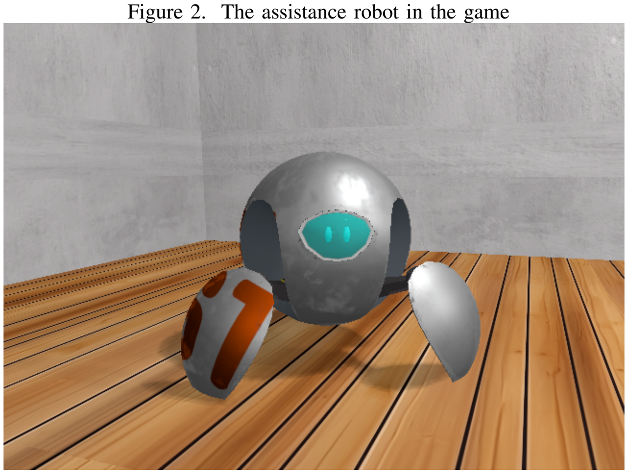


Figure 2. The assistance robot in the game

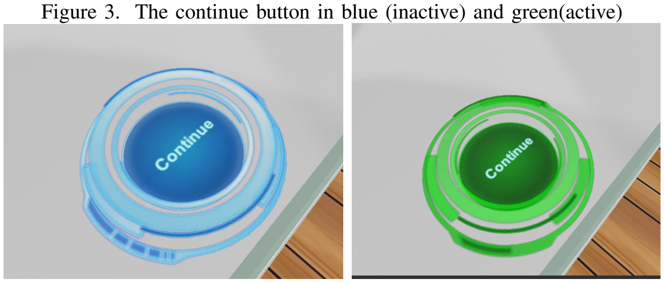


Figure 3. The continue button in blue (inactive) and green(active)

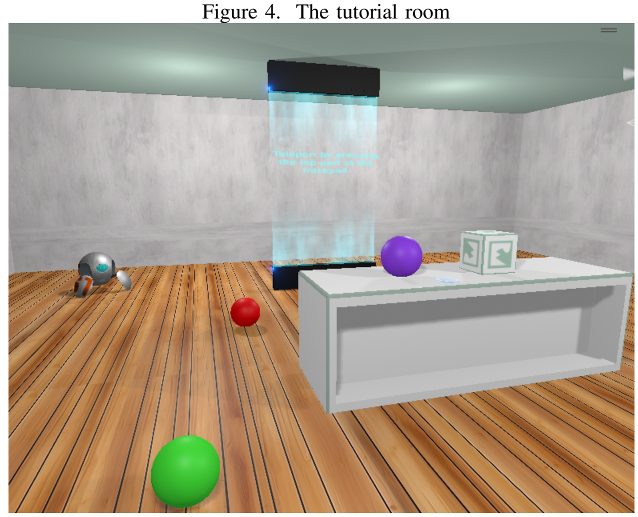


Figure 4. The tutorial room

B. Basic Flow

The participant starts in the tutorial room. In this room, the user learns about the controls in the game and how the general flow of the tasks works. Once the participant completes the tutorial, one of our experiment conductors performs basic eye calibration to ensure accurate recordings of the eye tracking data. Once the calibration is complete, they can continue to the first task room. The order of the remaining task rooms that the participant needs to perform is random. This ensures there is no bias in the experiment where participants are much more experienced in one room than in another. Once they finish all rooms, the application informs the user that they have completed the experiment, and then the application closes itself.

C. Tutorial Room

The tutorial room is used to familiarize participants with the application's controls and task format. It does not contain any specific task that relates to personal data. In this room, the player learns how to teleport within the room using the controller, how to pick up objects, and how to interact with the AI assistance robot that is present in each room. A screenshot of the tutorial room is depicted in figure 4.

D. Gender Room

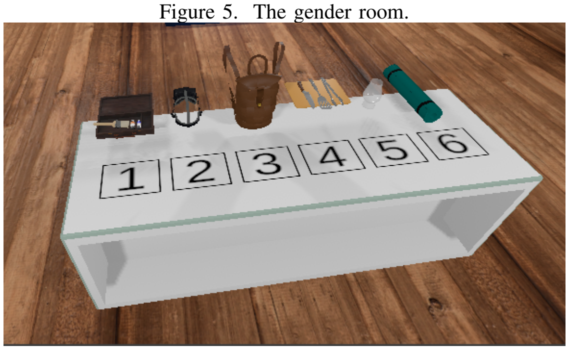
This room was designed to explore whether a participant's gender could be inferred from their gaze behavior as well as stated preferences. The underlying hypothesis was informed by prior research on gaze behavior and gender differences. For instance, Wang et al. [4] demonstrated that gender can be predicted from subtle differences in saccadic amplitudes, fixation durations, and overall scanning strategies. Similarly, studies examining attentional bias in visual decision-making [6] suggest that gaze patterns systematically vary across demographic and behavioral traits, indicating that such differences may generalize beyond specific task contexts.
The task in this room required participants to evaluate six different objects placed on a virtual table and to **rank them from 1 (most interesting) to 6 (least interesting)** by physically placing them in the corresponding numbered positions on the table. This design served two purposes: first, to elicit natural gaze patterns reflecting how participants explore and compare visual stimuli, and second, to investigate whether gaze duration correlates with subjective preference.
For this task, we selected objects that are typically associated with one gender but are not overly obvious and could appeal to both genders. This approach helps prevent participants from guessing the task's purpose. The objects that we chose that are associated with the male gender are:

1) Barbeque (BBQ) equipment: While both genders can enjoy cooking, BBQ is often associated as a dish that men in particular enjoy.
2) Whiskey glass: There are cultural stereotypes that link whiskey consumption with masculinity, toughness, and traditional male social settings such as bars or cigar lounges.
3) VR glasses: Men are typically associated with enjoying technology and video games more than women

The objects that we used that are more associated with the female gender are:

1) Leather backpack: Women are culturally stereotyped to be more into fashion and accessories than men
2) Yoga mat: Yoga and Pilates are sports that are particularly popular with women.
3) Box with acrylic paint: There is a cultural stereotype that women tend to be more interested in the arts.

Figure 5. The gender room.



The six selected objects were chosen to balance recognizability, moderate gender association, and visual diversity. Prior research has shown that gender differences in gaze behavior and interest often appear when stimuli evoke culturally gendered associations but remain subtle enough to avoid explicit cues [1], [4]. Therefore, we selected items such as a yoga mat or BBQ tools as proxies for stereotypically “female” and “male” domains, respectively, while including neutral options such as a paint box or backpack for comparison. This design allows testing whether gaze and ranking differences emerge spontaneously from interest rather than from explicit awareness of the task’s purpose. The combination of slightly gendered and neutral objects thus provides both a manipulation and an internal control condition for evaluating gender-linked attention patterns.

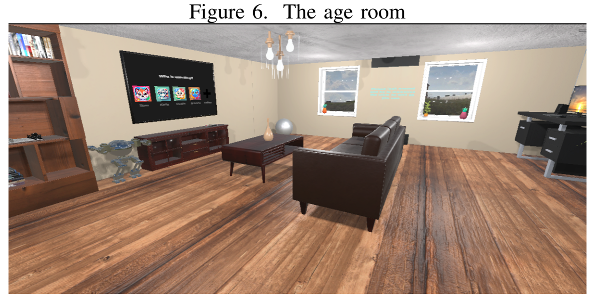
E. Age Room

The age room, as seen in III-E, appears to be a normal living room. The purpose of the room, as the name suggests, is to guess the participant’s age. Our hypothesis is that younger people’s eye movement and reaction time are quicker, and also their short-term memory is better. For this purpose, we decided to create a room where participants are asked to memorize the room as well as they can within a certain time span (set to two minutes). The participants are allowed to freely move around in the room within the given time frame. If the participant thinks they have memorized the room sooner, they can also press a green button in the room to skip the remaining time. Once the time frame is over or the participant chooses to skip the remaining time, the room will be reloaded. In the reloaded room, a number of objects are misplaced. The participants are then asked to move the objects back to their original position. The participants are not given any hints on how many objects have been misplaced or which objects they could potentially move. Once they think they are done, they can continue to the next room by pressing the green button in the room.

The following misplacements were implemented:

- A jacket from the coat hanger was placed on the couch.
- A plant that was originally on the windowsill was moved to a different location.
- The video game console in front of the television was moved to the computer desk.
- The character profiles shown on the television screen were swapped.
- Two panels of the multi-panel painting on the wall were interchanged.

Figure 6. The age room



F. The Ethnicity Room

This room was designed to investigate whether participants exhibit gaze preferences that may reflect the *own-race bias*, a phenomenon known from social and visual perception research. The own-race bias describes the tendency for individuals to allocate more visual attention to, and more easily recognize, faces that share familiar or self-similar features. While our setup does not directly measure face recognition performance, prior gaze-based studies have demonstrated that attention patterns toward social stimuli can reveal demographic and behavioral biases [1], [3].

