arXiv:2411.18438v1 [cs.HC] 27 Nov 2024

# Personalized Generative AI in VR for Enhanced Engagement: Eye-Tracking Insights into Cultural Heritage Learning through Neapolitan Pizza Making

KA HEI CARRIE LAU, Technical University of Munich, Germany
SEMA SEN, Technical University of Munich, Germany
PHILIPP STARK, Lund University, Sweden
EFE BOZKIR, Technical University of Munich, Germany
ENKELEJDA KASNECI, Technical University of Munich, Germany

Virtual Reality (VR) and Generative Artificial Intelligence (Gen-AI) are transforming personalized learning, particularly in intangible cultural heritage (ICH) education. However, designing immersive experiences that enhance engagement without overwhelming learners presents a challenge. This study examines the impact of personalized AI narration on user engagement and attention in a VR environment through eye-tracking metrics. In a controlled experiment with 54 participants, we explored three levels of personalization (high, moderate, none) in a Neapolitan pizza-making task, measuring attention and cognitive load through fixation duration, saccade duration, and pupil diameter. Results indicate that high personalization increased engagement by 64.1% over no personalization ($p < 0.001$). Furthermore, regression analysis reveals specific eye-tracking metrics significantly predict gameplay duration, underscoring eye-tracking's potential to capture real-time engagement. These findings support the use of eye-tracking to inform the development of adaptive VR learning experiences. Future work may integrate subjective assessments to better understand users' underlying motivations.

CCS Concepts: • **Human-centered computing → User studies**; **Virtual reality**; • **Computing methodologies → Artificial intelligence**; • **Applied computing → Interactive learning environments**.

Additional Key Words and Phrases: Generative AI, VR, Eye-Tracking, Intangible Cultural Heritage, Education Technologies

## 1 Introduction

Intangible cultural heritage (ICH) encompasses a wide array of traditional practices, with 'cuisine' particularly representing not only the food itself but also the associated rituals and cultural identities they embody [10, 14, 40, 77, 78]. Recognized by UNESCO in 2017, Neapolitan pizza-making exemplifies ICH through its reflection of the cultural values and skills of Naples, Italy [11, 62, 68]. However, globalization and the fast-food industry's influence have significantly challenged the transmission of traditional knowledge, especially among younger generations [29, 57, 81]. As fast-paced lifestyles prioritize convenience, fewer people have the patience for the hands-on processes that traditional practices require. This growing gap makes the preservation and effective transmission of these cultural elements more critical than ever.

The advancement of new technologies such as Virtual Reality (VR) and Generative Artificial Intelligence (Gen-AI) offer innovative solutions to these challenges by providing immersive and adaptive learning experiences. VR enables learners to interact dynamically with cultural practices in a highly immersive environment, while Gen-AI personalizes content to individual preferences, enhancing engagement and learning effectiveness. Despite their potential, traditional cultural institutions like Galleries, Libraries, Archives, and Museums (GLAM) often struggle to provide dynamic and interactive experiences necessary for immersive, skill-based learning required in ICH [15, 21, 37, 51, 65]. These institutions typically rely on static formats that struggle to engage

Authors' Contact Information: Ka Hei Carrie Lau, Technical University of Munich, Munich, Germany, carrie.lau@tum.de; Sema Sen, Technical University of Munich, Munich, Germany, sema.sen@tum.de; Philipp Stark, Lund University, Lund, Sweden, philipp.stark@keg.lu.se; Efe Bozkir, Technical University of Munich, Munich, Germany, efe.bozkir@tum.de; Enkelejda Kasneci, Technical University of Munich, Munich, Germany, enkelejda.kasneci@tum.de.

diverse audiences and support varied learning styles, highlighting the need for more engaging and personalized educational tools [32, 34].

Theories such as Self-Determination Theory (SDT) [59] and Cognitive Load Theory (CLT) [79] offer a foundational framework for understanding how personalized AI strategies can enhance learning in VR environments. As VR applications increasingly aim to keep learners motivated and cognitively engaged, SDT emphasizes fulfilling psychological needs for autonomy, competence, and relatedness to foster intrinsic motivation and sustained engagement [59]. In this context, personalized Gen-AI in VR meets these needs by offering tailored guidance and interactions, providing targeted feedback aligned with individual goals, and thereby enhancing engagement, retention, and motivation [82]. Meanwhile, CLT focuses on managing cognitive resources to optimize learning [71]. By personalizing content through Gen-AI, the experience reduces extraneous cognitive load (effort spent on irrelevant information) and enhances germane cognitive load, directing cognitive resources to essential learning. This balance allows learners to process information effectively without feeling overwhelmed, creating a more engaging and meaningful learning experience.

To explore how Gen-AI-driven personalization in VR can enrich cultural heritage learning, we designed "Neapolitan Pizza VR," a virtual kitchen inspired by the "Cooking as Inquiry" approach [8, 38]. This setup allows users to learn pizza-making techniques based on their culinary style. Integrating Gen-AI into VR enables dynamic adjustments to educational content based on user background and interactions, making the learning experience more immersive and tailored to individual needs.

This study addresses the following research questions:

- **RQ1**: How effective are personalized AI strategies in VR environments for enhancing engagement in learning Neapolitan pizza-making?
- **RQ2**: In what ways do AI personalization strategies affect user cognitive load, attention, and engagement, as measured by eye-tracking metrics, within a VR-based cultural heritage setting?

In a between-subjects design with 54 participants, we evaluated the effects of three personalization levels (High, Moderate, and No personalization) on cognitive load, attention, and engagement in learning Neapolitan pizza-making. Our findings indicate that personalized AI significantly enhanced user engagement and attention, as evidenced by eye-tracking metrics, fostering further interest in cultural activities. These results underscore the impact of AI-driven personalization on cultural heritage education in VR and offer practical insights for educators. Our contributions are threefold:

(1) We developed an immersive VR kitchen centered on Neapolitan pizza-making, incorporating personalized Gen-AI not only to provide guided learning and support cultural heritage preservation but also to enhance user engagement through tailored interactions.
(2) We find that personalization captures user attention more effectively, as evidenced by time to first fixation and saccade duration metrics, while not significantly increasing cognitive load.
(3) We demonstrate that eye-tracking metrics, specifically mean fixation and saccade durations, are reliable predictors of gameplay duration, indicating these measures can effectively capture user engagement in VR learning contexts.

## 2 Related Work

### 2.1 Personalized Narrations for Cultural Education

Gen-AI has expanded the possibilities for personalized, learner-centered education, moving away from traditional one-size-fits-all approach. Through adaptable, tailored content, Gen-AI has shown promise in enhancing motivation, engagement, and critical thinking in educational contexts [22, 28, 39, 48, 74]. Recent studies increasingly explore the synergy of VR and Gen-AI to create immersive, dynamic learning environments that adapt to individual user needs [6, 23, 64, 70].

In cultural heritage contexts, personalization is especially valuable, as learners engage more deeply with material that reflects their individual cultural backgrounds and preferences. Projects like meSch [51] and PEACH [69] have led efforts to integrate personalization into cultural heritage, exploring adaptive learning paths and user profiling to enhance visitor engagement. Research has shown that personalized virtual tours and interactive exhibits increase engagement by aligning content with users' interests and prior knowledge [4, 52, 58, 80].

Recent advancements in large language models (LLMs) have further enabled virtual agents' ability to deliver real-time, context-aware insights and engage users in personalized dialogues within cultural heritage settings. Initiatives like "Ask Dali" [75] and "Awaken Sleeping Beauties" [17] demonstrate how conversational AI can deepen engagement by enabling interactions with historical personas. When embedded in VR environments, these models effectively provide a multi-perspective exploration of cultural heritage [36, 49]. However, maintaining a balance in personalization remains challenging, as overly specific content may overwhelm users due to too many choices and information overload, while insufficient personalization risks disengagement [25, 55].

Building on this groundwork, our study assesses three levels of personalization (high, moderate, and none) to examine their effects on engagement, attention, and cognitive load in a VR-based cultural heritage setting. Using eye-tracking data, we obtain a nuanced view of how personalization impacts focus and mental effort, addressing a gap in the application of eye-tracking within personalized VR learning environments for cultural heritage learning.

