Bringing Everyday Objects to Life in Augmented Reality with AI-Powered Talking Characters

Nozomu Iwai
Preferred Networks
Tokyo, Japan
iwai.nozomu.r5@elms.hokudai.ac.jp

Fabrice Matulic
Preferred Networks
Tokyo, Japan
fmatulic@preferred.jp

Figure 1: AI-powered AR characters embodying everyday objects. (a) User wearing an AR headset interacting with characters representing various kitchen utensils. (b) User engaged in conversation with characters for a pair of sneakers and a fizzy drink, through a smartphone. (c) A mouse character commenting on the environment, demonstrating context awareness (speech bubble added for illustrative purposes). (d) A toothpaste character expresses anger in reaction to the user saying it is too minty, illustrating dynamic emotional responses.

Abstract

We explore the potential of Augmented Reality (AR) and AI to turn everyday household objects into interactive, talking characters. With an AR headset or a smartphone, the user views these animated characters overlaid on physical objects and participates in dynamic dialogues with them thanks to language, vision and speech AI. These AR characters are imbued with unique personalities reflective of their object's function and characteristics as well as respond to environmental changes and user interactions, resulting in rich and relatable conversations grounded in the real world. A preliminary user evaluation with a HoloLens 2 and an Android smartphone confirms the potential of our approach. Our ultimate goal is to support fully automatic augmentations of any physical object with engaging, personality-rich and context-aware AR characters to enable new styles of conversations.

CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**; • **Computing methodologies** → **Natural language processing**.

Keywords

Mixed Reality, Large Language Models, AI Conversational Agents

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3719978

1 Introduction

Humans have a tendency to attribute human characteristics and communication abilities to inanimate objects, such as talking to plants, believing it will aid their growth, or naming personal possessions to instil a sense of companionship [16]. Such kind of anthropomorphism is not merely whimsical, it can also enhance emotional well-being and comfort [13], as well as mitigate feelings of loneliness [5]. The emotional attachments formed with household appliances, for instance, can elevate mundane chores into more enjoyable activities [17]. When explicitly given facial features, this bond can be further enhanced, as illustrated by the popular Henry vacuum cleaner in the UK, with its distinctive design featuring cartoon eyes and a smiling face contributing to making it a cultural icon [15].

Technology has progressively tapped into this anthropomorphic tendency. Modern personal assistants like Siri and Alexa, alongside smart home appliances, leverage natural speech commands to enable more intuitive interactions with household devices. With the advent of generative AI, in particular large language models (LLMs), these devices are now capable of engaging users in more fluid and less constrained verbal interactions [6]. These advances underscore a broader trend towards embedding intelligent, conversational characteristics into the objects and environments we regularly interact with [2].

The convergence of AI, AR, and our inherent tendency to anthropomorphise objects present a unique opportunity to further enrich how we interact with household items beyond simple functional agents that respond to information requests and commands [11]. In this late-breaking work, we propose a novel approach and system that leverage AR and AI to transform ordinary objects into interactive, talking characters with human-like characteristics. With an

AR headset or a smartphone, users view animated characters overlaid onto physical objects, and engage in dynamic natural dialogues with them (Figure 1), thanks to advanced language, vision, and speech models. These AR characters are not mere chatbots, they are endowed with personalities reflective of the functions and attributes of their associated physical objects as well as are responsive to environmental changes and user interactions. This contextual information obtained from large vision models and detection modules is seamlessly integrated into the dialogue experience and influences the characters' responses. Furthermore, characters converse not only with the user but also between themselves, generating original conversations potentially providing new added value compared to current mostly one-on-one interactions with voice assistants.

We implement a proof-of-concept prototype, which we deploy on a HoloLens 2 and an Android smartphone. A preliminary user evaluation with common household items reveals the potential of our approach to make interactions with everyday objects more enjoyable and emotionally engaging. We discuss technological challenges and future avenues to explore in order to ultimately be able to automatically transform any object in any environment into interesting, personality-rich and context-aware characters.

