# Speaking with Objects: Conversational Agents' Embodiment in Virtual Museums

Irene Lopez Garcia*∗
Bauhaus-Universität Weimar

Ephraim Schott*†
Bauhaus-Universität Weimar

Marcel Gohsen‡
Bauhaus-Universität Weimar

Volker Bernhard §
Bauhaus-Universität Weimar

Benno Stein¶
Bauhaus-Universität Weimar

Bernd Froehlich‖
Bauhaus-Universität Weimar


Figure 1: Two embodiment concepts for conversational agents in a museum: animated objects and an abstract humanoid guide.

**ABSTRACT**

Conversational agents in virtual environments are an established approach for immersively conveying the information and narratives of museums and cultural heritage while expanding their accessibility to a wider and remote audience. The rapid development of large language models and text-to-speech technologies has raised the agents' conversational level significantly, which allows their use for proactive guidance of visitors. This raises the vital question of how such agents should be visually represented to promote knowledge transfer in immersive virtual environments. In this paper, we compared two representation concepts for agent embodiments in the context of a virtual museum by examining a stylized humanoid guide and a novel animism-based approach that enables users to talk to exhibited objects. Our work addresses the challenge of naturally introducing a virtual educational environment to users and encouraging their interest and engagement with the content. A user study ($N = 29$) revealed high usability and similar presence scores for the experience with each of the embodiments. A majority of participants showed a preference for the animated objects. In terms of user experience, they evoked significant stimulation and high levels of engagement. Our results suggest that agents that show emotions through appropriate word choice influence engagement levels. Based on our findings, we recommend humanoid guides for delivering general background information, while animated objects promote detailed questions about their own stories and a more stimulating exchange.

**Index Terms:** Virtual Reality, VR, Agent Embodiment, Conversational Agent, Virtual Museum, Virtual Tour

## 1 INTRODUCTION

The 3D digitization of museums and historical sites has opened new vistas for cultural engagement beyond their physical confines. The growing accessibility of virtual reality (VR) devices allows a broader range of distant visitors to explore digitized museums virtually and is transforming their familiar format. With the advancement of large language models (LLMs) and text-to-speech technologies, the design of cultural learning experiences is transitioning from traditional text panels and monotonous audio guides towards more engaging and interactive forms of presentation. Embodied conversational agents herald a paradigm shift from unidirectional audio snippets to interest-driven dialogues between visitors and agents. The design and representation of these virtual interlocutors emerge as crucial factors in enhancing the overall engagement and educational value of virtual museum visits.

The use of humanoid embodied guides has established itself as the prevailing form of content narration in virtual museum applications and is recommended as a representational form for conversational agents [53]. Related work indicates that fidelity, realism, and overall style of these conversational agents' embodiment play a critical role in user immersion and presence, as well as the extent of knowledge acquisition [23, 41]. Schmidt et al. [42] suggest that thematically

*e-mail: irene.lopez.garcia@uni-weimar.de; *contributed equally
†e-mail: ephraim.schott@uni-weimar.de; *contributed equally
‡e-mail: marcel.gohsen@uni-weimar.de
§e-mail: volker.bernhard@uni-weimar.de
¶e-mail: benno.stein@uni-weimar.de
‖e-mail: bernd.froehlich@uni-weimar.de

2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 979-8-3315-1647-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISMAR62088.2024.00042

979-8-3315-1647-5/24/$31.00 ©2024 IEEE
DOI 10.1109/ISMAR62088.2024.00042

fitting guides additionally improve user experience and credibility. Despite these advantages and the widespread adoption of humanoid guides, we believe that their structured guidance might reduce the natural curiosity of users and limit their agency. Furthermore, research on stylized embodied guides whose appearance fits into the context of a museum is scarce [37, 56].

Therefore, this paper introduces speaking objects as a novel embodiment concept for conversational agents and compares it to a stylized humanoid guide representation (see Figure 1). Six objects in a virtual museum space were uniquely animated and equipped with emotion-induced response capabilities. Additionally, we drew inspiration from previous work to design a stylized guide inspired by Oskar Schlemmer, which thematically matched the art and design context of our space. A user study ($N = 29$) was conducted to evaluate the two different embodiment representations, consisting of two guided tours gathering quantitative and qualitative data. During each tour, participants interacted with a conversational agent, represented by either objects or the guide, and were able to access additional information about the space and objects via oral conversations.

Our work is motivated by a collaboration with a cultural foundation that is exploring innovative ways to communicate its museum information. With the transition of museums to accessible virtual formats, the question arises as to how and by whom guided tours should be conducted. While humanoid conversational agents provide a scalable solution, talking objects could offer visitors a more flexible alternative with greater agency. In this context, our primary research question investigates whether talking objects are accepted as viable alternatives to traditionally embodied guides. Additionally, we explore the usability of our virtual museum visits, the user experience, co-presence, user preferences for both types of embodiment, and the level of engagement they foster with the content.

Our research resulted in the following main contributions:

- Empirical evidence from a quantitative user study ($N = 29$) demonstrating that animated objects are a viable alternative to humanoid embodied agents, achieving similar presence scores.
- Findings that animated objects enhance the user experience in terms of stimulation and novelty.
- Results indicating that both embodiment modalities promote user engagement with different topics.
- Indications suggesting that emotions in responses of conversational agents can increase engagement.
- Finally, we provide design guidelines, recommending contexts for employing humanoid guides versus animated objects.

In summary, our museum application demonstrates high user-friendliness and allows visitors to engage deeply with and learn about the virtual space in an interactive and personalized manner.

## 2 Related Work

The immersion facilitated by stereoscopic systems makes VR ideal for exploring and learning within 3D environments, as it can evoke a strong sense of "being there". This phenomenon, typically referred to as presence [47], is known to enhance the quality of the user experience [8] and is influenced by various factors [47]. Pivotal factors are agency, an individual's perception that they are the initiators of actions within an immersive virtual environment [20, 46, 47], and the self-embodiment of a user [10, 15], which has been demonstrated in studies such as the "rubber hand illusion" [7] and further explored in VR contexts [22, 48]. Convincing self-embodiment can lead to behavior that is more congruent with the virtual environment [34] and maintain avatar integrity–known as the Proteus effect [16, 55]. The embodiment of other users or agents also plays a crucial role, affecting perceptions of their actions and contributing to a sense of co-presence [6]. The following sections present prior research and concepts of embodiment that informed and inspired our development of a single-user museum application with a conversational agent.

### 2.1 Embodiment and Appearance

Avatar and agent embodiments play a pivotal role in fostering both presence and social presence in virtual environments [46]. Effective representations are crucial for creating a sense of closeness and supporting collaboration in VR, as they encapsulate the principles outlined by Gutwin et al. [17] – identifying users (who), their actions (what), and their locations (where). Invisible agents fail to convey this essential information, and research in both augmented reality (AR) and VR has shown that visible representations are preferred. Visible embodiments enhance confidence, trust, and social presence by providing vital communication cues [23, 40].