### 2.2 Eye-Tracking in VR: A Window to Human Cognitive Processes

Advancements in VR technology with high-resolution eye-tracking capabilities now enable researchers to observe subtle cognitive processes in real-time and non-intrusively, enhancing educational experiences [2, 7]. VR's immersive nature also supports behavioral studies that may be impractical or ethically challenging in real-world settings. In this context, eye-tracking within VR is a valuable tool for analyzing human behavior, widely used to measure cognitive load in diverse settings, such as driving simulations [47], medical procedures [76], and work-safety training [26]. In cultural heritage contexts, eye-tracking has been applied primarily to study engagement with tangible heritage like art galleries [19] and architecture [41]. While previous studies have focused on engagement and information processing, eye-tracking can objectively measure user attention and engagement, aligning with theoretical frameworks like SDT and CLT. Fixation duration, for instance, can reveal how well SDT's core learning needs are met by tracking sustained engagement and attention [59, 72], while CLT's emphasis on managing cognitive resources [71] aligns with metrics such as pupil diameter and saccade duration, which indicate cognitive load without overwhelming learners [5].

In VR educational settings, eye-tracking metrics are valuable for understanding student behavior. Fixation duration and fixation count can indicate levels of attention and information processing, while saccade metrics provide insights into visual scanning efficiency and cognitive engagement. Objects of interest (OOI) help reveal how students allocate their visual attention [7, 18]. Additional eye-tracking metrics have been applied in adaptive gameplay experiences, where longer fixation and saccade durations can signify increased cognitive load [63]. Some studies also associate pupil dilation with cognitive load, though this metric requires careful interpretation in VR due to its sensitivity to external factors like lighting [63, 66]. Research by [63] further highlights time-to-first-fixation (TTFF) and saccade amplitude as indicators of engagement. Shorter TTFF suggests content that quickly captures attention, while saccade amplitude can reflect moments of surprise or intrigue within an adaptive environment.

Although eye-tracking in VR provides valuable insights, it may not fully capture the complexity of cognition and behavior alone. In this study, we combine eye-tracking data with subjective assessments to provide a more nuanced understanding of user engagement, evaluating whether VR environments promote sustained, focused interaction with ICH content for more personalized educational experiences.

## 3 Method

In this section, we describe the demographics of our participants, apparatus, experimental design, technical implementation, procedure, data processing, measurements, and analysis. The Institutional Review Board (IRB) of the Technical University of Munich granted approval for this user study, ensuring adherence to ethical research standards.

### 3.1 Participants

The study involved 54 participants from diverse demographic backgrounds. Gender distribution was nearly balanced with 22 male participants (41%), 31 female participants (57%), and one non-binary participant (2%). Ages ranged from 18 to 54, with the majority in the 18–24 age group (46%) and the 25–34 age group (48%). Smaller percentages were in the 35–44 (4%) and 45–54 (2%) age brackets.

In terms of educational background, 15% of participants held a high school diploma, 40% held a bachelor’s degree, 43% a master’s degree, and 2% a doctoral degree. Most participants (70%) had prior experience with VR, while 30% were new to it.

Engagement with cultural heritage activities varied, with 9% engaging very frequently, 13% frequently, 41% rarely, 33% very rarely, and 4% having no prior engagement. Cooking frequency at home was also diverse, with 50% cooking very frequently, 29.6% frequently, 18.5% rarely, and 1.9% never.

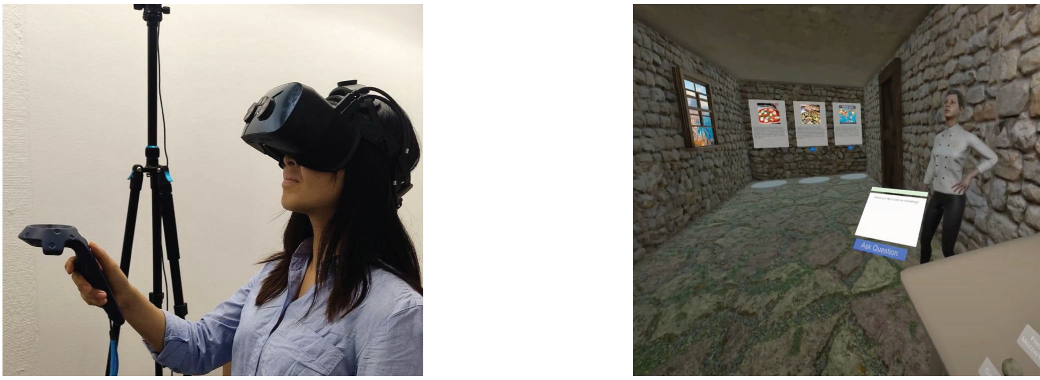
Eligibility criteria required participants to be at least 18 years old, have normal or corrected-to-normal vision, and fluency in English. Individuals with a history of severe motion sickness were excluded from the study. Each participant received a €10 voucher for their involvement at the end of the experiment.

### 3.2 Apparatus

The VR setup used in this study is shown in Figure 1a. It consisted of a Varjo VR-3[1] (Model HS-6) headset, paired with HTC Vive Controller 2.0 and HTC Vive Steam VR Base Station 2.0. The Varjo VR-3 offers a 115° field of view, a 90 Hz refresh rate, a screen resolution of 1920 × 1920 per eye, and is equipped with a built-in eye tracker that operating at a 200 Hz sampling rate.

The interactive VR game was developed using Unity (Version 2021.3.33f1). Key software extensions included the Varjo XR Plugin[2] (Version 3.6.0) for Varjo-specific support, the XR Interaction Toolkit[3] (Version 2.5.3), and XR Plugin Management[4] (Version 4.4.0), both essential for Unity-based VR development. Additionally, the OpenAI Unity[5] package (Version 0.2.0) was integrated to enable personalized AI interactions within the Unity game engine via OpenAI application programming interface.

[1] https://varjo.com/products/varjo-vr-3/, last accessed on 21 June 2024
[2] https://github.com/varjocom/VarjoUnityXRPlugin, last accessed on 19 August 2024
[3] https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@3.0/manual/index.html, last accessed on 19 August 2024
[4] https://docs.unity3d.com/2023.2/Documentation/Manual/com.unity.xr.management.html, last accessed on 19 August 2024
[5] https://github.com/srcnalt/OpenAI-Unity.git, last accessed on 19 August 2024




(a) Experiment setup. (b) Inside the VR experience.

Fig. 1. Overview of VR setup for data collection. (a) A participant using the Varjo VR-3 headset with HTC Vive controllers to engage with the VR experience. (b) Interaction scene showing the avatar within the VR environment.

### 3.3 Experimental Design

This study employed a between-subjects design with 54 participants randomly assigned to one of three conditions: no personalization (control), moderate personalization, and high personalization. The goal was to explore the impact of varying levels of personalized narration on cognitive load, visual attention, and user engagement.

The independent variable was the level of personalization, structured as follows: the **high personalization** condition dynamically adjusted narration based on both user interactions within the VR environment and demographic data from pre-assessments, the **moderate personalization** condition adapted narration according to ingredient choices selected by individual users, and the **no personalization** condition presented a standardized, non-adaptive narration. These levels were chosen to represent a spectrum of personalization, from fully adaptive to static, enabling a direct comparison of their effects on the dependent variables. This controlled setup ensures that observed differences in user responses can be directly attributed to the personalization level.

The dependent variables were cognitive load, visual attention, and user engagement within the personalized learning environment. Cognitive load was assessed through eye-tracking metrics associated with cognitive effort (e.g., pupil diameter, fixation duration and number of fixation). Visual attention was measured by TTFF, saccade duration and saccade amplitude). User engagement was evaluated through interaction log data in VR such as gameplay duration and questionnaire responses on immersion and interest.

The experiment was divided into three stages, illustrated in Figure 2. In the initial **(1) Onboarding Stage**, participants were greeted by a virtual agent and guided through the selection of pizza toppings from a list of 12 ingredients, including both traditional and non-traditional options.

In the main **(2) Gameplay Stage**, participants followed traditional steps in Neapolitan pizza-making, including dough preparation, mixing, kneading, and baking. Gen-AI provided real-time, adaptive instructions based on each participant's profile, with the level of personalization varying by experimental condition.

Throughout the experience, participants had the opportunity to explore three **(3) Posters** related to the history and cultural significance of Neapolitan pizza. The experiment concluded once participants successfully completed their personalized pizza. The design followed a linear structure, guiding participants through the steps to bake the pizza in sequence, with personalization levels tailored to each experimental condition.