Based on these findings, our hypothesis for this room was that participants would spend more time looking at non-player characters (NPCs) who share similar ethnic features and would be more likely to initiate interaction with them. The task was therefore designed as a simple social interaction scenario: participants were asked to freely observe six NPCs of different ethnicities for several seconds and then start conversations with two of them. Once two NPCs were selected, the participant could not interact with the others. After finishing both conversations, the participant could proceed to the next room by pressing the green button.

For this room, we created NPCs representing six ethnic backgrounds:

1) German (representing White/European)
2) Mexican (representing Latin American)
3) Iraqi (representing Middle Eastern)
4) Singaporean (representing East/Southeast Asian)
5) Black American (representing African American/Mixed)
6) South African (representing dark-skinned African)

The NPCs were positioned randomly in each session. Each NPC had a distinct name and a short biographical description to enable natural and context-rich conversations. The conversations selections were protocolled for later analysis to investigate whether ethnicity similarity predicts interaction choice and gaze duration.

G. The Weight Room

This room was designed to investigate whether a participant's body weight can be inferred from their gaze behavior toward food stimuli. Previous studies have shown that individuals with higher body mass index (BMI) tend to exhibit longer fixation durations on food items, particularly those that are energy-dense or high in calories [7]–[9]. These findings suggest that visual attention and preference strength are positively correlated with food energy content, especially among overweight individuals.

In our implementation, participants were presented with four different foods displayed on a virtual table:

1) Salad (25–50 kcal/100 g, typically 200 kcal per portion; healthy, low-calorie)
2) Grilled salmon (150–200 kcal/100 g, typically 300–440 kcal per portion; healthy, low-calorie)
3) Rotisserie chicken (230–250 kcal/100 g, typically 600–750 kcal per portion; high-fat, high-calorie)
4) Pizza (260–290 kcal/100 g, typically 900–1200 kcal per portion; high-fat, high-calorie)

The task required participants to "eat" the foods in the order of their personal preference—placing each item onto a plate in front of them to trigger an eating animation and sound. This task enabled us to measure both attentional patterns (gaze duration) and behavioral preferences (eating order) to determine whether they correlate with participants' BMI.

The room was designed to resemble a modern kitchen environment rather than a laboratory setting to increase ecological validity and evoke a naturalistic, appetizing atmosphere.

The four food items were selected to represent two contrasting categories—low- versus high-calorie—while keeping familiarity and visual appeal constant. Calorie values were derived from the USDA FoodData Central database [10]: salad (25–50 kcal/100 g), grilled salmon (180 kcal/100 g), rotisserie chicken (240 kcal/100 g), and pizza (270 kcal/100 g). This composition follows prior research showing that gaze duration and preference strength correlate positively with calorie content, particularly among individuals with higher BMI [6], [7]. Presenting two healthy and two unhealthy foods thus enables comparison of attentional and preference biases under controlled visual conditions.

1) Problems with BMI: While BMI is an imperfect proxy for body composition, it remains a commonly used and easily measurable indicator for categorizing participants' weight status. For the purposes of this study, BMI was considered sufficient to test the hypothesis regarding attentional bias toward high-calorie foods.

IV. Current Controls

The project uses the Varjo XR-3 headset for immersive mixed reality. The HTC Vive Controller is used as the primary input device. The trackpad of the controller is used for movement. The sensor in the trackpad recognizes the relative location of the player's finger when touching it, which translates into a two-dimensional interpretation of the direction and velocity with which the player wishes to move.

Furthermore, the player triggers teleportation by pressing the (north) upper region of the trackpad, as described in Section IX-C. By pressing the western / eastern (left/right) region of the trackpad, the player can turn around by 90°. The trigger button at the back of the controller is used for all interactions with objects defined in this report. When the trigger is pressed, the game interprets the action as an attempt to interact with the closest *Interactable Object*.

Additionally, voice input is activated and deactivated through the controller interaction. When the player presses the grip button located at the sides of the controller, the application starts recording microphone input. Pressing it again stops the recording. This method provides users with an intuitive way to control when their voice is captured, ensuring privacy and reducing unintended background recordings. During the production phase, the keyboard could simulate any of the inputs described above. However, in the final version of the game, only the VR setup can be used to play the experiment.

Moreover, there are certain controls of interests to researchers conducting the experiment: By pressing the space bar on the keyboard, headset calibration is triggered. Also, the 'K' key on the keyboard activates the button that participants need to press in order to exit the tutorial room. If deactivated, it is not possible to continue to the first room.

V. Removed or Altered Features

During the development of the application and experimental setup, several features were either removed, simplified, or altered for practical, technical, or ethical reasons. These decisions were made to streamline the user study, ensure feasibility, and maintain compliance with ethical standards.

A. Controller Usage

Although the application supports the use of two controllers for interaction, we limited participants to a single controller during the experiment. This decision was made to reduce complexity and ensure more consistent input handling across participants. The second controller was disabled to minimize potential confusion and simplify data logging.

B. Ethical Constraints on Attribute Prediction

We initially considered including a room designed to explore the prediction of sexual orientation based on gaze behavior. From an implementation perspective, this would have been similar to our room focusing on ethnicity-based attention patterns. However, due to the sensitivity of the topic and the potential ethical concerns around user profiling, the idea was ultimately discarded. Any future exploration of this or similar attributes would require extensive ethical review and user consent procedures.

C. Switch from Microsoft Rocketbox to Ready Player Me

In early versions of the application, we used Microsoft Rocketbox avatars for NPCs. While Rocketbox provides high-quality 3D characters, their customization options are extremely limited: clothing and appearance are fixed and represented only as texture maps. To achieve greater visual diversity and better control over the outward appearance of NPCs, we switched to the Ready Player Me platform. This allowed us to create customizable avatars with varied outfits and styles, better supporting the visual design needs of each room and making it easier to implement controlled variations in NPC appearance.

D. Removal of Quick Turn Feature

The project also includes an optional rotation feature that allows participants to rotate their view by fixed increments of 90° to the left or right. This functionality was designed to help players quickly reorient themselves within the virtual environment without relying on physical body rotation, making it especially useful for setups with limited physical space.

When enabled, the feature allows players to press the left or right region of the controller's trackpad to rotate their viewpoint by 90° in the corresponding direction. The functionality is implemented in the `LaserPointer` script, which is attached to the controller GameObjects and handles user input events.

The feature is currently deactivated in the experimental build. To activate it, the developer must uncomment the function calls `rotateLeft()` and `rotateRight()` within the `Update()` method of the `LaserPointer` script. Once active, pressing the left or right area of the trackpad will trigger the respective rotation, enabling smooth incremental orientation adjustments.

The feature was disabled in the current version of the experiment to ensure consistent gaze direction logging and to prevent unwanted head-turning behavior that could affect eye-tracking accuracy. Moreover, participants already found the other controls to be challenging enough. The quick turn feature led to unwanted inputs of the participants when trying to teleport.

E. Robot Navigation and Follow Behavior

An additional feature implemented in the project, though currently disabled, allows the in-game robot to dynamically follow the participant throughout the virtual environment. When this functionality is active, the robot continuously tracks the player's position and automatically walks toward their updated location whenever the player teleports. This creates a more interactive and responsive experience during conversations or task instructions.

The behavior is based on Unity's `NavMesh` system. The robot uses a `NavMesh Agent` component to calculate a navigable path across the floor surface, allowing smooth motion while avoiding obstacles. To enable this functionality, the following steps are required:

1) Activate the `NavMesh` on the floor object to define the traversable area.
2) Enable the `NavMesh Agent` component on the robot GameObject.
3) In the robot's `ChatGptRobot` script, check the parameter `Robot Should Move`.

Once these settings are applied, the robot automatically follows the participant after each teleport action, maintaining an appropriate conversational distance. This feature was deactivated for the current experiment to minimize motion distractions and maintain focus on the eye-tracking tasks.

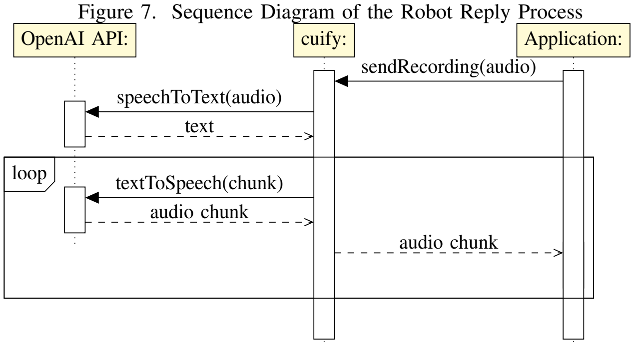
VI. Interaction with other APIs

In our project, we rely on a number of other APIs. The interaction with these APIs and their usage within the project are introduced in this section.

A. Generation of Robot Replies

The robot's replies are generated using the OpenAI API. In particular, we used the **OpenAI GPT-4o-mini model** (with streaming enabled). However, the application does not directly communicate with the API, but communicates via a *cuify* server that runs in a Docker container. The cuify server was developed by members of the Chair of Human-Centered Technologies for Learning at TUM. A quick description follows in IX-G. Cuify is an application that allows for speech-to-speech generation of text in Unity. The main benefit of using cuify is that it allows for splitting generated text into chunks and generating the audio per text chunk, which speeds up the process and shortens the reply time of the robot in the Unity environment. In our application, we record the voice of our participants when the microphone is activated. The recording is then sent to the cuify server. Cuify requests for a Speech-To-Text transcription of the audio to the OpenAI API. This transcription is then further split into text chunks by cuify. For each text chunk, cuify requests a Text-To-Speech audio from the OpenAI API again and then sends the audio bits back to the Unity application. This is much faster than generating one big MP3 file per audio reply. Furthermore, the transcribed text of the user audio and the generated reply text are also being protocolled on the cuify server for analysis purposes. Figure 7 depicts this process as a sequence diagram.