Although humanoid forms are generally recommended [23, 24, 53], they are not without challenges. The uncanny valley effect, by which imperfect human-like embodiments may cause eeriness, can make stylized or abstract forms preferable [32, 41, 49]. Nonetheless, realistic embodiments offer substantial benefits that often outweigh said effect, as they improve users' subjective experiences and enhance immersion [24, 56].

Weidner et al. [53] provide an extensive review of embodiments in AR and VR, identifying five major categories for the rendering style of avatars and agents: abstract, cartoon, stylized, robot, and realistic. Their review reveals that almost 70% of studies employ full-body visualizations [53] and states that in comparison, full-body avatars significantly affect task performance, user experiences, and social presence. The realism of avatars shows similar effects, with believable interactions being key to engaging users with the content [26, 53]. However, in educational settings or during tasks where users need to maintain focus, realistic rendering seems to have less of an impact [41, 45, 53].

#### 2.1.1 Humanoid Embodiment of Conversational Agents

Voice assistants have been successfully implemented in various domains due to their intuitive and straightforward usability [3, 26]. In designing a conversational tour guide for a museum context, we specifically focused on the embodiment of these agents. Research by Kim et al. [23] demonstrates that humanoid embodiments can significantly enhance a user's confidence in the agent's capabilities. Locomotion and gestures further amplify engagement and social presence. Schmidt et al. [42] observed similar effects, noting that humanoid agents that thematically matched their educational environment were particularly effective.

Rzayev et al. [41] compared the effectiveness of invisible, robotic or realistic agents to a real tour guide in a museum. Their results showed that realistic agents had a positive effect on co-presence and that the appearance of their agents had no significant impact on learning outcomes. Interestingly, participants preferred both audio guides and robotic embodiments, despite their respective invisibility or unmatching appearance. This finding is supported by a similar result from Woodworth et al. [54]. Considering these insights, we also opted for a stylized design. However, we chose a figure that aesthetically aligns with the art context of our museum.

#### 2.1.2 Non-Humanoid Embodiment of Conversational Agents

While the majority of research on VR embodiments focuses on full-body humanoid forms, there is a notable lack of studies exploring deviating [52] or non-anthropomorphic shapes and their impacts in VR environments [2, 19, 27, 33]. Surprisingly, in the context of conversational agents, various embodiments (including humanoid, non-humanoid, actual human, and mixed or invisible agents) have not demonstrated significant differences in learning outcomes [41, 45]. Furthermore, Schroeder et al. [45] discovered that anthropomorphic features are not essential for agents to foster social agency. While some studies have examined user behavior towards simple geometric shapes such as pillars [18, 19] and cubes [50], to the best of our knowledge, there has been no research focusing on

animated realistic objects, akin to those seen in Disney films like "The Sorcerer's Apprentice" or "Beauty and the Beast".

We believe that adopting a form of animism that allows inanimate objects to speak, could significantly enhance the agency over the content and the engagement of museum visitors. This led us to integrate animated speaking objects into our VR application, seeking to foster a more interactive and immersive learning environment.

### 2.2 Enhancing Engagement and Learning

Our work was inspired by research from Kiesel et al. [21], who examined the nature of questions users asked while viewing a 360-degree panorama of a museum. This study utilized two narrative styles to present information about the exhibits: a neutral third-person and a more personal first-person. Notably, only 5% of the queries were directed at the exhibits in a manner that suggested the presence of another being (e.g., "What are you made of?"), indicating a perceived absence of social presence that we attribute to the lack of immersion and social cues.

Social cues can enhance the presence of conversational agents. Feine et al. [13] categorize these into verbal, visual, auditory, and invisible cues. In virtual museums and other educational settings, employing these social cues in conversational agent embodiments can be particularly effective, as they evoke social responses from users [14]. According to social agency theory, which is employed in the field of pedagogical agents [28, 29, 37, 44], interacting with a computer system with the assumption that it is a social being can foster user engagement and learning [4, 29].

To enhance the social agency of our object embodiment, we integrated a variety of cues and emotions, as detailed in Section 3.2. Our assumption was that users would directly engage with speaking objects that display social cues and express different emotions, leading to enhanced user engagement and a higher number of queries.

## 3 Development

This section outlines the development of two concepts of embodiment representations for conversational agents in VR, each tailored to the context in which they were to act as information providers.

### 3.1 Humanoid Guide Embodiment

For the guide, we aimed for a stylized but humanoid embodiment design that has a connection to the historic room. We chose a stylized over a realistic design, since multiple works [41, 53] showed that users preferred abstract or stylized representations in learning environments due to less distraction. The humanoid form would still provide basic communication cues such as gaze direction, controlled head movement, and mouth movement.

The visual design for the guide's embodiment is based on Oskar Schlemmer's costume design for the *Triadisches Ballett*, in particular the figure called the *Golden Sphere*. This design satisfies the desired humanoid traits whilst maintaining an aesthetic and historical link to the theme of the room (i.e., the Triadisches Ballett and the room were featured at the Bauhaus exhibition in 1923). Despite the difference in their creators and the distinctive nature of these works, the Golden Sphere's geometric volumes are grounded on the same Bauhaus school artistic principles that gave reason to the room's design. This alignment between guide and exhibition is inspired by [42] to seek greater engagement by evoking trust in first-hand knowledge and creating a congruous experience.

The figurine's geometric volumes invite playful animation, a feature that would not be achievable with a human-like agent. Notably, the spherical thorax allows the figurine to morph into a ball and scale down, facilitating smooth transitions in and out of the room.

The figurine's height was intentionally adjusted to be on the smaller end of the human scale due to considerations around its original design proportions. A larger figure would surely hinder user navigation. Additionally, some modifications were incorporated from Schlemmer's original design. Hands were added to enable the guide to point out interesting objects. Following the findings in literature on the positive effect of human-like gestures (see Section 2.2), the guide turns to the user whenever someone changes positions in the room, blinks occasionally, and points to the objects she is referring to. She also walks to the objects when they are selected, not only to visually emphasize the center of discussion, but also to mimic the normal behavior of human guides in real-world scenarios.

The guide's walk uses ready-made animation loops from *mixamo.com* [31], which can be applied to any humanoid rig out of the box. The walking paths are calculated and applied using Unity's AI Navigation module [51]. The resulting character's velocity is matched to the corresponding animation through an Animation Blend Tree. The guide's mouth movement syncs with her voice using *Oculus Lipsync* [30], a supplementary plugin provided by Meta. This plugin aligns the sound with a viseme, which is a specific shape of the mouth in the model. The guide's audio source utilizes spatial audio to uphold the authenticity of the virtual space.