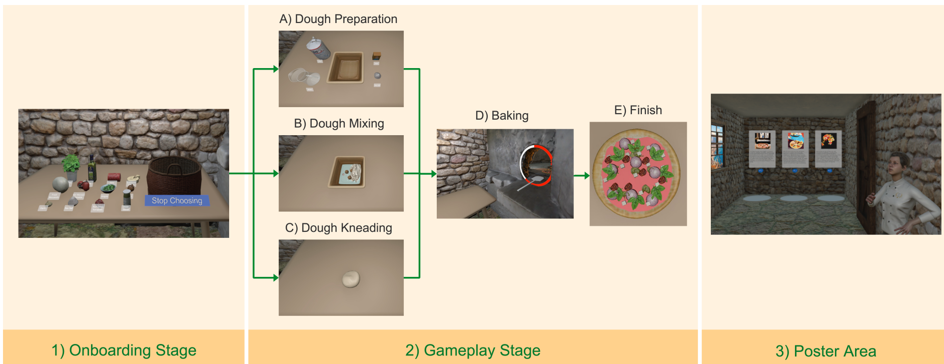



Fig. 2. The experimental setup includes: (1) the Onboarding Stage with ingredient selection and narration options, (2) the Gameplay Stage for hands-on pizza-making, and (3) the Poster Area featuring personalized cultural content. Levels of personalization—none, moderate, and high—determine the extent of interaction with the avatar and posters.

### 3.4 Technical Implementation

The technical implementation of 'Neapolitan Pizza VR' involved three main components: developing an immersive VR environment, designing a virtual coach, and enabling context-aware personalization.

For the VR environment, we created an interactive simulation that guided participants through the traditional steps of Neapolitan pizza-making, including ingredient selection, dough handling, topping choices, and wood-fired oven baking, as shown in Figure 2.

The Virtual Coach was powered by OpenAI's GPT-4 model, configured as a culturally informed guide to enhance educational authenticity. To ensure accurate guidance and minimize AI-generated inaccuracies, we applied prompt engineering with few-shot learning techniques [24, 60]. These prompts provided **Role definition** and **Instructional guidelines**, framing the AI as a "cultural ambassador" within the VR setting and specifying tone and response style [43].

For context-aware personalization, we used content from a massive open online course (MOOC) on Neapolitan pizza-making[6]. GPT-4[7] generated culturally relevant responses and image prompts for DALL-E[8], guided by participants' pre-questionnaire responses and ingredient selections. This personalization extended to dynamically generated posters within the VR environment, where GPT-4 and DALL-E collaboratively produced customized text and images aligned with each user's style and learning preferences.

To maintain accuracy, the virtual coach's responses were controlled through stage-specific prompts limited to the MOOC content. Figure 3 presents the architecture of the virtual agent, incorporating GPT-4, OpenAI Whisper[9] for speech-to-text (STT), the OpenAI Audio API[10] for

[6] https://www.federica.eu/federica-pro/pizza-revolution/, last accessed on 18 October 2024
[7] https://openai.com/index/gpt-4-research/, last accessed on 21 June 2024
[8] https://openai.com/index/dall-e-3/, last accessed on 17 October 2024
[9] https://openai.com/index/whisper/, last accessed on 21 June 2024
[10] https://platform.openai.com/docs/guides/text-to-speech/quickstart, last accessed on 19 August 2024

text-to-speech (TTS), and DALL-E for generating educational posters. Additionally, Luma AI[11] was used to create 3D ingredient models, ensuring cultural authenticity throughout the experience.

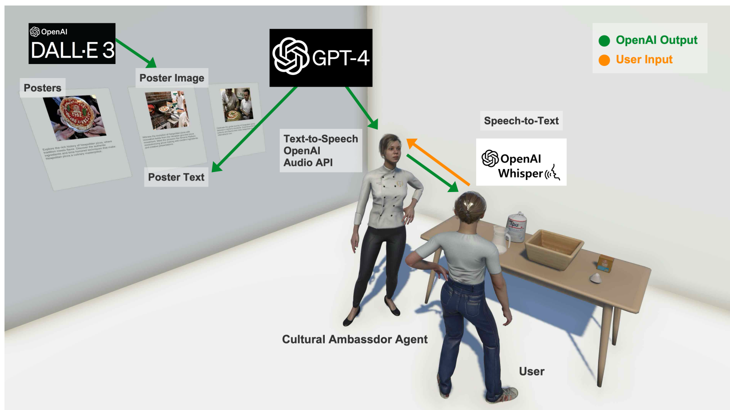



Fig. 3. Overview of the architecture for creating a personalized experience. A simplified VR environment is shown here for illustration only. Figure 1b presents the actual environment used.

### 3.5 Procedure

At the recruitment stage, participants completed a pre-questionnaire to gather demographic information, including age, gender, education level, VR experience, familiarity with the cultural content, and preferred culinary style.

Prior to starting the experiment, participants were informed of their right to withdraw at any time without consequence if they felt unwell. After a brief introduction, participants signed a consent form and were provided with an overview of the experiment's goals before proceeding. All participants were informed that the study involved making Neapolitan pizza, but they were not explicitly told that a Gen-AI agent was operating in the background. Additionally, they were informed that cultural posters were available to view, though viewing them was optional and not explicitly required.

During the experiment, participants wore a VR headset and remained standing throughout (see Figure 1a). The experiment began with a 5-point calibration phase using the Varjo headset. Following calibration, the investigator pressed the "Enter" button on the headset to initiate the actual experiment and data collection. Participants were given a few moments to explore the VR scene and familiarize themselves with the controllers.

Finally, a post-assessment was conducted. Participants were asked about their interest in the study topic and cultural heritage activities, perception of the agent's usability, a knowledge quiz on pizza-making steps, and perceived realism within the VR environment. Each session took $\approx 30$ minutes, including preparation, the experiment, and completion of the post-questionnaire.

### 3.6 Data Processing

To ensure reasonable eye-tracking quality, we removed all individuals with a tracking ratio lower than 80% in both eyes from the sample. For the remaining sample, we identified eye movement events using a Velocity Identification Threshold (I-VT) adapted for VR eye-tracking analysis [27, 61]. By exploiting gaze and head direction, we were able to calculate gaze and head velocity and determine

[11] https://lumalabs.ai/dream-machine, last accessed on 21 June 2024

events of fixation and saccades based on predefined thresholds. For the thresholds, we used previous VR experiments with similar analysis as an orientation [18, 67]. The chosen thresholds are displayed in Table 1.

Furthermore, the pupil diameter variables were also preprocessed. A subtractive baseline correction was performed since pupil diameter is considered idiosyncratic [66]. For a person-specific baseline, the median of the combined pupil diameter for five seconds at the start and end of the experiment was calculated. Further, gaze-ray casting [1, 7] was performed during the experiment, using the gaze and head information to obtain gaze target information. We defined three different OOI, namely the virtual avatar, the chat interface, and the posters.

For the eye-feature calculations, we ensured that all count measures were normalized by the experiment time and that all duration measures were stated as average values per second (e.g., mean fixation duration) or as ratios. These steps were necessary so that the duration of the experiment did not trivially influence the eye movement values.

Table 1. Head and eye movement event identification thresholds.

| **Event** | **Velocity ($v$)** | **Duration ($\Delta$)** |
|---|---|---|
| **Fixation** | $v_{\text{head}} < 7°/\text{s}$<br>$v_{\text{gaze}} < 30°/\text{s}$ | $\Delta_{\text{fixation}} > 100\,\text{ms}$<br>$\Delta_{\text{fixation}} < 700\,\text{ms}$ |
| **Saccade** | $v_{\text{gaze}} > 60°/\text{s}$ | $\Delta_{\text{saccade}} > 30\,\text{ms}$<br>$\Delta_{\text{saccade}} < 80\,\text{ms}$ |

### 3.7 Measurements

In this study, we focused on eye-tracking metrics and self-reported respond to assess cognitive load, attention, and user engagement. A pre-assessment questionnaire drawing from [33], was used to determine each participant's personalization pathway. This questionnaire included 5-point Likert scale items (ranging from 1, "not a motivation," to 5, "very strong motivation") to classify participants' backgrounds in cultural heritage (e.g., "How much do the following reasons motivate you to engage with cultural heritage activities?"). Baseline knowledge of Neapolitan pizza and familiarity with VR were also measured. Additionally, items adapted from [45] assessed participants' culinary style, distinguishing traditional versus innovative preferences, using a 5-point Likert scale (e.g., "I enjoy incorporating unique or authentic food experiences into my travel.").