Figure 7. Sequence Diagram of the Robot Reply Process



One more notable thing about our cuify usage is that we adjusted the cuify source code to fit our needs. Thanks to Kadir Burak Buldu (burak.buldu@tum.de), cuify now allows the application to send a text with the recorded audio. This text will also be forwarded to the OpenAI API as additional information to use when generating the robot replies. This way, we can give the API information about the current state of the Unity scene and the task for the user.

B. NPC Audio Generation

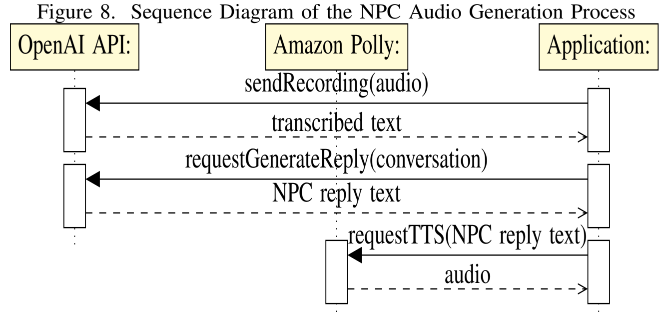
As Cuify only allows for speech-to-speech generation, and the OpenAI API does not offer the possibility to generate audio with distinct accents, we implemented a custom solution combining both OpenAI and Amazon Polly for dialogue generation and text-to-speech synthesis. Distinct accents were necessary for our project, as the NPCs were designed to represent diverse cultural and linguistic backgrounds.

For this purpose, we used the **OpenAI GPT-4o-mini model** (with streaming enabled) for text generation, paired with the **OpenAI Whisper model** for speech-to-text transcription. The system prompt for the LLM defined each NPC's conversational behavior (e.g., a friendly nurse from Singapore), while the recorded user audio was sent directly to the OpenAI API without using the Cuify server described in Section VI-A.

The resulting text output was subsequently sent to **Amazon Polly (Generative Engine, 2024 release)** for text-to-speech processing. The final audio returned by Amazon Polly was then played as the NPC's speech in Unity.

As with all interactions in the experiment, the full dialogue logs between users and NPCs were recorded and stored for later analysis. Figure 8 displays the process described in this section.

Figure 8. Sequence Diagram of the NPC Audio Generation Process



C. Protocolling speech

As mentioned in section VI-A and VI-B, we protocol all non-static speech that takes place during the experiment. Whenever the player talks to the robot in the environment or an NPC in the ethnicity room, we protocol the text generated by OpenAI's STT service of the participants' recorded microphone input, as well as the LLM-generated answer. The protocol is then saved as one file per conversation partner per room.

D. Eye Tracking

For eye tracking, we used the Varjo XR-3 headset in combination with the official Varjo Unity XR Plugin[1]. Our implementation closely follows the example provided in the `EyeTrackingExample.cs` script[2], which provides access to real-time gaze data, including fixation points and eye convergence.

We extended this functionality by logging additional metadata alongside each gaze sample. Specifically, our data schema includes columns for:

- the name of the object currently held in the left and right hands/controllers (or one side if only one controller is used),
- the name of the object the user is currently looking at,

[1] https://developer.varjo.com/docs/unity-xr-sdk/eye-tracking-with-varjo-xr-plugin (accessed: 20 September 2025)

[2] https://github.com/varjocom/VarjoUnityXRPlugin/blob/master/Samples~/HDRP/EyeTracking/Scripts/EyeTrackingExample.cs (accessed: 20 September 2025)

– timestamps for each gaze event, and
– frame-based scene context information (e.g., active task and room identifier).

This enriched data format allows us to correlate gaze behavior with interaction context, supporting more fine-grained analysis of user attention and task performance.

E. Task-specific tracking

In addition to gaze data and object interaction logging, we implemented task-specific tracking mechanisms tailored to each room's experimental hypothesis.

In the **gender room**, we used custom logic to monitor which numbered fields on the table were occupied or unoccupied at any given moment, including which specific object was placed or removed. This provided insight into participants' object placement behavior and spatial attention.

In the **age room**, we recorded the position of every movable object in the scene on every frame. This included both objects that were meant to be manipulated as part of the task and those that were technically movable but irrelevant to the goal. By capturing this fine-grained motion data, we could later reconstruct how each participant handled the objects and assess whether they moved wrong items, which could serve as an indicator of age-related behavioral patterns.

For the **ethnicity room**, we logged the full text of all spoken interactions between the player and the NPCs. This allowed us to analyze linguistic patterns, question choices, and conversation structure in relation to the presented ethnic diversity of the characters.

Finally, in the **weight room**, we tracked the precise sequence in which players placed different food items on their plates. The order of selection provides an additional layer of behavioral data beyond gaze fixation, potentially indicating implicit preferences or decision-making strategies that correlate with the participant's body weight.

VII. User Study

This section describes the empirical study conducted to investigate whether personal attributes such as age, gender, ethnicity, and body weight can be inferred from eye tracking data alone in a virtual reality environment. The subsections below follow the standard structure for reporting human-subject experiments.

A. Participants

A total of 69 individuals participated in the experiment, recruited via the ExperimenTUM platform at the Technical University of Munich (TUM). One dataset was excluded due to unreliable eye-tracking performance caused by reflections from glasses, resulting in a final sample of 68 valid participants. The participants represented a diverse range of ethnic backgrounds and body types, though the majority were in their twenties—a distribution typical for ExperimenTUM studies. Participants received a compensation of €7.50 in the form of an Amazon voucher for completing the study. Before participation, all individuals read and signed a written informed consent form approved by TUM, granting permission to use anonymized eye-tracking and interaction data for research purposes. Participants were informed that only gaze and interaction data would be recorded, and that no personally identifying information such as images or audio would be stored.

B. Apparatus

The experiment was implemented as a Unity-based VR application. Participants wore a **Varjo XR-3** headset, which provides integrated high-precision eye tracking at 200 Hz, and used an **HTC Vive controller** as the primary input device. The Unity project utilized several APIs and frameworks:

– **Unity Engine** [11] for real-time rendering, physics, and scene management.
– **SteamVR** [12] for hardware interfacing and controller bindings.
– **Vive Input Utility vive** for hand and controller input.
– **OpenAI API openai** and **AWS Polly** [13] for generating natural-sounding speech and dialogue for AI-driven NPCs.
– **cuify framework cuify** (running in Docker) for handling speech-to-speech communication between Unity and external APIs.
– **Ready Player Me** for generating customizable avatars used as NPCs in the ethnicity room.

C. Procedure

Upon arrival, participants were welcomed and briefed about the purpose of the study. They then provided informed consent and filled out a short demographic questionnaire. The experimenter guided each participant through a tutorial room to familiarize them with the controls and VR environment. Following this, the Varjo XR-3 eye tracker was calibrated using a five-point calibration procedure. Participants then completed the four task rooms in random order. Each room included clear on-screen and spoken instructions displayed via a virtual "info wall" and an interactive AI robot. The experimenter monitored progress remotely but did not intervene. After completing all rooms, participants filled out an online questionnaire including the NASA-TLX [14], System Usability Scale (SUS) [15], and the Igroup Presence Questionnaire (IPQ) [16], as well as self-reported data on height, weight, and ethnicity. The total duration per session was approximately 25–30 minutes.

D. Measurements

Each task collected a distinct set of dependent measures:

- **Gender Task:** Self-reported object ranking order
- **Weight Task:** Sequence of selected food items, derived "unhealthy preference score" (range 1–5), and associated gaze data.
- **Ethnicity Task:** NPC selection order and conversation logs to assess interaction preferences across ethnic groups.
- **Age Task:** Reconstruction time and average object placement error (Euclidean distance) during the spatial memory task.

Additional subjective metrics (workload, usability, presence, motion sickness) were collected via the post-experiment questionnaire to contextualize behavioral results.

E. Data Processing

Raw gaze data were exported from the Varjo XR Plugin at 200 Hz and synchronized with Unity's scene context via timestamps. Each gaze event included fixation point, hit object name, and active room identifier. Pre-processing steps included:

1) Removal of invalid samples with missing or zero gaze coordinates.
2) Exclusion of participants with more than 15% invalid samples (1 participant removed).
3) Aggregation of gaze samples per object or per task phase (e.g., fixation time per item, mean error per object).
4) Calculation of derived metrics such as unhealthy preference score and mean spatial error.

All processing and analysis scripts were written in Python 3.11 using `pandas` and `numpy`.