### 3.2 Animated Objects Embodiment

The animated objects embodiment was developed to represent a subjective, first-person perspective on the information in the space. In this virtual exhibition, as in the majority of cultural exhibitions, the data imparted is connected to the objects in the room. Their design, their history, and their mutual connection are a major part of what makes the room interesting. Embracing an animistic concept, we used six objects as embodiments to tell their own story. These objects are famous exhibits from the director's office of Walter Gropius comprising his *F51 armchair*, *tapestry*, *Gropius' desk*, *rug*, *desk chair* and the *Wagenfeld lamp*. Images of the models of these objects can be seen in Figure 2.

To embody the conversational agents, the mesh of each object was separated from the original reconstructed model of the room. Necessary modifications were made, for example, to give a perceivable volume to the otherwise flat textiles. Rigs and animations were then created using 3D modeling software. All objects share the same set of animated behaviors: greet, talk and idle state. However, each implementation is unique, based on the object's shape, size, and position in the room. Smaller, delicate items like the Wagenfeld lamp and the desk chair exhibit more noticeable movements. In contrast, bulkier items like the desk and armchair have animations limited to lighter, upper portions. For the textiles, the rug and tapestry, only specific sections were animated. The idle state animation is only displayed by an object upon selection, else it remains static. This prevents overloading the user with continuous movement in the room, and helps to keep the focus of attention on the current interlocutor. The greet animation is played randomly when no object is selected, to encourage the user to address the objects of interest. The talk animations are played when the agent speaks, and consist of an animation loop with added deformations based on the agent's voice's volume. Despite the lack of facial features, these animations provide the necessary visual feedback to allow users to identify them as animate beings and potential conversational partners. Individual spatial audio sources enhance this perception.

### 3.3 Conversational Agents

The conversational agents were implemented by connecting multiple services through requests to the OpenAI API [35]. In particular, Whisper (`v2-large`) [39], GPT-4 (`gpt-4-0125-preview`) [36] and TTS (`tts-1`) were employed. In general, the use of LLMs in educational settings must be carefully considered due to potential hallucinations. To enhance response accuracy, we prompted the chat model with detailed information, yielding responses that we deemed sufficiently precise for evaluating our embodiments.

To initiate a conversation, a participant has to press and hold a push-to-talk button, which activates the recording of the utterance. A

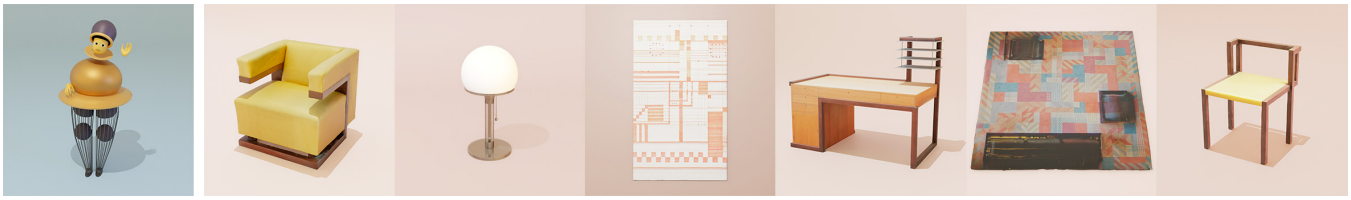

Figure 2: Developed context-based embodiments for a conversational agent. Far left: Humanoid Guide. From second left to right are the animated objects: F51 Armchair, Wagenfeld Lamp, Tapestry, Gropius' Desk, Rug, Desk Chair.

simple visual interface with text and color on the user's hand displays the current state of the voice system. Four states are supported: idle, listening, processing and speaking. An oral conversation with an agent entails

1. transcribing the spoken utterance of a user with Whisper,
2. prompting GPT-4 with the transcribed utterance, general and emotion-inducing instructions, information about the object and the location, and the chat history, and
3. synthesizing the generated response with TTS.

Distinct voices were assigned to each object to enhance their individuality and ensure they remained distinguishable.

The objects were infused with emotions that were selected based on the objects' history, fame, and visual characteristics by prompting the LLM accordingly. The emotions and character traits that were given to the objects are shown in Table 1.

The prompts to realize the humanoid guide agent contains curated information of the objects and the room. In contrast to the objects, the humanoid guide was not induced with emotions, but should rather act as a "friendly" guide.

| Object | Emotion & Trait | Object | Emotion & Trait |
|---|---|---|---|
| F51 Armchair | shy, introverted | Rug | sad |
| Wagenfeld Lamp | arrogant, self-centered | Tapestry | funny, proud |
| Gropius' Desk | serious, work-focused | Desk Chair | funny, playful |

Table 1: Emotions and character traits given to the objects.

### 3.4 Pilot Study

A pilot study with 10 participants was conducted to test these embodiments and gather user feedback. In this stage, besides the six objects described above, the room and two sections of the room could also act as embodiments. These sections are part of the architect's design of the space and play an important role in its concept. Since geometric deformations of the walls or floor could result in user discomfort, we opted for using subtle particle and light effects synchronized to the agent's voice volume. Each participant experienced a subset of all embodiments in two separate runs, exploring the room and asking questions about its content. Despite the highlight through particles and lights, users could hardly identify the space as an entity or point of discussion, often posing questions related to the objects within them. When discussing preferences, participants often favored the objects' embodiment. However, regarding space sections, users expressed doubt and generally preferred descriptions from the guide over the sections themselves. Due to the noticeably different impact of space embodiments compared to tangible objects on users, we excluded space sections from our main study.

## 4 User Study

The goal of this work is to analyze the effect that the proposed embodiment types of conversational agents in the context of virtual museums have on the user's engagement with the offered informational content, as well as on the quality of the user experience.

Although museum visits are often a social activity [1, 38], we based this study on a single-user scenario, in order to put emphasis on the user-agent conversation, and prevent inter-user social factors coming into play. In a within-subjects design, we asked participants to undertake two versions of a tour through a virtual replica of the director's office of Walter Gropius, using different embodiments for a conversational agent. Participants were asked to evaluate each version separately and finally compare both experiences using standard and particularly designed questionnaires.

### 4.1 Experimental Setup

The study was conducted in a computer lab behind closed doors, in a 4.5 m x 4.5 m free area. Although the virtual room is approximately 7.2 m x 5 m x 5 m, the physical interaction space was large enough to allow the users to walk freely around the objects of interests. No further virtual navigation was offered. As a head-mounted display, we used the Meta Quest 3 in native resolution (1,920 x 1,800) at 90 Hz in link mode. The application was built using Unity Engine and featured the model of the Gropius room, which was reconstructed from a 3D scan of the real site.

We divided the six considered objects into two distinct balanced groups, taking their physical characteristics into account: objects of similar sizes or materials were placed in opposite groups. For both groups, we defined room *tours* comprising a fixed sequence of objects: Tour A and B. Users could progress through the tour by issuing specific voice commands. After a brief introduction, the tour could be initiated with the command "start," which selected the first object. Once users focused on the selected object, they received a short description from either the guide or the object itself. Users could then initiate conversations by asking questions. The command "next object" allowed participants to move to the next item. The tour concluded with a goodbye message after the last object.