For eye-tracking metrics, the study focused on indicators of cognitive load and visual attention. Cognitive load metrics included pupil diameter, fixation duration, and fixation count. Visual attention was evaluated TTFF, saccade duration, and saccade amplitude. In eye-tracking literature, longer fixation durations generally indicate heightened interest and deeper cognitive involvement [53, 84], making fixation duration a valuable indicator of the learning process [46]. Saccade metrics, such as duration and amplitude, reveal search efficiency: shorter saccades suggest efficient scanning, while longer saccade durations and larger amplitudes indicate greater effort in locating relevant elements and more thorough scanning [50, 56]. TTFF and saccade duration together provide insights into the salience of learning materials, aiding in understanding improved visual search strategies [13, 30]. Additionally, pupil diameter is commonly associated with engagement and cognitive load [31, 35]. More research shows that in a working memory task, where attention allocation is required, pupil responses might reflect differently [3].

User engagement was further measured through VR interaction data, specifically gameplay duration, as an indicator of engagement. The VR experience was designed with consistent input materials across conditions, with variations only in tone of voice and adaptive responses based on ingredient selections and questionnaire responses. Post-assessment items were adapted from [32, 42] to measure changes in attitudes, interest in cultural heritage, and perceived realism. These items, rated on a 5-point Likert scale (e.g., "After the user study, would you be more interested in participating in cultural heritage activities?"), included two additional items regarding the likelihood of using similar systems for cultural exploration. The post-assessment also included three open-ended quiz items to measure knowledge gain (e.g., "Describe the main steps involved in making the Neapolitan pizza you created.").

By integrating objective eye-tracking metrics with subjective questionnaire responses, this study aims to provide a comprehensive view of how personalized AI narration in VR affects immersive cultural heritage learning experiences.

### 3.8 Analysis

For each eye-tracking metric, we conducted statistical analyses to evaluate differences across the three experimental conditions: None, Moderate, and High Personalization. Prior to selecting the statistical tests, we assessed normality and homogeneity of variances using the Shapiro-Wilk test [44] and Levene's test [20], respectively. Metrics meeting both assumptions were analyzed with one-way ANOVA to examine group differences, followed by Tukey's HSD for post-hoc comparisons when significant $p < 0.05$. For metrics failing either assumption, we used the Kruskal-Wallis test as a non-parametric alternative, with significant results further examined through pairwise Mann-Whitney U tests, applying the Holm-Bonferroni correction.

## 4 Results

Using our user study data, in this section, we report our results measured by the following dependent variables: Cognitive Load, Attention, and Engagement for different personalization strategies. Below, we analyze these metrics in detail, focusing on how each personalization level (no, moderate, and high) impacted the respective outcomes.

### 4.1 Cognitive Load

Cognitive load was assessed using three primary eye-tracking metrics: pupil diameter, fixation duration, and number of fixation, each serving as an indicator of cognitive effort across the personalization levels.

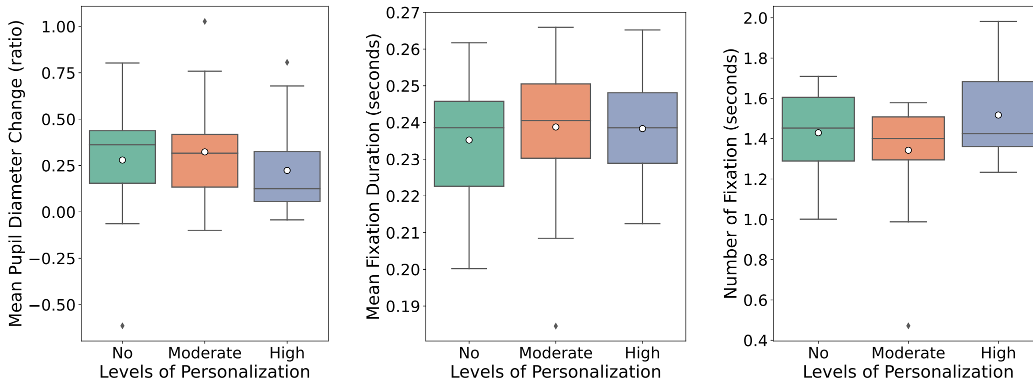
While no statistically significant differences were found among the No, Moderate, and High personalization conditions, as shown in Figure 4, descriptive statistics indicated slight trends. For instance, pupil diameter tended to increase in the Moderate personalization level, with values of ($M = 0.32$, $SD = 0.29$) for No personalization, ($M = 0.28$, $SD = 0.31$) and ($M = 0.22$, $SD = 0.25$) for High.

Similarly, fixation duration, reflecting the time spent on points of interest, showed a minor increase in the High personalization condition. Specifically, the mean fixation duration was ($M = 0.23$, $SD = 0.17$) in the No personalization condition, ($M = 0.24$, $SD = 0.02$) in Moderate, and ($M = 0.24$, $SD = 0.01$) in High.

Finally, fixation count demonstrated a non-significant increase in the High personalization group, with values of ($M = 1.43$, $SD = 0.23$) for No personalization, ($M = 1.34$, $SD = 0.26$) for Moderate, and ($M = 1.52$, $SD = 0.22$) for High.

### 4.2 Visual Attention

Visual attention metrics were examined using three primary measures: time to first fixation (TTFF), mean saccade duration, and mean saccade amplitude.




(a) Pupil diameters. (b) Mean fixation duration. (c) Number of fixation.

Fig. 4. Overview of key metrics used to evaluate cognitive load among three conditions.

Figure 5a illustrates TTFF across the three conditions (No, Moderate, and High personalization). A Kruskal-Wallis test, conducted due to non-normal data, revealed a statistically significant effect of personalization on TTFF ($H(2) = 8.73$, $p = 0.013$). Post-hoc pairwise Mann-Whitney U tests with Holm-Bonferroni correction indicated significant differences between the No and High personalization groups ($p = 0.022$, corrected $p = 0.045$) and between the Moderate and High groups ($p = 0.009$, corrected $p = 0.026$). A larger effect size was observed for the Moderate vs. High comparison (Hedges' $g = 0.943$). Descriptive statistics indicate that High personalization had the shortest mean TTFF ($M = 0.02$, $SD = 0.04$), while the No and Moderate personalized conditions had longer times ($M = 0.12$, $SD = 0.16$) and ($M = 0.34$, $SD = 0.46$), respectively.

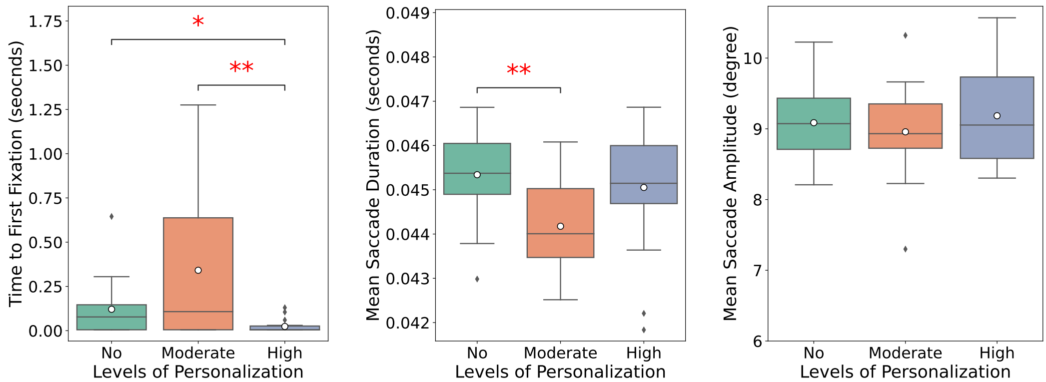
Mean saccade duration, shown in Figure 5b, exhibited variation across conditions, with statistically significant differences between the No and Moderate personalized groups ($F(2, 51) = 4.72$, $p = 0.013$, Hedges' $g = 1.065$). However, no significant differences were found between the No and High personalized groups ($p = 0.747$) or between Moderate and High ($p = 0.077$). Descriptively, the No personalized condition had the highest mean saccade duration ($M = 0.045$, $SD = 0.001$), followed by High ($M = 0.045$, $SD = 0.001$) and Moderate personalization ($M = 0.044$, $SD = 0.001$).

As shown in Figure 5c, mean saccade amplitude, the distance of eye movements between fixations revealed no significant differences across conditions. Descriptive statistics show mean saccade amplitude was similar across conditions: No personalization ($M = 9.09$, $SD = 0.50$), Moderate ($M = 8.96$, $SD = 0.65$), and High ($M = 9.18$, $SD = 0.68$).