F. Data Analysis

The collected data were analyzed using a combination of non-parametric testing and descriptive visualization. For the gender task, object ranking differences between male and female participants were examined using the **Mann–Whitney U test**. This test was chosen because the data were ordinal (rank-based) and the sample size per group was small, violating the assumptions of normality required for parametric tests such as the independent-samples t-test.

A **two-tailed** test was applied, as no specific directional hypothesis was made about which gender would rank particular object categories higher or lower. All statistical analyses were conducted with a significance level of $\alpha = 0.05$, and the corresponding effect size (r) was reported to indicate the magnitude of observed differences.

To complement the inferential analysis, **descriptive and visual methods** were employed, including boxplots and violin plots that illustrate distributional patterns and individual variability across all experimental tasks. All statistical computations were performed using the `scipy.stats` library in Python, and all visualizations were generated with `matplotlib` and `seaborn`.

VIII. Results

This section presents the results of the four experimental tasks conducted in virtual reality, each designed to investigate whether participants' gaze behavior and interaction preferences reveal patterns related to demographic or psychological attributes. Analyses combine self-reported data (e.g., object rankings or BMI) with behavioral metrics recorded in VR, such as task performance and gaze-based interactions. Each subsection corresponds to one of the task rooms—*gender*, *weight*, *ethnicity*, and *age*—and summarizes both the quantitative findings and their interpretation in the context of prior psychological or human–computer interaction research.

A. Results of the Gender Task

In this task, participants ranked six objects according to their personal preferences (1 = liked most, 6 = liked least). The selected objects represented stereotypically "male" or "female" interests while remaining sufficiently ambiguous to avoid revealing the purpose of the task. Our aim was to determine whether gaze-based preferences align with self-reported rankings and whether these preferences differ systematically between genders.

To assess gender differences, we conducted a non-parametric Mann-Whitney U test for each object. Five of the six comparisons yielded statistically significant differences ($p < 0.05$):

- **Whiskey glass** ($p = 0.0254$): Men ranked the whiskey glass higher than women, consistent with its traditional association with masculinity.
- **BBQ tools** ($p = 0.0424$): Women ranked BBQ tools lower than men, reflecting the stereotype that grilling is more often associated with male leisure activities.
- **Yoga mat** ($p = 0.0386$): Women ranked the yoga mat higher, aligning with the perception of yoga as a more female-oriented activity.
- **VR headset** ($p = 0.0204$): Men showed a stronger preference for the VR headset, potentially reflecting greater male engagement with gaming and technology.
- **Backpack** ($p = 0.0273$): Women ranked the backpack higher, possibly indicating an aesthetic or practical appreciation more common among female participants.

For the **paint box**, no significant gender difference was found ($p = 0.2081$), suggesting that artistic interests may be more evenly distributed across genders within this participant sample. This opposes our hypothesis that women tend to be more interested into arts.

Table I provides the complete results of the gender preference task, illustrating the average ranking position assigned to each object by men and women.

Table I
AVERAGE OBJECT RANKINGS BY GENDER (LOWER = MORE PREFERRED)

Gender	Whiskey Glass	Paint Box	BBQ Tools	Yoga Mat	VR Headset	Backpack
Man	4.45	3.34	3.52	4.21	1.62	3.86
Woman	5.13	2.87	4.13	3.38	2.51	2.97

Overall, these results show that even with relatively subtle object choices, gender-related patterns in preferences emerge. Items such as the VR headset and BBQ tools reflect traditional male-coded preferences, while the yoga mat and backpack align more with female-coded ones. The paint box, however, appears to serve as a neutral item, hinting at a possible shift away from conventional gender associations in certain domains. Together, these findings support the hypothesis that both gaze-based and explicit rankings can capture culturally embedded gender differences in object preference.

B. Weight Room: Food Preference and BMI

The *weight room* examined whether gaze and choice behavior align with participants' body weight. Four foods were presented: *salad* and *salmon* (lower calorie) and *pizza* and *roasted chicken* (higher calorie). Participants "ate" items in order of preference by placing them on a plate in VR, which triggered an animation and sound.

At the end of the study, participants self-reported height and weight, from which BMI categories were computed. The distribution was: **45** healthy (BMI 18.5-25), **3** underweight (BMI < 18.5), **16** overweight (BMI 25-30), and **4** obese (BMI > 30). Figure 9 illustrates these counts across BMI categories.

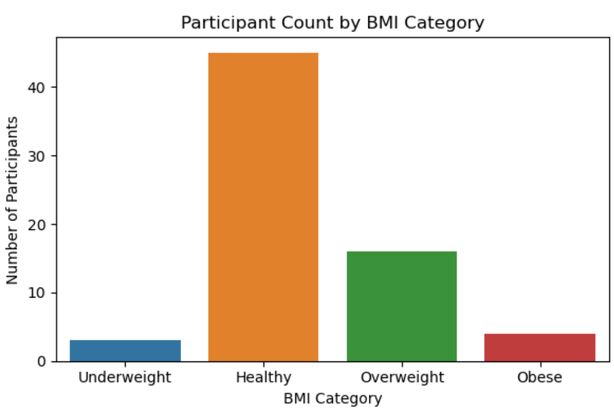


Figure 9. Participant count by BMI category.

a) Unhealthy preference score: To quantify choices, we defined an *unhealthy preference score* capturing how highly the two high-calorie items (pizza, chicken) were ranked:

– 1st place: 3 points, 2nd: 2 points, 3rd: 1 point, 4th: 0 points.

Summed across the score of the two high-calorie items, scores range from 1 (when ranked the place 3 & 4) to 5(when ranked 1 & 2); higher values indicate stronger preference for high-calorie foods.

The unhealthy preference score transforms ordinal food rankings into a continuous metric representing attraction to energy-dense foods. Similar additive indices are widely used in nutritional psychology to quantify tendencies toward high-calorie stimuli [6]. By assigning higher weights to earlier selections, this score integrates both order and valence, allowing meaningful comparisons across BMI categories while maintaining interpretability.

Figure 10 shows a split *violin plot* with overlaid points (colored by gender), which makes small-n groups visible while conveying distributional shape.

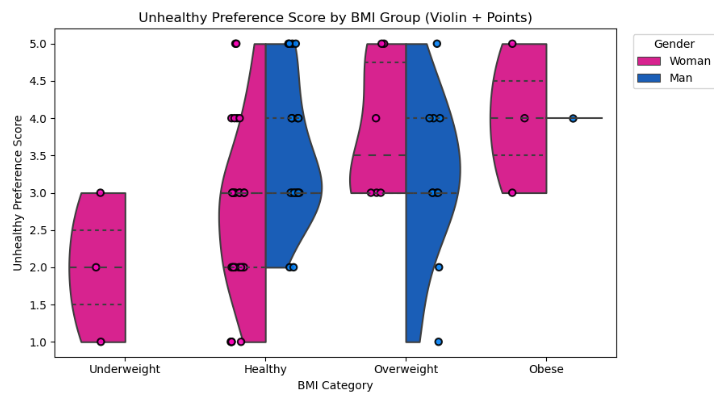


Figure 10. Unhealthy preference score by BMI category and gender

A clear pattern emerges: underweight participants cluster at the lower end of the scale, with none scoring above 3, indicating a consistent preference for lower-calorie foods. In contrast, obese participants appear exclusively at the upper end, with none scoring below 3, reflecting strong and uniform preferences for calorie-dense foods. Healthy and overweight groups display a broader spread, suggesting more heterogeneous preferences overall, with a mild upward shift in scores from the healthy to the overweight range.

For completeness, Figure 11 plots individual BMI values against unhealthy scores (color-coded by gender).

Taken together, these results support the hypothesis that food-choice behavior in VR correlates with BMI. Despite the controlled, minimal stimulus set, we observe robust patterns consistent with prior behavioral findings that overweight and obese individuals tend to show stronger preferences for high-calorie foods.

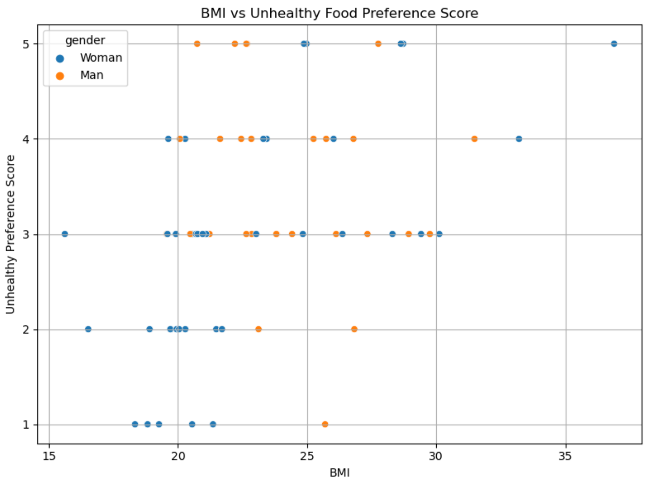


Figure 11. Scatter of BMI vs. unhealthy preference score

C. Ethnicity Room: NPC Selection Preferences

The *ethnicity room* was designed to study whether participants display systematic preferences when choosing social interaction partners of different ethnic backgrounds. In this task, six NPCs were presented, each representing a distinct broad racial category (*African, American, Arab, European, Asian, Latinx*). Participants were instructed to choose two NPCs to initiate a short conversation with. Moreover, we asked the participants to state their own ethnicity in the questionnaire at the end of the study. Figure 12 summarizes the broad participant distribution. The majority of participants identified as **Asian** ($n = 27$) or **White** ($n = 20$), followed by **Other/Unspecified** ($n = 13$), **Hispanic or Latino/a** ($n = 5$), and **Black or African** ($n = 3$). A more fine-grained distribution is provided in Figure 13, showing the diversity within the Asian and White subgroups, such as South Asian ($n = 12$), East Asian ($n = 8$), European ($n = 20$), Middle Eastern/North African ($n = 6$), and several smaller subgroups.