In addition, a simplified virtual reconstruction of the hallway outside the main room with three basic animated objects (cube, sphere, tetrahedron) acted as a warm-up scene, mimicking the dynamics and visual cues of the study tour.

### 4.2 Task and Conditions

Our participants were instructed to take part in the guided tour as they would in any museum or cultural space. Their task was to learn as much as possible about the history and design of the room and the objects with which they were presented. To initiate a conversation, users were tasked to ask at least one question per object in the tour.

The design of the study followed a 2x2 within-subjects design. As an independent variable, we compared two levels of embodiment modality: *guide* and *objects*. Another independent variable is the ordered subset of presented objects (tour), introduced to avoid repetition and consequent tediousness that could affect user engagement in the second treatment. Tour A featured the F51 armchair, tapestry, and the Wagenfeld lamp, while Tour B had the Gropius desk, rug, and desk chair. Each participant had both tours, each with a different embodiment. The order of the embodiment modalities and the tours was counterbalanced between participants.

Figure 3: Virtual museum room with the embodied conversational agent and guidance visualizations.

### 4.3 Procedure

Participants were welcomed to the laboratory where they were briefed about the experiment's topic, procedures, the data collected, and their rights. After providing informed consent and completing a demographic questionnaire, they put on the HMD and started the warm-up phase in the virtual hallway. Here, they were briefed on the task, introduced to the system, the voice commands and the guidance visualizations (Figure 3): a semi-transparent arrow highlighting the object of interest and a set of virtual footprints that indicated an optimal viewing location. During this phase, the participants got familiar with both embodiments. They first met the stylized humanoid agent and practiced posing questions. Then the moderator hid the guide and users talked to the animated objects.

When there were no more questions about the interaction or task, they were led to the virtual historic room, where the humanoid guide provided a brief introduction to the room and its significance. Upon the user's "start" command, the first object in the tour would be selected. In the objects embodiment condition, the guide disappeared with a short animation.

After each completed tour, the participants filled in a digital form evaluating their experience. The form was based on standard questionnaires and contained three sections: social presence, usability, and user experience. Finally, participants completed a closing custom questionnaire comparing both experiences and detailing the content they remembered. The study lasted 45 to 60 minutes in total.

### 4.4 Dependent Variables

During the study, our system logged the participants' interactions (button press and release) and conversations with the conversational agents. These include the transcriptions of the user's utterances from Whisper and the responses of the agent voiced through the TTS model. Audio and screen recordings acted as a fallback. By removing voice commands from the utterances, we determined the number of questions and comments asked to the agent.

Our evaluation included three questionnaires for the quantitative assessment of our system. The SUS questionnaire evaluates the usability of a system based on 10 statements on a 5-point Likert scale, resulting in a score from 0 to 100 [9]. The UEQ uses a 7-point scale for 26 pairs of opposing adjectives, and its analysis results in scores from -3 to 3 on 6 different scales [43]. Additionally, participants rated 5 statements for social presence on a 7-point Likert scale, based on the questionnaire from Bailenson et al. [5]. In the final questionnaire, participants indicated their preferred embodiment modality for both enjoyment and learning and justified their choices. Additionally, they were given a list of topics curated by experts on the room's history and asked to identify those they recalled hearing about during the tours. We opted for this approach in order to obtain an estimate of the knowledge acquired in correlation with the number of questions asked. Although this approach does not provide such reliable results, we chose it because it does not affect the intrinsic interest and engagement of the users as a knowledge test might.

### 4.5 Hypotheses

Our main goal was to understand the effects of different embodiments of conversational agents on the engagement of users with the provided knowledge in the virtual knowledge space. We assumed that the originality of talking objects in contrast to a more traditional embodiment of a guide would encourage users to have more conversations. This should be reflected in the number of inquiries they address to the objects.

**H1**. The average amount of questions per object in a tour will be greater during the *objects* embodiment modality of the conversational agent.

We expected that the amount of informational topics the participants were exposed to, would be directly linked to the amount of questions they asked.

**H2**. The number of questions asked per participant will be in direct correlation to the amount of topics they remember hearing about during their visits.

Lastly, in the evaluation of the different embodiment modalities, we assumed that participants would value each differently. However, the exact nature of the results for the different attributes being assessed was unpredictable, given that personal preferences could play a role. Thus, our hypothesis is formulated undirected.

**H3**. The mean scores for SUS, UEQ Scales and Social Presence will be different dependent on the conversational agent embodiment modality.

### 4.6 Participants

29 participants (16 male, 12 female, 1 diverse) between 23 and 35 years of age (Mean (M) = 27.38, Standard Deviation ($\sigma$) = 3.70) took part in the experiment. All of them were recruited from the student and research body of the Media Informatics department of the university and related programs. 14 participants declared being regular HMD users, 12 were familiar with HMDs and 3 had no prior experience with them. 12 of the participants stated not using voice assistants, 12 only occasionally, and 5 use them very often. All participants were compensated with a 10 Euro voucher.

## 5 Results

In this section, we present the results of our study, starting by the quantitative data gathered and following with the qualitative data obtained from the final questionnaire.

### 5.1 Quantitative Data

For our quantitative analysis, we discarded the data from one of the 29 participants, who stated having a phobia of puppets and animated objects. By their own statement, they felt unease during the animated objects part of the study and did not consider themselves able to make a fair comparison between embodiments. Their comments are later considered in our qualitative analysis.

Considering our hypotheses, we analyzed the effects of embodiment modality on the collected quantitative measures concerning engagement (number of questions) and questionnaires scores. Given the size of our sample, we performed Shapiro-Wilk tests, which could not ensure the normality of the sample data. Consequently, we conducted Wilcoxon signed-rank tests and when finding significance, calculated effect sizes using rank-biserial correlation coefficient. We interpret the size of these effects using Cohen's [11] r thresholds: .10 (small), .30 (medium), and .50 (large).

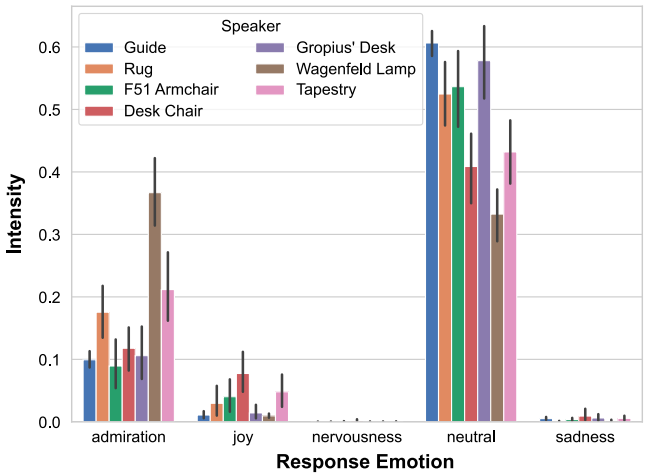


Figure 4: Automatically detected intensities of emotions transported by responses of the embodied conversational agents.