### 4.3 Engagement

Engagement was assessed through total play duration and interest in continuing cultural heritage activities, as depicted in Figure 6a and Figure 6b. A significant overall effect was found ($H(2) = 25$, $p < 0.001$). Results indicate that the High personalization group had a significantly longer play duration than the No personalization group ($p < 0.0001$, Holm-adjusted $p < 0.0001$) and similarly, the Moderate personalization group also had a significantly longer play duration compared to No personalization group ($p < 0.001$, Holm-adjusted $p < 0.001$). No significant difference was found between the High and Moderate personalization groups ($p = 0.740$, Holm-adjusted $p = 0.740$).

Further, significant differences were observed in participants' interest in continuing cultural heritage activities, such as visiting museums, participating in local workshops, or volunteering in cultural events. The Kruskal-Wallis test revealed a statistically significant effect of condition on this



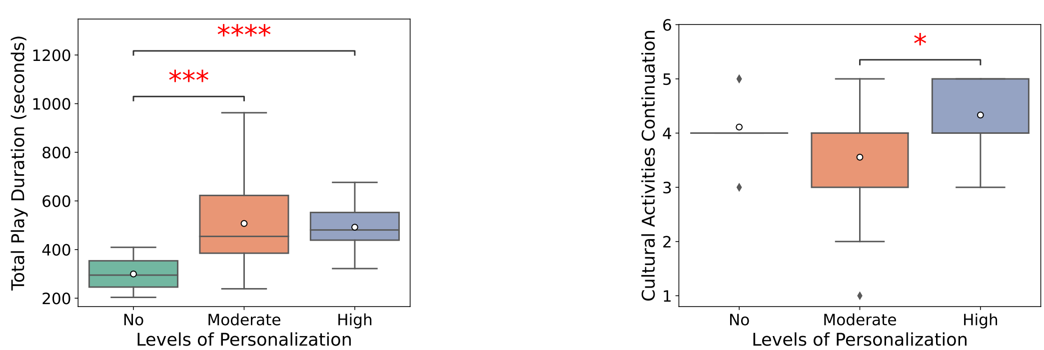

(a) Time to first fixation.

(b) Mean saccade duration.

(c) Mean saccade amplitude.

Fig. 5. Overview of key metrics used to evaluate attention among three conditions. Significance levels are represented as * for $p < .05$, ** for $p < .01$, *** for $p < .001$, and **** for $p < .0001$.

measure ($H(2) = 6.23$, $p = 0.044$). Post-hoc pairwise Mann-Whitney U tests with Holm-Bonferroni correction indicated no significant differences between the No and Moderate personalization groups ($p = 0.111$, Holm-adjusted $p = 0.222$, Hedges' $g = 0.62$) or between the No and High personalization groups ($p = 0.262$, Holm-adjusted $p = 0.262$, Hedges' $g = -0.34$). However, a significance was observed between the Moderate and High personalization groups ($p = 0.023$, Holm-adjusted $p = 0.068$, Hedges' $g = -0.83$), suggesting a moderate effect size. Descriptive statistics indicated that participants in the High personalization condition reported the highest mean interest in engaging with cultural heritage activities ($M = 4.33$, $SD = 0.69$), compared to the No ($M = 4.11$, $SD = 0.58$) and Moderate ($M = 3.56$, $SD = 1.10$) conditions.




(a) Total play duration.

(b) Interest in continuing cultural activities.

Fig. 6. Overview of key metrics used to assess engagement in the VR pizza-making task. Significance levels are represented as * for $p < .05$, ** for $p < .01$, *** for $p < .001$, and **** for $p < .0001$.

#### 4.3.1 *Prediction of Gameplay Duration.*

To evaluate eye-tracking metrics as indicators of engagement in VR learning, we performed a multiple regression analysis to assess how specific eye-tracking features predict gameplay duration. Table 2 presents the regression model with coefficients, standard errors, $t$-values, and $p$-values for each predictor, including mean pupil diameter, mean fixation duration, mean saccade duration, number of fixations, and mean saccade smplitude.

The overall model was statistically significant, $F(5, 48) = 3.209, p = 0.014$, explaining approximately 25% of the variance in gameplay duration ($R^2 = 0.251$). Notably, mean fixation duration ($b = 3737.29, p = 0.007$) and mean saccade duration ($b = -43480.00, p = 0.015$) were significant predictors, suggesting that longer fixations and shorter saccades are associated with extended gameplay duration.

In contrast, mean pupil diameter, number of fixation, and mean saccade amplitude did not significantly predict gameplay duration ($p > 0.05$ for each).

Table 2. Results of the multiple regression analysis with gameplay length as the dependent variable.

| **Model summary** | | | | R-squared = 0.251 |
|---|---|---|---|---|
| Dep. Var.: Gameplay length | | | Adjusted | R-squared = 0.172 |
| | Estimate | Std. Error | t-value | p-value |
| (Intercept) | 1791.22 | 763.84 | 2.345 | 0.023 |
| Mean pupil diameter | 97.62 | 71.19 | 1.371 | 0.177 |
| Mean fixation duration | 3737.29 | 1332.16 | 2.805 | 0.007 |
| Mean saccade duration | -43480.00 | 17300.00 | -2.511 | 0.015 |
| Number of fixation | -188.47 | 93.51 | -2.015 | 0.049 |
| Mean saccade amplitude | -5.79 | 34.68 | -0.167 | 0.868 |

## 5 Discussion

In an era where technology increasingly demands our attention, defining balanced personalization is essential to create meaningful, engaging experiences. This study addressed this challenge by exploring how adaptive AI in VR can offer individualized, culturally rich learning pathways without overwhelming users. Guided by our research questions, we discuss our findings in two main areas: enhancing engagement through personalized narration and understanding cognitive processing via eye-tracking metrics.

To answer **RQ1**, our findings indicate that personalized AI strategies are highly effective in increasing user engagement in a VR-based cultural heritage learning setting, specifically in the context of Neapolitan pizza-making.

As shown in Figure 6, the highest level of personalization not only boosted immediate engagement but also promoted ongoing cultural interest, as reflected in both eye-tracking metrics and qualitative engagement measures. Figure 6b further indicates that higher personalized narrations resonate more deeply with users, fostering sustained interest beyond the VR experience.

The predictive relationship observed between eye-tracking metrics (e.g., fixation duration and saccade patterns) and gameplay duration underscores that personalized narration can lead to extended interaction times in VR. This prolonged engagement not only leverages the immersive qualities of VR but also fosters intrinsic motivation through personalized learning pathways. Grounded in SDT, the personalization content likely satisfies the psychological needs for relevancy, autonomy and competence, which are essential for promoting intrinsic motivation and sustained engagement and ultimately encouraging users to continue exploring cultural content.

To address **RQ2**, we examined how personalized AI narration in VR influences cognitive load and attention through eye-tracking metrics. While no statistically significant differences were observed for pupil diameter, fixation duration, or fixation count, descriptive trends provided insights. The slight increase in pupil diameter in the Moderate personalization condition suggests higher cognitive load as users adapt to moderately tailored content, whereas its decrease in the High personalization condition indicates smoother cognitive processing, likely due to closer alignment with user preferences, as mentioned in research [84].

Visual attention metrics further highlight personalization's impact. Figure 5a shows significantly shorter TTFF in the High personalization condition, suggesting that tailored content effectively captures attention, orienting users quickly to relevant VR elements, aligns with research like [12, 83]. Additionally, reduced saccade duration in the Moderate condition points to enhanced scanning efficiency with personalized narration. Similar saccade amplitudes across conditions indicate that personalization did not disrupt exploratory behavior, reflecting a balanced engagement approach.

In summary, the findings suggest that well-calibrated personalization fosters engagement and attention while maintaining manageable cognitive load levels, enhancing both the depth and quality of user interactions. These insights are valuable for the future development of adaptive VR learning environments, especially in cultural heritage contexts, where balancing cognitive demands with meaningful engagement is crucial.

## 5.1 Advantages and Limitations

The use of VR to simulate skill-based ICH, such as Neapolitan pizza-making, is crucial, as accessing authentic ingredients and traditional cooking equipment is often challenging in the real world. Additionally, traditional cuisine can be unforgiving of errors, as sourcing ingredients can be time-consuming and costly [9]. Our study explores whether such a remote educational setup can not only enhance learners' understanding and engagement but also foster a lasting, meaningful, and immersive experience. Recognizing the sensitivity of personal information revealed by eye-tracking data, we approached this measurement with care, analyzing time-dependent changes in visual behavior in a manner that respects privacy while providing insights into how adaptive education shapes perception in heritage-based creative processes.