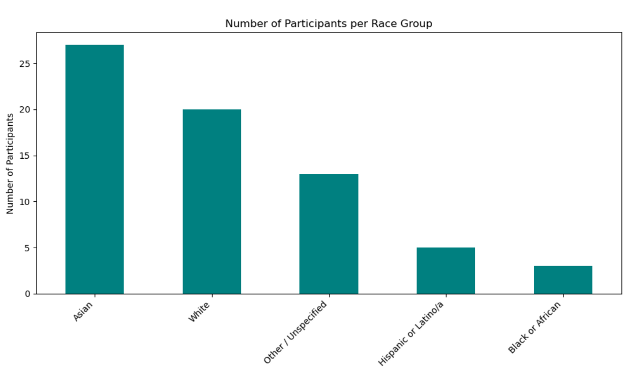


Figure 12. Broad ethnicity distribution of participants.

a) Results: A detailed breakdown of which ethnic group interacted with which NPCs can be found in Appendix A. The **European NPC** was selected most frequently overall, across nearly all participant groups, regardless of their own ethnicity. Conversely, the **African and African American NPCs** were consistently chosen least often. While this may suggest implicit social preferences, alternative explanations such as differences in **clothing style** or **facial expression design** should also be considered.
Breaking down the data by first and second choices reveals additional nuances. As shown in Figure 23, the European NPC was the most common **first choice**, particularly among White and Black/African participants. The second choice distribution (Figure 24) was more varied: some groups, such as Hispanic/Latino/a participants, showed a more balanced spread across NPC ethnicities, while others (e.g., Asian participants) maintained a consistent preference for European or Mexican NPCs.

b) Interpretation: The results indicate a clear general bias toward the **European NPC**, independent of participants' self-reported ethnicity. NPCs of African descent were consistently under-selected, whereas the Mexican and Japanese NPCs were moderately popular among Asian and Hispanic/Latino/a participants. Although such outcomes may partly reflect implicit social preferences, they could also stem from visual or stylistic elements of NPC design. Future work should therefore control for visual design variables to isolate ethnicity as a factor of interest and further explore cross-cultural patterns in social choice behavior.

D. Age Task

The *age room* was designed to measure short-term memory and spatial recall in a VR environment. Participants were first presented with a furnished 3D living room scene and instructed to memorize the arrangement of objects. The task consisted of two phases:

1) **Memorization (Phase 1):** Participants had up to two minutes to study the room. They were allowed to skip ahead once they felt prepared. In practice, most participants used nearly the full duration, resulting in little variance for this phase.
2) **Reconstruction (Phase 2):** The scene was presented again, but with several objects misplaced. Participants were asked to reposition the objects in their original locations, with no time limit imposed. Both completion time and placement accuracy were recorded.

a) Results: Figure 14 shows the relationship between age and task duration. While Phase 1 exhibited little variance, Phase 2 revealed a clear age-related trend: older participants required significantly more time to complete the reconstruction task. This finding aligns with prior research on age-related decline in processing speed and working memory.

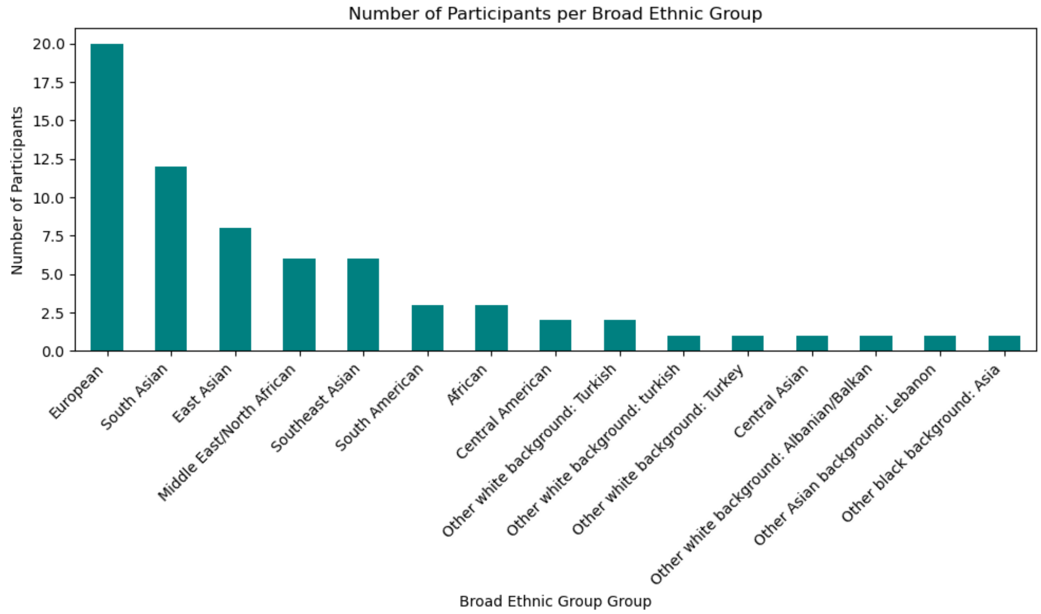


Figure 13. Detailed ethnicity distribution of participants.

Figure 15 illustrates the relationship between age and reconstruction accuracy, measured as average object error (Euclidean distance from the correct position). The regression indicates that older participants not only took longer but also made larger errors when rearranging objects. Although variance was relatively high in younger participants, the overall trend suggests that memory-related performance declined with age even in this short, task-oriented VR setting.

b) Interpretation: These results demonstrate that VR-based memory and recall tasks are sensitive to age-related performance differences. The combination of longer completion times and higher placement errors among older participants highlights the potential of eye-tracking and interaction data in VR to serve as non-invasive indicators of cognitive decline. Future work could refine this task with more controlled object sets or adaptive difficulty to better capture subtle age-related effects.

E. Statistical Assumptions and Justification

All statistical tests and visual analyses used in this study are based on explicit methodological assumptions that are reasonable given the nature of our data and study design. Because each experimental task produced ordinal or small-sample data, we selected non-parametric and distribution-free methods to avoid violations of normality and homoscedasticity assumptions.

a) Non-parametric testing: For the gender task, we used the *Mann–Whitney U test* to examine whether object rankings differed between male and female participants. In this task, participants were asked to rank six objects from 1 (most interesting) to 6 (least interesting). The test compares the distribution of ranking positions across gender groups to identify potential gender-based preferences.

The Mann–Whitney U test assumes that the dependent variable (here: ranking position) is at least ordinal and that observations are independent across groups. These conditions are met because each participant provided exactly one ranking per object, and gender categories were mutually exclusive. The test further assumes that the two group distributions have a similar shape; visual inspection of the ranking data using violin plots confirmed comparable spreads, supporting the validity of this assumption. This non-parametric method was therefore more appropriate than a t-test, which requires interval-scaled data and normally distributed residuals.

b) Regression and correlation.: Linear regression models used in the age-related analyses assume a monotonic relation between predictors (age) and dependent variables (task duration or spatial error) and independence of residuals. Since each participant contributed only one data point per metric and visual inspection showed no strong curvature or clustering, a simple linear model was considered valid. Although normality of residuals cannot be fully guaranteed with small n, linear regression remains robust for approximately symmetric data, especially when used for exploratory visualization rather than inference.

c) Independence and sampling.: All participants performed the tasks individually and in randomized order. No participant completed the same task twice, and the eye-tracking calibration was reset between sessions. These measures ensure independence of observations across participants and across tasks. Because each task

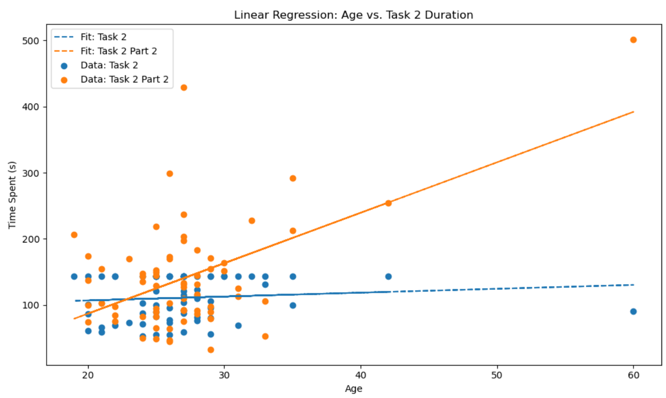


Figure 14. Linear regression of time spent on memorization (Phase 1) and reconstruction (Phase 2)