### 5.1.1 Conversation Analytics

In total, the participants voiced 577 utterances. From these, we excluded commands, incorrectly transcribed utterances, and those uttered after the tour's closing message, resulting in 402 questions for further analysis. Table 2 lists detailed statistics about these questions. 210 questions addressed the *objects* and 192 the *guide* embodiment. Among the latter, 37 were asked before any object was selected, thus only 155 were asked during the actual tour.

**Number of questions (H1):** A Wilcoxon signed-rank test showed that participants asked significantly more questions to the conversational agents embodied by objects than to the agent embodied by the humanoid guide during the curated tour ($Z = -2.544$, $p = 0.011$, $r = 0.153$). However, including questions that were asked before the tour showed no statistically significant difference ($Z = -1.165$, $p > 0.05$). Hence, **H1** is only partially supported.

To analyze if individual objects evoked more questions across conditions, we conducted Wilcoxon signed-rank tests of which the results are detailed in Table 2. Only the Wagenfeld Lamp displayed significant differences ($Z = -2.7308$, $p = 0.039$, $r = 0.124$), but a slight trend towards the object embodiment was observed. To put this into perspective with objects' expressed emotions, we conducted an emotion detection experiment. We used a RoBERTa model [25] that was fine-tuned on GoEmotions [12], a Reddit dataset with about 211,000 multi-labeled posts annotated with 28 emotions. From these emotions, we selected five that should reflect the emotions that we induced into the agents, which are *admiration*, *joy*, *nervousness*, *neutral*, and *sadness*. This emotion classifier labels each of these five emotions on the test split of GoEmotions with $F1 = 0.70$, $F1 = 0.60$, $F1 = 0.21$, $F1 = 0.65$ and $F1 = 0.55$, respectively.

Figure 4 presents the detected emotion intensities of responses (i.e., the class probabilities given by the classifier) from all conversational agents. According to a Kurskal-Wallis test (emotion intensities were not normally distributed), each of the five emotions showed significant differences between the generated responses from our conversational agents. The guide showed the least emotions, which is reflected in the neutral probability of responses being significantly higher for the guide than for the desk chair, rug, tapestry and Wagenfeld lamp ($p < 0.05$). The Wagenfeld lamp was overall the least neutral or joyful, but significantly exceeds all other agents in terms of admiration. In this particular case, the high admiration responses are in fact self-admiring responses due to the induced arrogance. Furthermore, the induced "funniness" of the desk chair shows a significant effect in comparison to the guide, F51 armchair, Gropius' desk, rug and Wagenfeld lamp even though the overall joy intensity is rather low. A similar but insignificant effect is visible for the tapestry. Moreover, the induced seriousness of the Gropius' desk translated well into the neutrality of the responses. Unfortunately, the negative emotions were not picked up by the agents. The nervousness that should be reflected by the shy F51 armchair and the sadness induced into the rug showed very low intensities.

Table 2: The sum (Σ), mean (M), and standard deviation (σ) of questions asked by participants in both conditions. The *p*-values are calculated by a Wilcoxon signed-rank test between conditions.

| | **Guide** | | | **Objects** | | | **Significance** | |
|---|---|---|---|---|---|---|---|---|
| | Σ | M | σ | Σ | M | σ | *p*-value | *r* |
| All | 192 | 6.9 | 4.4 | 210 | 7.5 | 4.6 | 0.244 | 0.271 |
| During Tour | 155 | 5.3 | 3.2 | 210 | 7.2 | 4.8 | **0.011** | 0.153 |
| On Selected Object | 143 | 4.9 | 2.4 | 208 | 7.2 | 4.7 | **0.002** | 0.110 |
| *While object is selected* | | | | | | | | |
| F51 Armchair | 28 | 2.0 | 1.2 | 32 | 2.3 | 1.4 | 0.506 | 0.162 |
| Tapestry | 24 | 1.7 | 0.7 | 36 | 2.6 | 1.9 | 0.136 | 0.157 |
| Wagenfeld Lamp | 21 | 1.5 | 0.8 | 44 | 3.1 | 2.5 | **0.039** | 0.124 |
| Gropius' Desk | 33 | 2.4 | 3.4 | 31 | 2.2 | 1.6 | 0.297 | 0.205 |
| Rug | 25 | 1.8 | 1.0 | 34 | 2.4 | 1.6 | 0.159 | 0.205 |
| Desk Chair | 24 | 1.7 | 1.7 | 33 | 2.4 | 1.6 | 0.229 | 0.152 |

Further analysis went into the type of questions users posed to each embodiment. We annotated the questions with an adapted version of the schema by Kiesel et al. [21]. An important difference we state in our study, is that 96.86% of the questions to the animated objects, excluding those directly asking about their creator (e.g., "Why did Walter Gropius go to the U.S.?"), used the second person ("What's your size?"). Regarding the topics of the questions, the guide-embodied agent answered 49 requests (25.52% of the 192 total questions) about topics or objects other than the one selected, whereas the objects only answered 2 questions that were not about themselves or their creators. This results in 143 and 208 *questions on selected object* for each embodiment, respectively, showing a significant yet small positive effect of objects ($Z = -3.036$, $p = 0.002$, $r = 0.110$).

**Remembered topics (H2):** From the list of 18 topics given, participants remembered hearing between 4 and 12 of the topics ($M = 9.00$, $\sigma = 2.58$). We compared these answers to the amount of questions the users asked, and found a low value of the Pearson correlation coefficient between them ($r = 0.11$). Thus, we found no evidence that supports **H2**.

### 5.1.2 Evaluation Scores

We calculated a social presence score as the sum of the individual Likert scales, as proposed by Bailenson et al. [5]. The mean score for the guide was 2.50 ($min = -11$, $max = 13$, $\sigma = 5.25$) and for the animated objects 2.79 ($min = -7$, $max = 12$, $\sigma = 5.85$). No significant difference between embodiments' scores could be determined. Analysis of the five individual questions adjusted with Bonferroni show significantly higher scores for animated objects ($Z = -3.104$, $p = 0.002$, $r = 0.135$) in question 4: "The [guide OR objects] appeared to be sentient, conscious, and alive to me". Figure 5 shows the distributions for each question.

For the SUS, we obtained an average score of 86.43 ($min = 60$, $max = 100$, $\sigma = 10.72$) for the guide, and 88.48 ($min = 70$, $max = 100$, $\sigma = 9.51$) for the animated objects. A Wilcoxon signed-rank test showed no significant effect of embodiment on the scores.