While eye-tracking offers valuable, objective measures of cognitive load and attention, it has limitations. Eye-tracking reveals 'where' and 'when' participants focus but does not capture the underlying reasons for their focus. Future research could incorporate retrospective reviewing in conjunction with eye-tracking [16, 54, 73] to gain deeper insights into user perceptions of personalized VR interactions, including aspects of learning, relevance, and enjoyment.

Moreover, the creative nature of Neapolitan pizza-making may impact eye-tracking metrics, as participants follow step-by-step instructions while simultaneously processing guidance from the virtual agent. This concurrent engagement might contribute to variability or 'noise' in certain eye-tracking metrics, as the task requires both focused attention to procedural details and flexibility in creative decision-making. Future iterations of this research would expand eye-tracking metrics to better capture stages of creativity, potentially through metrics that distinguish between convergent (focused) and divergent (creative or exploratory) attention, providing a more nuanced understanding of cognitive engagement in personalized learning environments.

## 5.2 Privacy and Ethics Statement

This study received Institutional Review Board approval from the Technical University of Munich and adhered to strict ethical standards. Eye-tracking data were anonymized, with participants fully informed about data handling and privacy. Recognizing the sensitivity of eye-tracking data, strict privacy safeguards were implemented, with secure storage limited to authorized personnel. Additionally, VR personalization poses ethical considerations around fairness and potential biases. This study aimed to enhance engagement while ensuring equitable and unbiased experiences.

Future research should continue to address the responsible use of AI personalization in VR to protect user rights and privacy, fostering positive educational outcomes.

## 6 Conclusion

This study demonstrates that personalized Generative AI in VR can meaningfully enhance sustained engagement and attention in cultural heritage learning. By leveraging eye-tracking metrics, including fixation duration, pupil diameter, and saccade amplitude, we show how culturally adaptive VR experiences promote deeper engagement and attentiveness. Additionally, eye-tracking proves to be a valuable predictor of interaction behaviors, such as gameplay duration, underscoring its utility in real-time adaptive systems. These insights highlight the potential of VR learning environments to dynamically adjust content to user backgrounds, effectively accommodating cognitive and attentional needs. Ultimately, this approach fosters higher engagement and supports an ongoing connection to cultural education, promoting a more continuous dialogue in the cultural space.

## References

[1] Najood Al Ghamdi and Wadee Alhalabi. 2019. Fixation Detection with Ray-casting in Immersive Virtual Reality. *International Journal of Advanced Computer Science and Applications* 10 (01 2019). https://doi.org/10.14569/IJACSA.2019.0100710

[2] Domitile Lourdeaux Alexis D. Souchet, Stéphanie Philippe and Laure Leroy. 2022. Measuring Visual Fatigue and Cognitive Load via Eye Tracking while Learning with Virtual Reality Head-Mounted Displays: A Review. *International Journal of Human–Computer Interaction* 38, 9 (2022), 801–824. https://doi.org/10.1080/10447318.2021.1976509

[3] Ruben D. Vromans Alwin de Rooij and Matthijs Dekker. 2018. Noradrenergic Modulation of Creativity: Evidence from Pupillometry. *Creativity Research Journal* 30, 4 (2018), 339–351. https://doi.org/10.1080/10400419.2018.1530533

[4] Eoin Bailey, Séamus Lawless, Cormac Hampson, Alexander Connor, Mark Sweetnam, Owen Conlan, and Vincent Wade. 2012. CULTURA: Supporting Enhanced Exploration of Cultural Archives through Personalisation. (01 2012).

[5] Hyejin Bang and Bartosz W. Wojdynski. 2016. Tracking users' visual attention and responses to personalized advertising based on task cognitive demand. *Computers in Human Behavior* 55 (2016), 867–876. https://doi.org/10.1016/j.chb.2015.10.025

[6] Efe Bozkir, Süleyman Özdel, Ka Hei Carrie Lau, Mengdi Wang, Hong Gao, and Enkelejda Kasneci. 2024. Embedding Large Language Models into Extended Reality: Opportunities and Challenges for Inclusion, Engagement, and Privacy. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 38, 7 pages. https://doi.org/10.1145/3640794.3665563

[7] Efe Bozkir, Philipp Stark, Hong Gao, Lisa Hasenbein, Jens-Uwe Hahn, Enkelejda Kasneci, and Richard Göllner. 2021. Exploiting Object-of-Interest Information to Understand Attention in VR Classrooms. https://doi.org/10.1109/VR50410.2021.00085

[8] Jennifer Brady. 2011. Cooking as Inquiry: A Method to Stir Up Prevailing Ways of Knowing Food, Body, and Identity. *International Journal of Qualitative Methods* 10, 4 (2011), 321–334. https://doi.org/10.1177/160940691101000402

[9] Morris Brako. 2024. Preserving culinary heritage: Challenges faced by Takoradi technical university food service lab students. *World Journal of Advanced Research and Reviews* 21 (03 2024), 1536–1545. https://doi.org/10.30574/wjarr.2024.21.3.0882

[10] Ronda L Brulotte and Michael A Di Giovine. 2016. *Edible identities: Food as cultural heritage.* Routledge.

[11] Rossella Ceccarini. 2011. *Pizza and pizza chefs in Japan: A case of culinary globalization.* Vol. 31. Brill.

[12] Qi Chen, Xiaomei Yu, Nan Liu, Xiaoning Yuan, and Zhaojie Wang. 2020. Personalized Course Recommendation Based on Eye-Tracking Technology and Deep Learning. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*. 692–968. https://doi.org/10.1109/DSAA49011.2020.00079

[13] Anwesha Das, Zekun Wu, Iza Skrjanec, and Anna Maria Feit. 2024. Shifting Focus with HCEye: Exploring the Dynamics of Visual Highlighting and Cognitive Load on User Attention and Saliency Prediction. *Proc. ACM Hum.-Comput. Interact.* 8, ETRA, Article 236 (May 2024), 18 pages. https://doi.org/10.1145/3655610

[14] Laura Di Fiore. 2020. Heritage and food history: A critical assessment. In *Food heritage and nationalism in Europe.* Taylor & Francis.

[15] Andrea Dordio, Eva Lancho, María José Merchán, and Pilar Merchán. 2024. Cultural Heritage as a Didactic Resource through Extended Reality: A Systematic Review of the Literature. *Multimodal Technologies and Interaction* 8, 7 (2024). https://doi.org/10.3390/mti8070058

[16] Sanne Elling, L.R. Lentz, and Menno De Jong. 2012. Combining Concurrent Think-Aloud Protocols and Eye-Tracking Observations: An Analysis of Verbalizations and Silences. *Professional Communication, IEEE Transactions on* 55 (09 2012), 206–220. https://doi.org/10.1109/TPC.2012.2206190

[17] Eric Hal Schwartz. 2024. OpenAI built a virtual 1920s socialite for an exhibit at the Met in New York. https://www.techradar.com/computing/artificial-intelligence/this-ai-ghost-of-the-past-might-be-your-guide-through-museums-of-the-future. Accessed: 2024-08-16.

[18] Hong Gao, Efe Bozkir, Lisa Hasenbein, Jens-Uwe Hahn, Richard Göllner, and Enkelejda Kasneci. 2021. Digital Transformations of Classrooms in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 483, 10 pages. https://doi.org/10.1145/3411764.3445596

[19] Michael Garbutt, Scott East, Branka Spehar, Vicente Estrada-Gonzalez, Brooke Carson-Ewart, and Josephine Touma. 2020. The Embodied Gaze: Exploring Applications for Mobile Eye Tracking in the Art Museum. *Visitor Studies* 23, 1 (2020), 82–100. https://doi.org/10.1080/10645578.2020.1750271

[20] Joseph L Gastwirth, Yulia R Gel, and Weiwen Miao. 2009. The impact of Levene's test of equality of variances on statistical theory and practice. *Statist. Sci.* 24, 3 (2009), 343–360.