2 Related Work

Previous work explored various aspects of human-object interaction, anthropomorphism, and AR conversation agents.

Reddy et al explored the potential of everyday objects to engage in conversations using a more-than-human approach, considering various themes in human-object relationships to inspire designs of conversational agents in domestic settings [11]. While the work is entirely speculative (humans respond for the objects), it highlights the untapped potential of designing voice interfaces to foster richer, more meaningful engagements with everyday objects beyond purely utilitarian interactions.

Anthropomorphism, the attribution of human-like traits to objects, is a key psychological phenomenon supporting this concept. Saunderson and Kuo surveyed 125 adults about anthropomorphism and found that making objects more relatable in this manner helps people form social connections and cope with loneliness [14]. Osawa et al. showed that attaching human-like robotic body parts to household objects influences users' perceptions and enables more natural interactions [10]. Similarly, Marchetti et al. demonstrated that playful, anthropomorphic design elements in a zoomorphic floor cleaning robot facilitated its acceptance among care home residents with dementia [8]. Garcia et al. conducted a study of speaking objects in virtual museums and reported that participants prefer talking exhibits over humanoid guides, as the former encouraged detailed questions about the objects' stories and created more stimulating exchanges [3]. This notion of imbuing objects with human-like qualities has also been explored for games, with titles like "Date Anything" allowing players to interact with humanised versions of household objects in a virtual home [7].

Other research has focused on the visual representation of AR conversational agents. Reinhardt et al. examined how different levels of realism in visual representation affect the usability of AR agents, though without association to physical objects [12]. Zhu et al. explored the impact of representations of human and symbolic avatars in mixed reality on user interaction with GPT-3-powered agents, finding that human avatars enhance user recall and experience [19].

Recent advancements in AI have further enhanced the possibilities to create more dynamic and natural AR characters. Nagano et al. showcased a prototype system around a travel book powered by language and speech AI, with study participants wanting to build human-like connections with the book and expressing a desire to talk to their personal objects [9]. XR-Objects enables interaction with physical objects via pre-defined AR menus and simple speech queries, but lacks rich, character-driven engagement, real-time conversational capabilities and environment awareness [2].

Building on this prior work, we propose augmenting everyday objects via AR with human-like AI-driven characters that possess unique personalities based on the objects they represent and which dynamically react to environmental changes. Our approach advances beyond purely digital chatbots and informational AR applications that engage in one-to-one exchanges with the user, as we aim to offer richer, context-aware, and personalised multi-agent experiences anchored in the physical world.

3 AR Characters

Our aim is to generate digital faces for everyday objects of interest within the user's environment and facilitate dynamic conversations between the user and these characters, as well as between characters themselves. These conversations should further reflect the surrounding context and the user's interactions. Achieving this requires a sophisticated system capable of interpreting multiple input sources, including user speech, AR device pose, visual features of objects and the environment, and other user interactions (gaze, touch, gestures etc.). While low-latency realtime large language models have recently appeared to support natural conversations with chatbots, including using speech, current models and engines are mainly designed for 1-on-1 dialogues with the user and lack the ability to seamlessly integrate multi-agent conversation and environmental cues.

To address this limitation, we design an architecture incorporating dedicated tracking and detection modules that process diverse input sources and generate detailed context prompts capturing this rich context for a realtime large language model (LLM) (Figure 2). The LLM is instructed to continuously stream dialogue text in a format similar to a play script showing character names and their spoken lines (Figure 5). For our implementation, we use OpenAI's models and realtime API, which is stateful and automatically maintains conversation history across multiple turns.