The general UEQ scores showed positive results across scales (Figure 6). Compared to the UEQ proposed benchmark, the scales of Attractiveness, Perspicuity, Stimulation, and Novelty obtained means on the Excellent range, while Efficiency and Dependability

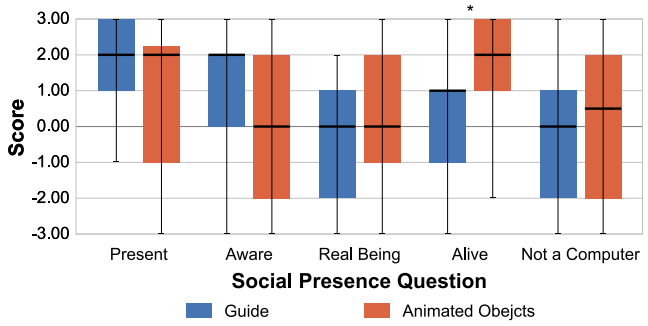


Figure 5: Box plots displaying the scores for the individual questions of the social presence questionnaire. Each question has been labeled with the according salient quality it assesses. * indicates $p < 0.05$.

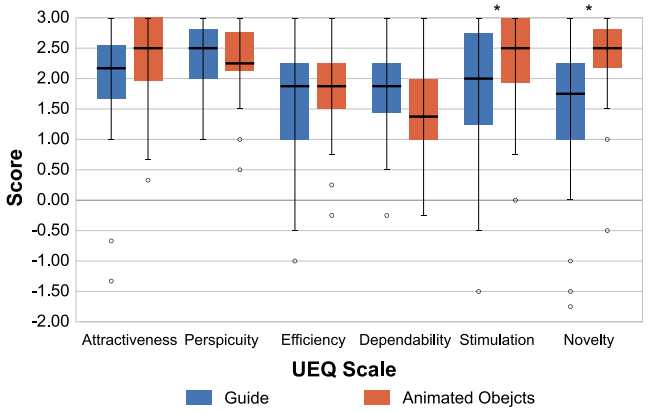


Figure 6: Box plots displaying the scores for each of the six scales in the UEQ per embodiment. * indicates $p < 0.05$.

had means in the Good range. When comparing subscale scores by condition, significant differences were found for the Stimulation ($Z = -3.783$, $p < 0.001$) and Novelty ($Z = -3.575$, $p < 0.001$) but their effects were very small ($r = 0.011$ and $r = 0.074$ respectively).

**SUS, UEQ, and Social Presence (H3):** Given that significance was only found in certain attributes within the three types of assessments, **H3** is only partially supported.

#### 5.1.3 User Preferences

Asked what version the participant had enjoyed more, 20 (71.43%) favored the objects embodiment, 7 the guide (25.00%) and 1 stated no preference. In terms of the learning outcomes, 9 (32.14%) believed the guide provided better results, 9 (32.14%) thought that of the objects and 10 (35.71%) found no difference between them.

### 5.2 Qualitative Data

We used open coding to analyze the qualitative data from the final questionnaire, which asked participants about their preferences for enjoyment and learning. In addition to their preferences, participants were requested to write a justification for their choice.

Most answers favored the objects. Among the reasons, several mentioned the "fun", "playful", or "exciting" nature of the experience (Comment Occurrence (CO) = 12). Participant (P) 09 contrasted this version to the guide's as "[being] in a dream where objects start talking and moving. It creates a better surprise and [caught] my attention." Many found the animations appealing and enticing (CO = 5). P02 highlighted the sync of movement and voice as "an amazing detail." Moreover, some users thought this unusual approach was attractive and interesting (CO = 3), yet others felt that the animation and novelty distracted from the information (CO = 3). P06, an admirer of the original room, stated that the animated objects detracted from the importance of the room's design.

In contrast, the guide was perceived as a more conventional solution; some found this rather boring (CO = 1) while others welcomed its familiarity, human-likeness and found it thus easier to ask questions (CO = 5). P07 wrote: "It feels easier to adjust to one 'persona' and it gets clearer during the time, what to expect from 'her'." This embodiment was also more helpful in guiding participants from object to object and was less distracting when talking (CO = 4). The guide's information was seen as valuable, objective, and neutral (CO = 4), whereas the objects' narratives were considered subjective yet interesting and succinct (CO = 3). The latter approach was appreciated for sparking curiosity and focusing on personal stories (CO = 6). The variety of voices and personalities was also mentioned by participants (CO = 6) as a reason for their increased engagement.

A few participants mentioned a sense of eerie from both embodiments. The former was critiqued by a lack of facial gestures and overall looks (CO = 3), while two users thought talking to objects was too unusual a situation and even frightening (CO = 2).

Most participants enjoyed the experiment and found it challenging to choose between the two versions. Some expressed their desire to see a combined version that integrates both the guide and animated objects, recognizing the unique advantages each modality provides.

## 6 Discussion

With the help of conversational agents, participants explored and learned effectively in our virtual museum. Both embodiment modalities fared similarly well, but some nuances were observed.

In this section, we draw our main findings from the collected data and state our recommendations for the design of agent embodiments.

### 6.1 User Preference

The majority of participants preferred the animated objects approach over the guide. Its novelty was often mentioned as a reason for that, which is supported by the results in the novelty scale of the UEQ. Similarly, higher values for the animated objects in the stimulation scale correspond to the participants' praises on the animations, their different characters, and their subjective viewpoint.

In terms of induced emotions, we could determine that the answers of the agents showed indeed different characteristics. We observed that more histrionic characters, like the lamp or the tapestry, engaged users longer, although significance could only be reported for the lamp. Participants asked the animated objects a larger amount of questions involving their thoughts and feelings (for example: "Do you like it here?", "Do you not feel part of the collection?") than the guide, who was perceived as monotonous in comparison. Thus, users understood the experience as having conversations and getting to know different "people", leading many to feel that the animated objects embodiment was more fun and interesting.

Conversely, some participants who said they liked the animated objects better felt that the guide was a better option for learning, with the main reasons being the objectivity of the information and fewer distractions. Others felt that the familiarity of a guide figure was calming and preferable. This leads us to suggest that although animated objects with induced emotions as embodiments provide an exciting experience for users, a balance with more conventional figures, as one of a guide who provides objective information, may be desirable in virtual learning environments.

### 6.2 Usability and Presence

Both embodiment conditions scored similar satisfactory results in the SUS questionnaire, suggesting that the embodied conversational agent in both modalities provides an easy-to-use interface to access the information of the virtual room.

In terms of social presence, both forms yielded similar positive average scores. However, further examination into the individual questionnaire items reveals contrasts between the embodiments, suggesting that each condition has varying effects, but that the differences cancel each other out in the combined score.

In particular, the question regarding the agent being "sentient, conscious and alive" had significantly different answers, with the animated objects scoring higher. We believe that this stems from the personal nature of the conversations with the animated objects and the noticeable differences of their perceived emotions, compared to the objective and balanced tone of the guide. This would suggest that the form or looks of the embodiment, in particular whether it is a humanoid or an object, are not as influential in the perception of consciousness in an agent as are combinations of distinguishable emotional characteristics.