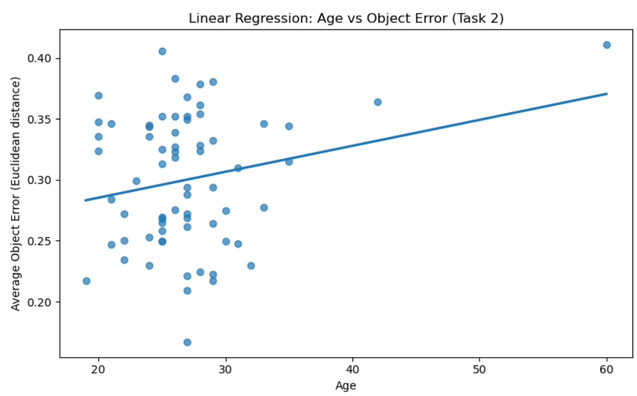


Figure 15. Linear regression of age vs. average object placement error

targeted a distinct behavioral domain (object preference, food choice, spatial memory, or social attention), cross-task correlations are minimal and do not violate the assumption of independent samples.

d) Visualization and interpretability.: The chosen visualization methods—split violin plots, scatter plots with regression lines, and boxplots—assume no particular data distribution. They are designed to display full variance and individual points rather than aggregated means, supporting transparent interpretation of heterogeneous samples, as recommended in behavioral and HCI research [3], [14].

e) Summary.: In summary, the applied statistical tests and visualization methods adhere to assumptions appropriate for small-to-medium sample behavioral data: independent observations, ordinal or continuous variables, and approximately similar distributional shapes across groups. These conditions were verified through data inspection before analysis, ensuring that the reported results are statistically valid and interpretable within the constraints of non-parametric inference.

IX. Used Tools

This section presents the core technologies and software frameworks employed in the development and execution of the experiment. Because the project was realized as an interactive 3D environment in virtual reality, the implementation relied primarily on **Unity** as the game engine, supplemented by VR-specific frameworks such as **SteamVR** and the **Vive Input Utility**. In addition, several APIs and cloud-based services were integrated to provide functionalities such as speech synthesis, avatar generation, and large-language-model interaction. The following subsections describe each tool in detail and outline its role in the project.

A. Unity

Unity [11] is one of the most widely used game engines for developing both 2D and 3D interactive experiences. It provides comprehensive functionality for rendering, physics, sound, and input handling. Game logic and interaction mechanics were implemented as Unity `C#` scripts within this environment.

B. SteamVR

SteamVR [12] provides the interface between Unity and VR hardware such as the HTC Vive and Varjo XR-3. It handles device pairing, tracking, and input bindings, allowing for seamless integration of head-mounted displays and controllers into the Unity environment.

C. Vive Input Utility

The Vive Input Utility [17] is a Unity plugin that facilitates the handling of Vive input devices. It enables the tracking of controller status, button presses, and hand gestures, and was used to implement player movement and interaction within the VR environment.

D. OpenAI API

The OpenAI API [18] provides access to advanced natural-language processing models that generate human-like text responses based on textual prompts. In this project, it was also used to convert text into audio via text-to-speech (TTS) generation for interactive dialogues.

E. Amazon AWS

We used **Amazon Web Services (AWS)** [13], specifically the **Amazon Polly** [19] service, to generate text-to-speech audio in multiple languages and regional accents. Amazon Polly is a cloud-based service that converts text into natural-sounding speech using deep learning, allowing us to provide participants with realistic audio instructions in several languages (e.g., British English, American English, German). This made the experimental environment more immersive and enabled linguistic variation without requiring local audio storage or pre-recorded files.

F. Ready Player Me

Ready Player Me [20] is a cross-platform avatar creation tool that allows developers to generate fully rigged, customizable 3D characters. In this project, it was used to create non-player characters (NPCs) with distinct appearances and outfits. Unlike Microsoft Rocketbox [21], which provides static meshes with limited customization, Ready Player Me enables more dynamic visual variety—essential for creating consistent yet diverse NPCs across the experimental scenarios.

G. cuify

cuify [22] is a framework developed by the Chair of Human-Centered Technologies for Learning at TUM. It simplifies the integration of speech-to-speech interaction within Unity projects and enables communication with various large language models, including those from OpenAI.

H. Docker

Docker [23] was used to host the cuify server, ensuring consistent deployment and dependency management across different systems. By containerizing services, Docker guaranteed that the same configurations and dependencies were maintained, regardless of the underlying operating environment.

X. Discussion

Our findings confirm that eye tracking data, even when collected in a minimal and controlled VR setting, can reveal meaningful information about a user's personal attributes. However, this very potential also highlights significant privacy and ethical concerns. In this section, we discuss the broader implications for privacy in virtual environments, methodological limitations of our experiment, and issues that should be approached with caution.

A. Privacy and Ethical Implications

Eye tracking provides continuous, fine-grained behavioral data that reflects not only where users look, but also cognitive and emotional processes occurring below conscious awareness. This makes gaze data uniquely sensitive [2], [3]. In immersive systems such as VR or AR headsets, this sensitivity is amplified by the duration and contextual richness of the data being collected. Even in our simplified setup—without avatars, voice, or hand tracking—it was possible to infer gender, weight, and approximate age with non-trivial accuracy.

Such capability raises concerns about involuntary profiling. In commercial VR ecosystems, gaze data could be monetized for targeted advertising, adaptive pricing, or behavioral manipulation [24]. The subtlety of gaze collection makes it unlikely that users are aware of what personal traits are being inferred, undermining informed consent. Moreover, anonymization does not guarantee safety, since gaze patterns themselves can act as biometric signatures [1].

Therefore, future XR platforms must treat gaze data as personally identifiable information. Developers and policy makers should implement privacy-by-design principles such as local data processing, consent dialogs, and real-time transparency indicators showing when gaze data is being logged or used for inference.

B. Methodological Limitations

While the experiment demonstrates proof of concept, certain limitations restrict the generalizability of the results. The participant pool consisted primarily of university students, resulting in a narrow BMI range and limited age diversity. This homogeneity may have amplified certain effects (e.g., between-group contrasts) while concealing others. Future studies should include a more balanced population sample, particularly with respect to body composition, age, and cultural background.

BMI, though a convenient metric, does not fully capture body composition or health status. It fails to distinguish between muscle and fat mass, which could lead to misleading interpretations. Similarly, while gaze-based gender differences were significant, they reflect cultural associations rather than biological determinants—meaning that results may vary across societies or generations.

Another limitation is the small number of participants in some demographic categories, which affects the statistical power of non-parametric tests. Although the Mann–Whitney U test is robust for small samples, broader datasets would improve reliability and support multivariate modeling in future research.

C. Additional Privacy Considerations

Beyond demographic inference, gaze data could also expose momentary emotional states, levels of stress, or cognitive load [6]. In real-world XR applications, this opens the door for manipulative design practices—e.g., adjusting advertisements when users appear fatigued or distracted. Even more critically, continuous gaze logging across multiple sessions could allow re-identification or longitudinal tracking of users without explicit consent.

Therefore, eye tracking data must be considered a biometric signal under frameworks such as the GDPR and CCPA. Any future deployment of gaze-enabled systems should provide clear opt-out mechanisms, limit retention time, and minimize raw data storage. Technical safeguards such as on-device processing and real-time anonymization represent key steps toward ethical gaze-based computing.

XI. Outlook

The results of this project illustrate both the potential and the risks of eye tracking in virtual environments. Going forward, the central question is not only *what* can be inferred from gaze, but *how* this knowledge should be used responsibly.

A. Future Research Directions

To strengthen the validity of our findings, future work should focus on increasing participant diversity and expanding the range of measurable traits. Integrating additional modalities such as hand tracking or facial expressions could improve predictive accuracy—but would also intensify privacy challenges. Therefore, subsequent studies should explicitly balance predictive performance with data minimization.

A natural next step is to explore user awareness and perception of gaze-based inference. Survey-based follow-ups could assess whether participants understand how much information is revealed through eye tracking and how comfortable they are with such analyses. Insights from these studies could guide the design of transparency interfaces and consent mechanisms for commercial VR platforms.

B. Ethical and Design Implications

The knowledge gained through this experiment should not be used for profiling, but for prevention. Our findings underline the need for XR systems that process gaze locally, store only aggregated metrics, and offer users control over when gaze tracking is active. Developers should clearly indicate when gaze data is used for adaptive interaction rather than for analysis.

At the policy level, our results can support ongoing discussions about categorizing gaze as biometric data and defining stricter handling requirements. Establishing standardized privacy guidelines for XR research—similar to those for medical or facial data—would ensure that the benefits of eye tracking are realized without compromising user autonomy.

C. Final Remarks

Ultimately, our project serves as both a technical proof of concept and a warning. The ability to infer personal traits from eye movements highlights a powerful intersection of behavioral science and machine learning—but also exposes new privacy risks that must not be ignored. Eye tracking should empower users, not profile them. Ensuring this balance will define the ethical success of future VR and metaverse technologies.