3.1 Physical Object Tracking

To anchor digital faces to objects, the AR device must recognise objects of interest in the user's environment and track their 6 degrees of freedom (DoF) pose in 3D space. While real-time 6 DoF tracking of unknown objects is possible with RGBD cameras [18], achieving this with monocular RGB input without prior models is challenging. Accuracy and robustness are generally lower compared to systems using depth sensors or pre-registered models, especially for objects with minimal texture or in poor lighting. Since common AR devices like smartphones and consumer smart glasses may lack depth

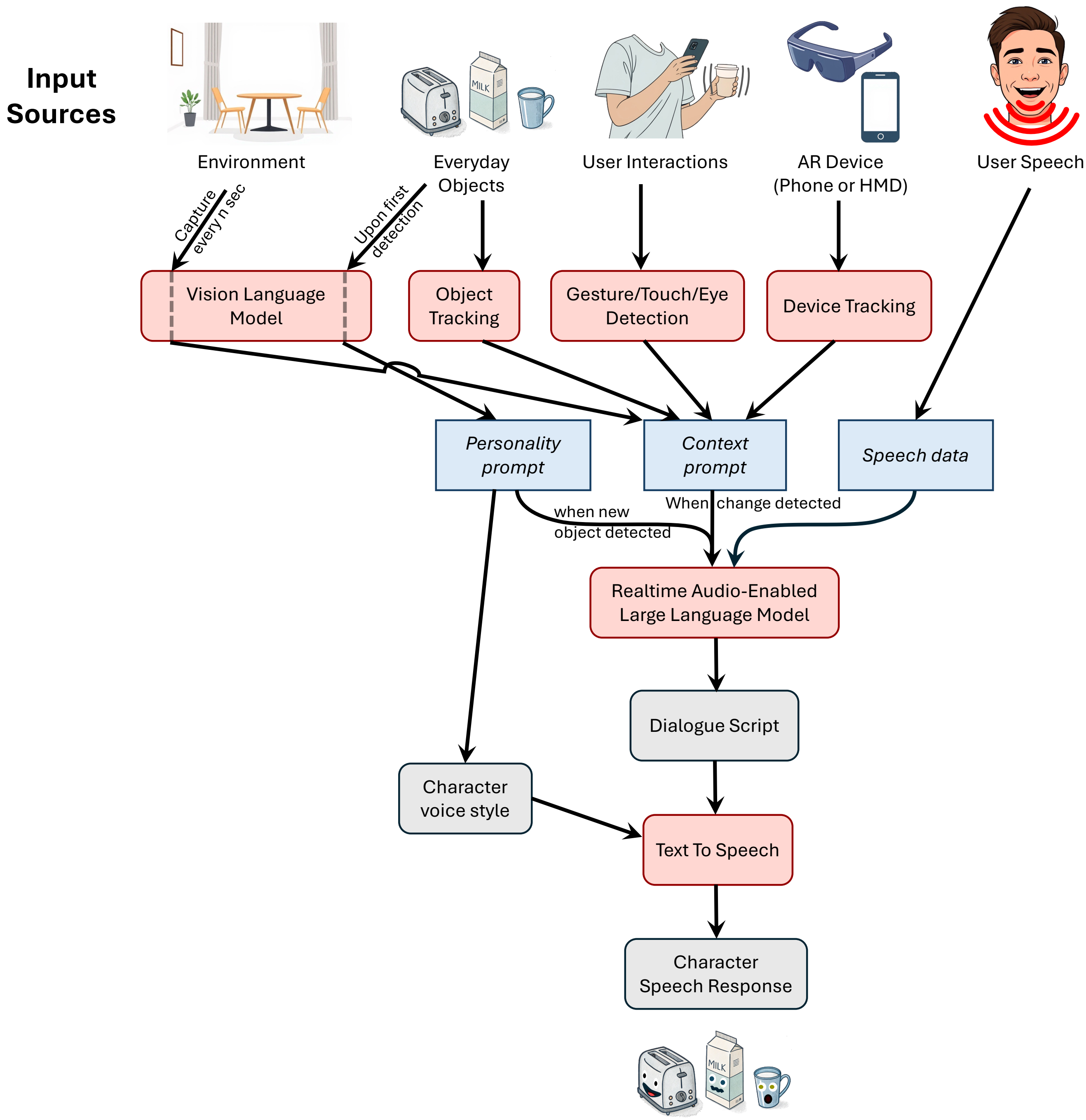

Figure 2: Flow diagram illustrating the generation of AR characters' responses based on their personalities and various input sources, including user speech, AR device pose, user interactions and image captures of the objects and the surrounding environment. This diverse input data is interpreted by various AI and tracking modules that create corresponding context prompts. Along with the user's speech data, character personality and context prompts are fed to a realtime large language model, which produces a dialogue script. This script is subsequently converted into speech using the character's associated voice style.