### 6.3 Diversity in Experiences

The analysis of the participants statements shows that although they agreed on certain characteristics of the embodiments, their positive or negative judgment of these characteristics differs across individuals. For example, the guide being a humanoid figure was compared to a common museum guide. For some people this meant a familiar situation and therefore one where they felt more comfortable. Others found the permanent presence of the figure unsettling, as if they were being watched. Similar differences in acceptance were observed for the animated objects: most participants thought their movements were enticing, while there was a participant that found them unsettling (albeit being what we considered an outlier case).

We believe these to be the natural diversity of experiences and interpretations that can be expected from any group of users. Therefore, we recommend giving users a certain amount of control over the choice of embodiment they shall interact with while exploring virtual museums. Means to exert this control would include the choice of embodiment, displaying guides on demand or during guidance-critical moments, and giving the user agency in the activation and deactivation of animated object as conversational partners.

### 6.4 Engagement and Content per Embodiment

In terms of engagement, our data reveal that 15 out of 28 participants asked 12 or more questions, doubling the proposed minimum of one question per item in the tour. In our question analysis, we reported that many questions directed at the guide did not concern the objects proposed in the tour, but rather general information about the room's design and history, as well as other less prominent objects. In contrast, the animated objects were mostly required to talk about themselves or occasionally about their relationship with other objects. This pattern suggests that the third-person perspective of humanoid guides might allow users to inquire more freely about a virtual space, whereas animated objects draw a user's attention to their own particular subject. It should be noticed that a large amount of the questions to the objects were of subjective character, asking about their thoughts and feelings and not factual information. At first glance, this might seem as an undesired effect for museum curators. However, our study showed that conversational agents, when prompted correctly, are able to deliver playful yet informational responses that engage the user with relevant content.

### 6.5 Limitations

Our study focused primarily on user experience and ease of use to evaluate agent embodiments. To simplify the evaluation process, we measured engagement based solely on the number of questions asked by participants. While this approach is practical, it may not capture all the nuances of user engagement. In addition, we did not assess participants' knowledge acquisition. This was partly due to potential overlap in information between conditions, which could complicate the analysis, and partly to avoid influencing users' intrinsic engagement. As another interesting possibility, we considered adding a realistically embodied tour guide; however, including such a condition would have required shortening the tours and expanding the number of conditions, which would have broken the study design and possibly diluted the results.

The use of text-to-speech synthesis resulted in the emotional expression of our objects happening mainly through word choice rather than intonation. However, as our emotion analysis shows, word choice alone cannot express all emotions. As a result, objects that were shy or sad simply gave short answers with normal intonation and thus appeared less expressive.

## 7 Conclusion

Our work fosters an enjoyable and educational virtual museum experience with an intuitive voice interface that allows users to retrieve background information in social conversations with a conversational agent. We present "speaking object" as a novel, previously unexplored embodiment concept for conversational agents and compare it to a stylized humanoid guide. For the animism-based approach, six objects were individually animated to emphasize their expressiveness and uniqueness when speaking. In addition to these visual cues, the conversational agent was instructed to integrate specific emotions into their responses to give each object its own persona through verbal cues. For the guide's embodiment, an artistic style was selected that matched the theme of the museum room.

A user study with 29 participants revealed that both the speaking objects and the guide were perceived as conscious conversational partners with decent social presence scores. The voice interface achieved high scores for usability and was well accepted by all participants. Notably, the innovative concept of speaking to objects enhanced the user experience significantly in terms of stimulation and novelty, and was generally preferred over the humanoid guide. In terms of engagement, both embodiments elicited a similar number of questions overall. Yet further evaluation showed nuanced differences. Our data indicate a slight increase in questions directed at objects that expressed stronger emotions, suggesting that more expressive personalities may raise engagement.

While our study yielded promising results for both agent representations, we observed a tendency for general questions related to the space to be directed more towards the guide, and more specific, object-related and even personal questions to the animated objects. This pattern suggests that a synergistic combination of both approaches could optimally stimulate interest and foster even greater interactive engagement. Moving forward, we plan to merge these approaches and explore the possibility of dynamically switching between both embodiment forms. This flexibility is crucial as some users may have strong preferences or aversions, including fear, towards certain embodiments.

Another observation was that some users asked very few or off-topic questions, highlighting the need for better cues towards relevant information. In the future, we want to explore the use of visual nudges to guide users effectively through other virtual knowledge spaces. Finally, we intend to evaluate different forms of object embodiment. Inspired by findings of Kim et al. [23], which showed that gestures and locomotion increased confidence and social presence for humanoid agents, we aim to incorporate anthropomorphic features such as eyes and a mouth into non-humanoid objects. We believe these features will enhance expressiveness and significantly boost engagement by displaying a broader range of emotions.

## Acknowledgments

This work is funded by the Thuringian Ministry for Economic Affairs, Science and Digital Society under grant 5575/10-5 (MetaReal) and the German Federal Ministry of Education and Research (BMBF) under the grant 16SV8716 (Goethe-Live-3D).

## REFERENCES


[1] T. A. Agency. Museums audience report: What audience finder says about audiences for museums, 11 2018. Accessed: 2024-03-28.

[2] C. R. Agnew. Feeling close to a crab-thing in virtual reality: Does avatar appearance always matter in forming meaningful connections? a case study. *Frontiers in Virtual Reality*, 3:889247, 2022.

[3] V. W. Anelli, T. D. Noia, E. D. Sciascio, and A. Ragone. Anna: A virtual assistant to interact with puglia digital library (discussion paper). volume 2400, 2019.

[4] R. K. Atkinson, R. E. Mayer, and M. M. Merrill. Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, 30(1):117–139, 2005.

[5] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7):819–833, 2003.

[6] F. Biocca, C. Harms, and J. Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*, pages 1–9, 2001.

[7] M. Botvinick and J. Cohen. Rubber hands 'feel'touch that eyes see. *Nature*, 391(6669):756–756, 1998.

[8] J. Brade, M. Lorenz, M. Busch, N. Hammer, M. Tscheligi, and P. Klimant. Being there again–presence in real and virtual environments and its relation to usability and user experience using a mobile navigation task. *International Journal of Human-Computer Studies*, 101:76–87, 2017.

[9] J. Brooke. Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 11 1995.

[10] L. E. Buck, S. Chakraborty, and B. Bodenheimer. The impact of embodiment and avatar sizing on personal space in immersive virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2102–2113, 2022.

[11] J. Cohen. *Statistical power analysis for the behavioral sciences*. Routledge, 2013.

[12] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi. GoEmotions: A Dataset of Fine-Grained Emotions. In *58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

[13] J. Feine, U. Gnewuch, S. Morana, and A. Maedche. A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, 132:138–161, 2019.