XII. Conclusion

This project demonstrates that eye tracking data collected in virtual reality can reveal far more about users than they might expect. Even without facial, vocal, or behavioral inputs, our results show that subtle gaze patterns can predict personal attributes such as gender, age, and body weight with measurable accuracy. These findings highlight both the scientific potential and the profound ethical challenges of gaze-based interaction.

From a technical perspective, our Unity-based experimental framework shows that minimal visual environments can already generate predictive signals about user traits. This suggests that gaze, often treated as a neutral input modality for foveated rendering or object selection, is in fact a soft biometric. As such, eye tracking systems should no longer be designed under the assumption that gaze data is anonymous or harmless.

From a privacy standpoint, the implications are substantial. In commercial XR ecosystems, where eye tracking may be active by default, similar inference mechanisms could enable user profiling, behavioral targeting, or adaptive advertising without explicit consent. Even when identifiers are removed, characteristic gaze patterns can make re-identification possible. For this reason, eye tracking data must be treated as personally identifiable and handled with the same level of protection as other biometric data under privacy regulations such as the GDPR or CCPA [1]–[3].

At the same time, the results underscore the potential for constructive applications—such as health diagnostics, cognitive assessment, and adaptive accessibility tools—if implemented ethically. Achieving this balance requires a privacy-by-design approach in which gaze data is processed locally, stored only in aggregated form, and collected transparently with clear user consent.

In conclusion, eye tracking is not merely a tool for interaction but a window into human identity, cognition, and behavior. Its integration into virtual and augmented reality technologies offers immense promise but also new forms of vulnerability. As researchers and developers, we bear the responsibility to ensure that these insights are used to empower, not exploit, the individuals behind the data. The future of XR privacy will depend not only on technological innovation, but on the ethical frameworks we choose to uphold.

References

[1] J. L. Kröger, O. H.-M. Lutz and F. Müller, “What does your gaze reveal about you? on the privacy implications of eye tracking,” in *Privacy and Identity Management. Data for Better Lives*, ser. IFIP Advances in Information and Communication Technology, M. Friedewald, S. Schiffner, S. Rung, M. Rost and K. Wadhwa, Eds., vol. 576, Springer, 2020, pp. 226–241. (accessed 02/09/2025).

[2] D. J. Liebling and S. Preibusch, “Privacy considerations for a pervasive eye tracking world,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing:*

Adjunct Publication, ACM, 2014, pp. 1169–1176. (accessed 21/09/2025).

[3] J. Steil, D. Sonntag and A. Bulling, “Privacy implications of eye tracking in virtual reality: A case study with htc vive,” in *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, 2019, pp. 1–26. (accessed 09/09/2025).

[4] L. Wang, Y. Shen, J. He and Y. Wang, “Gender classification using eye movement features in reading,” in *2018 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2018, pp. 1–7. (accessed 02/09/2025).

[5] Y. Sugano and A. Bulling, “Age estimation using a deep learning approach with gaze features,” in *Proceedings of the 2016 ACM Symposium on Eye Tracking Research & Applications*, 2016, pp. 271–274. (accessed 05/09/2025).

[6] D. J. Graham, A. K. Hoover and N. A. Ceballos, “Visual attention to food cues in obesity: An eye-tracking study,” *Obesity*, vol. 19, no. 9, pp. 1950–1955, 2011. (accessed 09/09/2025).

[7] L. Nummenmaa, J. Hirvonen, J. Hannukainen *et al.*, “Dorsal striatum and its limbic connectivity mediate abnormal anticipatory reward processing in obesity,” *PLoS ONE*, vol. 7, no. 2, e31089, 2012. (accessed 05/09/2025).

[8] E. H. Castellanos, E. Charboneau, M. S. Dietrich *et al.*, “Obese adults have visual attention bias for food cue images: Evidence for altered reward system function,” *International Journal of Obesity*, vol. 33, no. 9, pp. 1063–1073, 2009. (accessed 02/09/2025).

[9] J. Werthmann, M. Field and A. Roefs, “Attention bias for food is independent of restraint in healthy weight individuals—an eye tracking study,” *Appetite*, vol. 84, pp. 265–272, 2015. (accessed 05/09/2025).

[10] U.S. Department of Agriculture, *Fooddata central*, 2025. [Online]. Available: https://fdc.nal.usda.gov/ (accessed 18/09/2025).

[11] Unity Technologies, *Unity real-time development platform*, 2025. [Online]. Available: https://unity.com (accessed 18/09/2025).

[12] Valve Corporation, *Steamvr developer resources*, 2025. [Online]. Available: https://store.steampowered.com/steamvr (accessed 18/09/2025).

[13] Amazon Web Services, *Amazon web services (aws) cloud platform*, 2025. [Online]. Available: https://aws.amazon.com (accessed 18/09/2025).

[14] S. G. Hart and L. E. Staveland, “Development of nasa-tlx (task load index): Results of empirical and theoretical research,” in *Advances in Psychology*, vol. 52, Elsevier, 1988, pp. 139–183. [Online]. Available: https://doi.org/10.1016/S0166-4115(08)62386-9 (accessed 05/09/2025).

[15] J. Brooke, “Sus: A quick and dirty usability scale,” Digital Equipment Corporation, Reading, UK, Tech. Rep. TR-38, 1996. [Online]. Available: https://hell.meiert.org/core/pdf/sus.pdf (accessed 18/09/2025).

[16] T. Schubert, F. Friedmann and H. Regenbrecht, “The experience of presence: Factor analytic insights,” *Presence: Teleoperators & Virtual Environments*, vol. 10, no. 3, pp. 266–281, 2001. [Online]. Available: https://doi.org/10.1162/105474601300343603 (accessed 05/09/2025).

[17] HTC Corporation, *Vive input utility for unity*, 2025. [Online]. Available: https://developer.vive.com/resources/vive-sense/sdk/vive-input-utility/ (accessed 18/09/2025).

[18] OpenAI, *Openai api documentation*, 2025. [Online]. Available: https://platform.openai.com/docs (accessed 18/09/2025).

[19] Amazon Web Services, *Amazon polly: Text-to-speech service*, 2025. [Online]. Available: https://aws.amazon.com/polly/ (accessed 18/09/2025).

[20] Wolf3D, *Ready player me: Cross-platform avatar creator*, 2025. [Online]. Available: https://readyplayer.me (accessed 18/09/2025).

[21] Microsoft Corporation, *Microsoft rocketbox avatar library*, 2025. [Online]. Available: https://www.microsoft.com/en-us/research/project/microsoft-rocketbox/ (accessed 18/09/2025).

[22] Chair of Human-Centered Technologies for Learning, Technical University of Munich, *Cuify: Conversational user interfaces for xr environments*, 2025. [Online]. Available: https://www.ce.cit.tum.de/hctl/projects/cuify/ (accessed 18/09/2025).

[23] Docker Inc., *Docker: Accelerated container application development*, 2025. [Online]. Available: https://www.docker.com (accessed 18/09/2025).

[24] S. Mayer, T. Krüger and D. Widmer, “The metaverse and the behavioral economy: Eye tracking, predictive modeling, and the new frontier of data privacy,” *Journal of Virtual Worlds Research*, vol. 13, no. 2, 2020. (accessed 09/09/2025).