[21] Daniel Gorman, Simon Hoermann, Robert Lindeman, and Bahareh Shahri. 2022. Using Virtual Reality to Enhance Food Technology Education. *International Journal of Technology and Design Education* 32 (07 2022), 1–19. https://doi.org/10.1007/s10798-021-09669-3

[22] Hua Guo, Weiqian Yi, and Kecheng Liu. 2024. Enhancing Constructivist Learning: The Role of Generative AI in Personalised Learning Experiences. In *Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024) - Volume 1*. 767–770. https://doi.org/10.5220/0012688700003690

[23] Teresa Hirzle, Florian Müller, Fiona Draxler, Martin Schmitz, Pascal Knierim, and Kasper Hornbæk. 2023. When XR and AI Meet - A Scoping Review on Extended Reality and Artificial Intelligence *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 730, 45 pages. https://doi.org/10.1145/3544548.3581072

[24] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2023. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. arXiv:2311.05232 [cs.CL] https://arxiv.org/abs/2311.05232

[25] Yong-Ming Huang, C.-H Liu, and C.-Y Lee. 2012. Designing a personalized guide recommendation system to mitigate information overload in museum learning. *Educational Technology and Society* 15 (01 2012), 150–166.

[26] Idris Jeelani, Kevin Han, and Alex Albert. 2018. Automating and scaling personalized safety training using eye-tracking data. *Automation in Construction* 93 (2018), 63–77. https://doi.org/10.1016/j.autcon.2018.05.006

[27] Enkelejda Kasneci, Hong Gao, Suleyman Ozdel, Virmarie Maquiling, Enkeleda Thaqi, Carrie Lau, Yao Rong, Gjergji Kasneci, and Efe Bozkir. 2024. Introduction to Eye Tracking: A Hands-On Tutorial for Students and Practitioners. arXiv:2404.15435 [cs.HC] https://arxiv.org/abs/2404.15435

[28] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences* 103 (2023), 102274. https://doi.org/10.1016/j.lindif.2023.102274

[29] Kyoung Jin Kim, Jiyoon Yoon, and Min-Kyung Han. 2021. Young chefs in the classroom: promoting scientific process skills and healthy eating habits through an inquiry-based cooking project. *International Journal of Early Years Education* 30 (08 2021), 1–10. https://doi.org/10.1080/09669760.2021.1892597

[30] Nayeon Kim. 2024. Capturing Initial Gaze Attraction in Branded Spaces Through VR Eye-Tracking Technology. *International Journal of Human–Computer Interaction* 0, 0 (2024), 1–14. https://doi.org/10.1080/10447318.2024.2351717

[31] Jeff Klingner and Pat Hanrahan. 2008. Measuring the Task-Evoked Pupillary Response with a Remote Eye Tracker. *Eye Tracking Research and Applications Symposium (ETRA)*, 69–72. https://doi.org/10.1145/1344471.1344489

[32] Markos Konstantakis and George Caridakis. 2020. Adding Culture to UX: UX Research Methodologies and Applications in Cultural Heritage. *Journal on Computing and Cultural Heritage* 13 (02 2020), 1–17. https://doi.org/10.1145/3354002

[33] Markos Konstantakis, Yannis Christodoulou, Georgios Alexandridis, Alexandros Teneketzis, and George Caridakis. 2022. ACUX Typology: A Harmonisation of Cultural-Visitor Typologies for Multi-Profile Classification. *Digital* 2, 3 (2022), 365–378. https://doi.org/10.3390/digital2030020

[34] Markos Konstantakis, Konstantinos Michalakis, John Aliprantis, Eirini Kalatha, and George Caridakis. 2017. Formalising and evaluating Cultural User Experience. 90–94. https://doi.org/10.1109/SMAP.2017.8022675

[35] Feng-Yang Kuo, Chiung-Wen Hsu, and Rong-Fuh Day. 2009. An exploratory study of cognitive effort involved in decision under Framing—an application of the eye-tracking technology. *Decision Support Systems* 48, 1 (2009), 81–91. https://doi.org/10.1016/j.dss.2009.06.011 Information product markets.

[36] Ka Hei Carrie Lau, Efe Bozkir, Hong Gao, and Enkelejda Kasneci. 2024. Evaluating Usability and Engagement of Large Language Models in Virtual Reality for Traditional Scottish Curling. arXiv:2408.09285 https://arxiv.org/abs/2408.09285

[37] Sarah Baker Lauren Istvandity and Paul Long. 2024. Creative futures for cultural heritage: a typology of creative practice in the GLAM sector – towards a creative heritage approach. *Museum Management and Curatorship* 0, 0 (2024), 1–17. https://doi.org/10.1080/09647775.2024.2331444

[38] Kai-Sean Lee. 2023. Cooking up food memories: A taste of intangible cultural heritage. *Journal of Hospitality and Tourism Management* 54 (03 2023), 1–9. https://doi.org/10.1016/j.jhtm.2022.11.005

[39] Joanne Leong, Pat Pataranutaporn, Valdemar Danry, Florian Perteneder, Yaoli Mao, and Pattie Maes. 2024. Putting Things into Context: Generative AI-Enabled Context Personalization for Vocabulary Learning Improves Learning Motivation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 677, 15 pages. https://doi.org/10.1145/3613904.3642393

[40] Claude Lévi-Strauss. 2012. The Culinary Triangle. https://doi.org/10.4324/9780203079751-12

[41] Na Li, Shanshan Zhang, Lei Xia, and Yue Wu. 2022. Investigating the Visual Behavior Characteristics of Architectural Heritage Using Eye-Tracking. *Buildings* 12, 7 (2022). https://doi.org/10.3390/buildings12071058

[42] Miguel Melo, Guilherme Gonçalves, josé Vasconcelos-Raposo, and Maximino Bessa. 2023. How Much Presence is Enough? Qualitative Scales for Interpreting the Igroup Presence Questionnaire Score. *IEEE Access* 11 (2023), 24675–24685. https://doi.org/10.1109/ACCESS.2023.3254892

[43] Swaroop Mishra, Daniel Khashabi, Chitta Baral, Yejin Choi, and Hannaneh Hajishirzi. 2022. Reframing Instructional Prompts to GPTk's Language. In *Findings of the Association for Computational Linguistics: ACL 2022*, Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 589–612. https://doi.org/10.18653/v1/2022.findings-acl.50

[44] Nornadiah Mohd Razali and Bee Yap. 2011. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *J. Stat. Model. Analytics* 2 (01 2011).

[45] Andrew Moreo, Mark Traynor, and Srikanth Beldona. 2022. Food enthusiasts: A behavioral typology. *Food Quality and Preference* 96 (2022), 104369. https://doi.org/10.1016/j.foodqual.2021.104369

[46] Shivsevak Negi and Ritayan Mitra. 2020. Fixation duration and the learning process: an eye tracking study with subtitled videos. *Journal of Eye Movement Research* 13 (08 2020). https://doi.org/10.16910/jemr.13.6.1

[47] Oskar Palinko, Andrew L. Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating cognitive load using remote eye tracking in a driving simulator *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 141–144. https://doi.org/10.1145/1743666.1743701

[48] Pat Pataranutaporn, Valdemar Danry, Joanne Leong, Parinya Punpongsanon, Dan Novy, Pattie Maes, and Misha Sra. 2021. AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence* 3, 12 (01 Dec 2021), 1013–1022. https://doi.org/10.1038/s42256-021-00417-9

[49] Pat Pataranutaporn, Phoomparin Mano, Piyaporn Bhongse-Tong, Tas Chongchadklang, Chayapatr Archiwaranguprok, Lamtharn Hantrakul, Jirach Eaimsa-ard, Pattie Maes, and Pichet Klunchun. 2024. Human-AI Co-Dancing: Evolving Cultural Heritage through Collaborative Choreography with Generative Virtual Characters. In *Proceedings of the 9th International Conference on Movement and Computing* (Utrecht, Netherlands) *(MOCO '24)*. Association for Computing Machinery, New York, NY, USA, Article 14, 10 pages. https://doi.org/10.1145/3658852.3661317

[50] Martin Raubal Peter Kiefer, Ioannis Giannopoulos and Andrew Duchowski. 2017. Eye tracking for spatial research: Cognition, computation, challenges. *Spatial Cognition & Computation* 17, 1-2 (2017), 1–19. https://doi.org/10.1080/13875868.2016.1254634

[51] Daniela Petrelli, Luigina Ciolfi, Dick van Dijk, Eva Hornecker, Elena Not, and Albrecht Schmidt. 2013. Integrating material and digital: a new way for cultural heritage. *Interactions* 20, 4 (jul 2013), 58–63. https://doi.org/10.1145/2486227.2486239

[52] Daniela Petrelli, Nick Dulake, Mark Marshall, Hub Kockelkorn, and Anna Pisetti. 2016. Do it together: The effect of curators, designers, and technologists sharing the making of new interactive visitors' experiences. (2016).