[14] A. Felnhofer, T. Knaust, L. Weiss, K. Goinska, A. Mayer, and O. D. Kothgassner. A virtual character's agency affects social responses in immersive virtual reality: a systematic review and meta-analysis. *International Journal of Human–Computer Interaction*, pages 1–16, 2023.

[15] R. Fribourg, F. Argelaguet, A. Lécuyer, and L. Hoyet. Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view. *IEEE transactions on visualization and computer graphics*, 26(5):2062–2072, 2020.

[16] G. Gorisse, O. Christmann, S. Houzangbe, and S. Richir. From robot to virtual doppelganger: Impact of visual fidelity of avatars controlled in third-person perspective on embodiment and behavior in immersive virtual environments. *Frontiers in Robotics and AI*, 6:8, 2019.

[17] C. Gutwin and S. Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11:411–446, 2002.

[18] A. Huang, P. Knierim, F. Chiossi, L. L. Chuang, and R. Welsch. Proxemics for human-agent interaction in augmented reality. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, pages 1–13, 2022.

[19] T. Iachini, Y. Coello, F. Frassinetti, and G. Ruggiero. Body space in social interactions: a comparison of reaching and comfort distance in immersive virtual reality. *PloS one*, 9(11):e111511, 2014.

[20] C. Jicol, C. H. Wan, B. Doling, C. H. Illingworth, J. Yoon, C. Headey, C. Lutteroth, M. J. Proulx, K. Petrini, and E. O'Neill. Effects of emotion and agency on presence in virtual reality. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–13, 2021.

[21] J. Kiesel, V. Bernhard, M. Gohsen, J. Roth, and B. Stein. What is that? crowdsourcing questions to a virtual exhibition. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*, pages 358–362, 2022.

[22] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

[23] K. Kim, L. Boelling, S. Haesler, J. Bailenson, G. Bruder, and G. F. Welch. Does a digital assistant need a body? the influence of visual embodiment and social behavior on the perception of intelligent virtual agents in ar. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 105–114. IEEE, 2018.

[24] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM symposium on virtual reality software and technology*, pages 1–10, 2017.

[25] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.

[26] O. M. Machidon, M. Duguleana, and M. Carrozzino. Virtual humans in cultural heritage ict applications: A review. *Journal of Cultural Heritage*, 33:249–260, 9 2018.

[27] G. Makransky, P. Wismer, and R. E. Mayer. A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation. *Journal of Computer Assisted Learning*, 35(3):349–358, 2019.

[28] A. S. D. Martha and H. B. Santoso. The design and impact of the pedagogical agent: A systematic literature review. *Journal of educators Online*, 16(1):n1, 2019.

[29] R. E. Mayer. Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. *The Cambridge handbook of multimedia learning*, 16:345–370, 2014.

[30] Meta. Oculus LipSync, 2024. `https://developer.oculus.com/downloads/package/oculus-lipsync-unity/`.

[31] Mixamo. mixamo, 2022. `https://www.mixamo.com/`.

[32] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *IEEE Robotics & automation magazine*, 19(2):98–100, 2012.

[33] K. L. Nowak and F. Biocca. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 12(5):481–494, 2003.

[34] N. Ogawa, T. Narumi, H. Kuzuoka, and M. Hirose. Do you feel like passing through walls?: Effect of self-avatar appearance on facilitating realistic behavior in virtual environments. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–14, 2020.

[35] OpenAI. OpenAI Platform, 2024. `https://platform.openai.com/`.

[36] OpenAI, J. Achiam, S. Adler, S. Agarwal, L. Ahmad, and et al. Gpt-4 technical report, 2024.

[37] G. B. Petersen, A. Mottelson, and G. Makransky. Pedagogical agents in educational vr: An in the wild study. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2021.

[38] R. Prentice, A. Davies, and A. Beeho. Seeking generic motivations for visiting and not visiting museums and like cultural attractions. *Museum Management and Curatorship*, 16(1):45–70, 1997.

[39] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever. Robust speech recognition via large-scale weak supervision, 2022.

[40] J. Reinhardt, L. Hillen, and K. Wolf. Embedding conversational agents into ar: Invisible or with a realistic human body? In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 299–310, 2020.

[41] R. Rzayev, G. Karaman, K. Wolf, N. Henze, and V. Schwind. The effect of presence and appearance of guides in virtual reality exhibitions. In *Proceedings of Mensch Und Computer 2019*, pages 11–20. 2019.

[42] S. Schmidt, G. Bruder, and F. Steinicke. Effects of virtual agent and object representation on experiencing exhibited artifacts. *Computers &*

*Graphics*, 83:1–10, 2019.
[43] M. Schrepp. User experience questionnaire handbook, 09 2015.
[44] N. L. Schroeder and O. O. Adesope. A systematic review of pedagogical agents' persona, motivation, and cognitive load implications for learners. *Journal of Research on Technology in Education*, 46(3):229–251, 2014.
[45] N. L. Schroeder, O. O. Adesope, and R. B. Gilbert. How effective are pedagogical agents for learning? a meta-analytic review. *Journal of Educational Computing Research*, 49(1):1–39, 2013.
[46] J. Short, E. Williams, and B. Christie. *The social psychology of telecommunications*. Toronto; London; New York: Wiley, 1976.
[47] R. Skarbez, F. P. Brooks, Jr, and M. C. Whitton. A survey of presence and related concepts. *ACM computing surveys (CSUR)*, 50(6):1–39, 2017.
[48] M. Slater, D. Pérez Marcos, H. Ehrsson, and M. V. Sanchez-Vives. Inducing illusory ownership of a virtual body. *Frontiers in neuroscience*, 3:676, 2009.
[49] J.-P. Stein and P. Ohler. Venturing into the uncanny valley of mind—the influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, 160:43–50, 2017.
[50] Y. Sun, O. Shaikh, and A. S. Won. Nonverbal synchrony in virtual reality. *PloS one*, 14(9):e0221803, 2019.
[51] U. Technologies. AI Navigation, 2024. `https://docs.unity3d.com/Packages/com.unity.ai.navigation@2.0/manual/index.html`.
[52] I. Wang, J. Smith, and J. Ruiz. Exploring virtual agents for augmented reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.
[53] F. Weidner, G. Boettcher, S. A. Arboleda, C. Diao, L. Sinani, C. Kunert, C. Gerhardt, W. Broll, and A. Raake. A systematic review on the visualization of avatars and agents in ar & vr displayed using head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 2023.
[54] J. W. Woodworth, N. G. Lipari, and C. W. Borst. Evaluating teacher avatar appearances in educational vr. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1235–1236. IEEE, 2019.
[55] N. Yee and J. Bailenson. The proteus effect: The effect of transformed self-representation on behavior. *Human communication research*, 33(3):271–290, 2007.
[56] K. Zibrek, E. Kokkinara, and R. McDonnell. The effect of realistic appearance of virtual characters in immersive environments-does the character's personality play a role? *IEEE transactions on visualization and computer graphics*, 24(4):1681–1690, 2018.