Appendix
Boxplots of preferences of objects in Task 1 by gender

Figure 16. backpack preferences

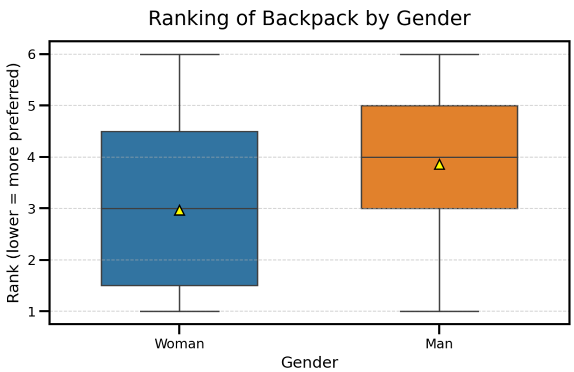


Figure 17. bbq tools preferences

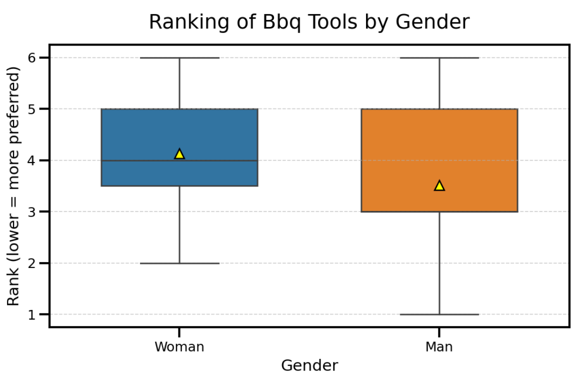


Figure 18. paintbox preferences

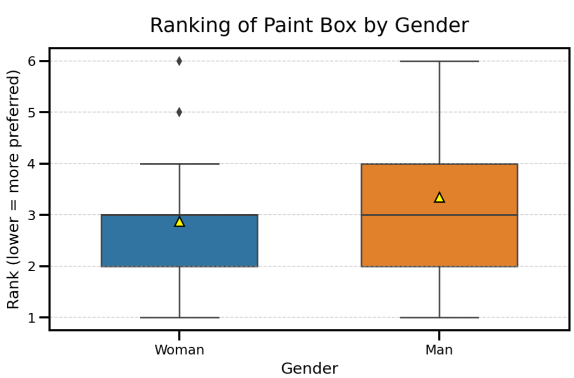


Figure 19. vr headset preferences

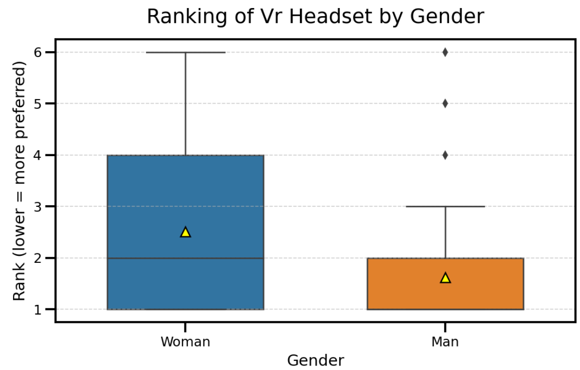


Figure 20. whiskey glass preferences

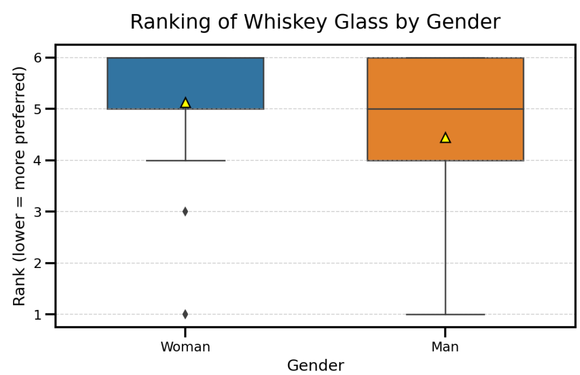
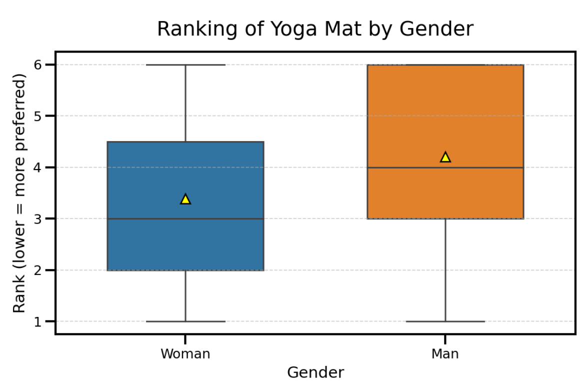


Figure 21. yoga mat preferences



Ethnicity Task - First and Second NPC Choices Detailed

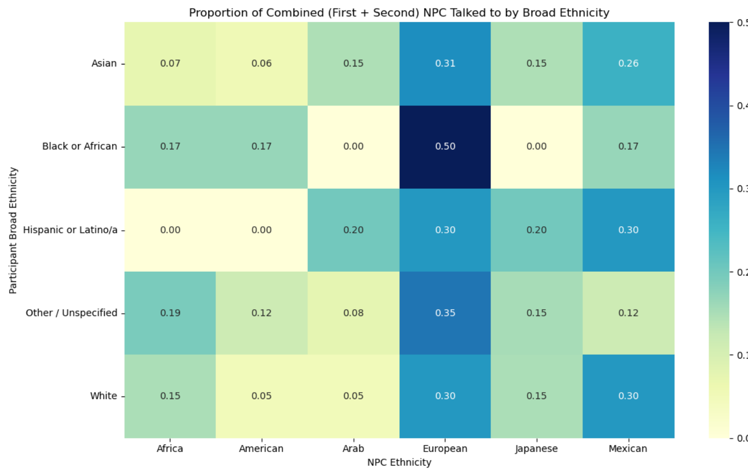


Figure 22. Combined (first + second) NPC choices.

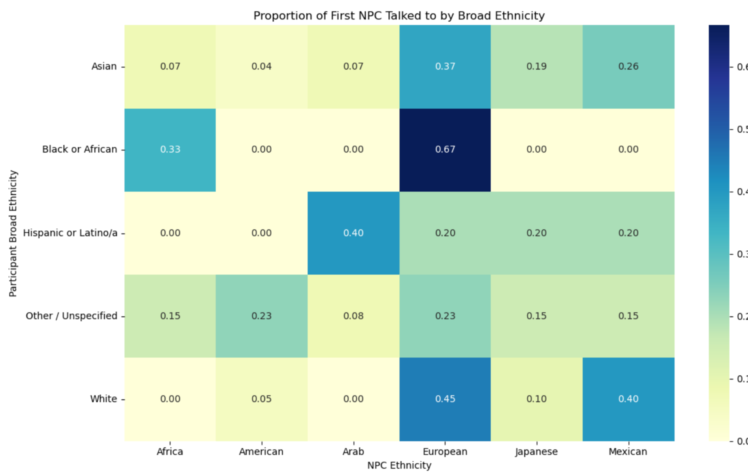


Figure 23. Proportion of first NPC choices across participant ethnicities.

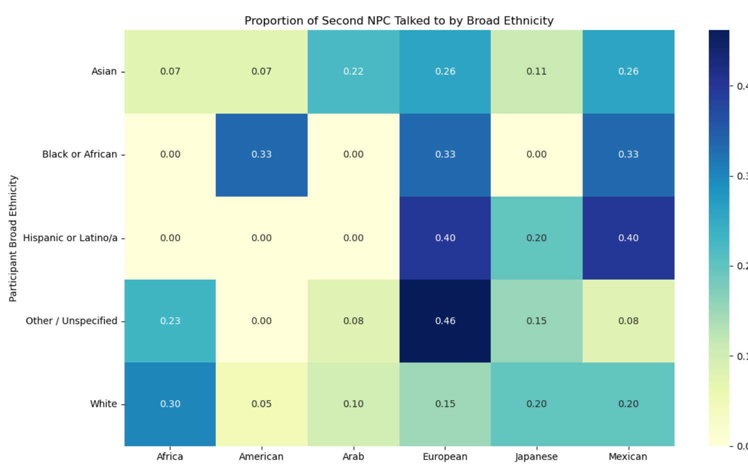


Figure 24. Proportion of second NPC choices across participant ethnicities.

Figure 25. NPC selection proportions across participant ethnicities. Each panel shows the proportion of participants from each ethnicity who selected NPCs of different ethnic backgrounds.

Assets Source

Description	Link / Source
Tutorial Assets	https://www.kodeco.com/9189-htc-vive-tutorial-for-unity
Assistance robot	https://assetstore.unity.com/packages/3d/characters/robots/robot-sphere-136226
Plate	https://assetstore.unity.com/packages/3d/props/interior/plates-bowls-mugs-pack-146682
Books in bookshelf	https://assetstore.unity.com/packages/3d/props/interior/qa-books-115415
House plants	https://assetstore.unity.com/packages/3d/vegetation/casual-plants-lite-pack-303173
Skybox	https://assetstore.unity.com/packages/2d/textures-materials/sky/city-street-skyboxes-vol-1-157401 Floor Tiles
https://assetstore.unity.com/packages/2d/textures-materials/4k-tiled-ground-textures-part-2-283704	
Whiskey glass	https://www.artstation.com/marketplace/p/6VkG7/glencairn-whiskey-glass-3d-model
Yoga mat	https://www.artstation.com/marketplace/p/jWJlg/pbr-yoga-mat-tiffany-blue
VR glasses	https://assetstore.unity.com/packages/3d/props/vr-headset-vol-2-161106
Backpack	https://www.artstation.com/marketplace/p/W8WwG/34models-of-bag-fimale-s-megapack-85-off-marvelous-clo3d-obj-fbx-only-4-99-for-24-hours?utm_source=artstation&utm_medium=referral&utm_campaign=homepage&utm_term=marketplace
Paintbox	https://assetstore.unity.com/packages/3d/props/painting-set-296016
Living room assets	https://assetstore.unity.com/packages/3d/environments/apartment-kit-124055
Age room jacket	https://sketchfab.com/3d-models/jacket-0ef825d04a05436f8a91c1ac489366cf
Kitchen props	https://assetstore.unity.com/packages/3d/props/furniture/kitchen-set-interior-263284
Pizza	https://assetstore.unity.com/packages/3d/props/food/seafood-pizza-set-11-1-types-310018
Big salad bowl	https://assetstore.unity.com/packages/3d/props/food/italian-salad-bowl-190766
Small salad bowl	https://assetstore.unity.com/packages/3d/props/food/mediterranean-salad-food-190533
Salmon	https://assetstore.unity.com/packages/3d/props/food/salmon-grilled-food-190507
Grilled chicken	https://assetstore.unity.com/packages/3d/props/food/chicken-pack-162713
Cutlery	https://assetstore.unity.com/packages/3d/props/food/cutlery-silverware-pbr-106932
Sci-Fi Button	https://assetstore.unity.com/packages/3d/gui/sci-fi-single-button-257195

The asset's web pages were last checked on 21.08.2025