[53] Anja Podlesek, Manja Veldin, Cirila Peklaj, and Matija Svetina. 2021. *Cognitive Processes and Eye-Tracking Methodology*. Springer International Publishing, Cham, 1–31. https://doi.org/10.1007/978-3-030-71535-9_1

[54] Michal Prokop, Ladislav Pilař, and Ivana Tichá. 2020. Impact of Think-Aloud on Eye-Tracking: A Comparison of Concurrent and Retrospective Think-Aloud for Research on Decision-Making in the Game Environment. *Sensors* 20, 10 (2020). https://doi.org/10.3390/s20102750

[55] Zhixin Pu and Michael Beam. 2024. The impacts of relevance of recommendations and goal commitment on user experience in news recommender design. *User Modeling and User-Adapted Interaction* 34 (06 2024), 925–953. https://doi.org/10.1007/s11257-024-09405-1

[56] Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124, 3 (1998), 372.

[57] George Ritzer. 2021. The McDonaldization of society. In *In the Mind's Eye*. Routledge, 143–152.
[58] Ivo Roes, Natalia Stash, Yiwen Wang, and Lora Aroyo. 2009. A personalized walk through the museum: The CHIP interactive tour guide. *Conference on Human Factors in Computing Systems - Proceedings*, 3317–3322. https://doi.org/10.1145/1520340.1520479
[59] Richard Ryan and Edward Deci. 2000. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *The American psychologist* 55 (01 2000), 68–78. https://doi.org/10.1037/0003-066X.55.1.68
[60] Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal, and Aman Chadha. 2024. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. arXiv:2402.07927 [cs.AI] https://arxiv.org/abs/2402.07927
[61] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) *(ETRA '00)*. Association for Computing Machinery, New York, NY, USA, 71–78. https://doi.org/10.1145/355017.355028
[62] United Nations Educational Scientific and Cultural Organization. 2017. *Art of Neapolitan 'Pizzaiuolo'*. Retrieved July 17, 2024 from https://ich.unesco.org/en/RL/art-of-neapolitan-pizzaiuolo-00722
[63] Jose Soler-Dominguez, Jorge D. Camba, Manuel Contero, and Mariano Alcañiz Raya. 2017. A Proposal for the Selection of Eye-Tracking Metrics for the Implementation of Adaptive Gameplay in Virtual Reality Based Games. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 369–380. https://doi.org/10.1007/978-3-319-57987-0_30
[64] Yanjie Song, Kaiyi Wu, and Jiaoyang Ding. 2024. Developing an immersive game-based learning platform with generative artificial intelligence and virtual reality technologies – "LearningverseVR". *Computers & Education: X Reality* 4 (2024), 100069. https://doi.org/10.1016/j.cexr.2024.100069
[65] Tales Souza and Joao Bernardes. 2016. The Influences of Culture on User Experience, Vol. 9741. 43–52. https://doi.org/10.1007/978-3-319-40093-8_5
[66] Philipp Stark, Tobias Appel, Milo J. Olbrich, and Enkelejda Kasneci. 2023. Pupil Diameter during Counting Tasks as Potential Baseline for Virtual Reality Experiments. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications* (Tubingen, Germany) *(ETRA '23)*. Association for Computing Machinery, New York, NY, USA, Article 17, 7 pages. https://doi.org/10.1145/3588015.3588414
[67] Philipp Stark, Efe Bozkir, Weronika Sójka, Markus Huff, Enkelejda Kasneci, and Richard Göllner. 2024. The impact of presentation modes on mental rotation processing: a comparative analysis of eye movements and performance. *Scientific Reports* 14, 1 (29 May 2024), 12329. https://doi.org/10.1038/s41598-024-60370-6
[68] Marialuisa Stazio. 2021. Verace Glocal Pizza. Localized globalism and globalized localism in the Neapolitan artisan pizza. *Food, Culture & Society* 24 (03 2021), 1–25. https://doi.org/10.1080/15528014.2021.1884400
[69] Oliviero Stock, Massimo Zancanaro, Paolo Busetta, Charles Callaway, Antonio Krüger, Michael Kruppa, Tsvi Kuflik, Elena Not, and Cesare Rocchi. 2007. Adaptive, intelligent presentation of information for the museum visitor in PEACH. *User Model. User-Adapt. Interact.* 17 (05 2007), 257–304. https://doi.org/10.1007/s11257-007-9029-6
[70] Ryo Suzuki, Mar Gonzalez-Franco, Misha Sra, and David Lindlbauer. 2023. XR and AI: AI-Enabled Virtual, Augmented, and Mixed Reality. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 108, 3 pages. https://doi.org/10.1145/3586182.3617432
[71] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12, 2 (1988), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7
[72] Iman Tahamtan. 2019. The Effect of Motivation on Web Search Behaviors of Health Consumers. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (Glasgow, Scotland UK) *(CHIIR '19)*. Association for Computing Machinery, New York, NY, USA, 401–404. https://doi.org/10.1145/3295750.3298969
[73] Sean A. Tanner, Mary B. McCarthy, and Seamus J. O'Reilly. 2019. Exploring the roles of motivation and cognition in label-usage using a combined eye-tracking and retrospective think aloud approach. *Appetite* 135 (2019), 146–158. https://doi.org/10.1016/j.appet.2018.11.015
[74] Olga Tapalova, Nadezhda Zhiyenbayeva, and Dmitry Gura. 2022. Artificial Intelligence in Education: AIEd for Personalised Learning Pathways. *Electronic Journal of e-Learning* 20 (12 2022), 639–653. https://doi.org/10.34190/ejel.20.5.2597
[75] The Dalí Museum. 2024. Ask Dali - An Interactive AI Experience. https://thedali.org/exhibit/ask-dali/. Accessed: 2024-07-18.
[76] Otto Tolvanen, Antti-Pekka Elomaa, Matti Itkonen, Hana Vrzakova, Roman Bednarik, and Antti Huotarinen. 2022. Eye-Tracking Indicators of Workload in Surgery: A Systematic Review. *Journal of Investigative Surgery* 35, 6 (2022), 1340–1349. https://doi.org/10.1080/08941939.2021.2025282

[77] UNESCO. 2004. Co-operation and coordination between the UNESCO Conventions concerning heritage: the Yamato Declaration on Integrated Approaches for Safeguarding Tangible and Intangible Cultural Heritage. World Heritage Committee, 7th extraordinary session, Paris, 2004. Document code: WHC.2004/CONF.202/CLD.21, WHC-04/7 EXT.COM/INF.9.

[78] UNESCO. 2024. Oral traditions and expressions including language as a vehicle of the intangible cultural heritage. https://ich.unesco.org/en/oral-traditions-and-expressions-00053 Retrieved July 10, 2024.

[79] Jeroen J. G. van Merriënboer and John Sweller. 2005. Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review* 17, 2 (01 Jun 2005), 147–177. https://doi.org/10.1007/s10648-005-3951-0

[80] Federica Lucia Vinella, Ioanna Lykourentzou, and Konstantinos Papangelis. 2020. Motivational Principles and Personalisation Needs for Geo-Crowdsourced Intangible Cultural Heritage Mobile Applications *(UMAP '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 362–369. https://doi.org/10.1145/3386392.3399284

[81] Richard Wilk. 2006. *Fast food/slow food: the cultural economy of the global food system.* Rowman Altamira.

[82] Qi Xia, Thomas K.F. Chiu, Min Lee, Ismaila Temitayo Sanusi, Yun Dai, and Ching Sing Chai. 2022. A self-determination theory (SDT) design approach for inclusive and diverse artificial intelligence (AI) education. *Computers & Education* 189 (2022), 104582. https://doi.org/10.1016/j.compedu.2022.104582

[83] Steffi Zander, Maria Reichelt, Stefanie Wetzel, Sven Bertel, and Frauke Kämmerer. 2015. Does Personalisation Promote Learners' Attention? An Eye-Tracking Study. *Frontline Learning Research* 3 (11 2015), 1–13. https://doi.org/10.14786/flr.v3i4.161

[84] Hedda Martina Šola, Fayyaz Hussain Qureshi, and Sarwar Khawaja. 2024. AI Eye-Tracking Technology: A New Era in Managing Cognitive Loads for Online Learners. *Education Sciences* 14, 9 (2024). https://doi.org/10.3390/educsci14090933