sensors, we implement two distinct object tracking approaches using only RGB images, each with its own trade-offs in terms of supported DoF and required setup:

1. **6 DoF Tracking with Pre-registration**: For robust 6 DoF tracking from monocular RGB images, we leverage AR frameworks like Vuforia. These frameworks require prior knowledge of the object's geometry and texture, typically obtained through a pre-registration process. This involves creating a "model target" (a 3D model of the object) using either a 3D scanning application or a CAD program. Alternatively, an "image target" (a 2D image of the object) can be registered, though this limits tracking to a single visible surface. While pre-registration enables accurate 6 DoF tracking, it introduces a setup phase that hinders spontaneous interaction with arbitrary objects in the environment.
2. **3 DoF Tracking without Pre-registration**: To enable more spontaneous interactions, we also implement a 3 DoF tracking method that does not require pre-registration. This approach utilises a 2D object detection engine (MediaPipe [4]) to identify objects of certain categories within the camera's view. By integrating this with AR frameworks like ARKit and ARCore, the object's 3D position can be estimated [2]. However, this method cannot track the object's orientation, preventing the character's face from being locked to a specific 3D surface and rotating with it. Additionally, this technique struggles to track moving objects (e.g. when picked up by the user), limiting the system's ability to support dynamic physical interactions.

In summary, our system integrates both approaches: Vuforia for accurate 6 DoF tracking with pre-registered objects, and 3 DoF tracking without pre-registration for more spontaneous engagement. This allows us to explore the trade-offs between tracking capabilities and ease of use in different interaction scenarios.

3.2 Context-Dependent Speech-To-Speech

Our system centres around a realtime large language model (LLM), specifically OpenAI's Realtime API (version 2024-12-17). This LLM drives the conversational interactions by generating dialogue scripts for the characters, as illustrated in Figure 5. It achieves this by processing two types of inputs: user speech and contextual information provided as text prompts. The generated dialogue is then converted into speech using OpenAI's Audio API (tts-1 model), which gives voice to the characters.

While user speech forms the primary input, the LLM also relies on a rich set of contextual cues to generate relevant and engaging responses. These cues include diverse information about the characters' personalities, user actions and changes in the surrounding scene. To capture and interpret this information, we use dedicated modules that monitor several input sources and translate them into textual descriptions suitable for the LLM.

3.2.1 Character Personalities. A key innovation in our system is the dynamic generation of unique personalities for each character. These personalities are automatically derived from the visual appearance and inferred function of the corresponding object (Figure 4). For instance, a pack of biscuits might be assigned a "sweet" personality and a cactus a "prickly" one.

This process unfolds as follows: when a new object is detected, an image is captured and analysed by a vision language model (OpenAI's GPT-4o-mini, which has vision capabilities). This model identifies the object and generates a detailed description of its potential personality traits (see Appendix for examples of prompts and outputs). This description is then incorporated into the prompt provided to the real-time LLM responsible for generating dialogue.

To further enhance the character's persona, the system also selects an appropriate voice style from a set of options (currently, the different voices supported by OpenAI's voice API). This selection is guided by the personality description, ensuring that the character's voice complements its personality traits.

3.2.2 Dynamic Environment Awareness. To further enhance the realism and interactivity of the AR characters, the system integrates dynamic awareness of the environment. This is achieved through the vision language model, which not only analyses the objects themselves but also their surroundings.

Upon initialisation, the model captures an image of the environment surrounding each object and generates a textual description. This provides initial context for the conversation (Figure 11). Furthermore, the model periodically analyses the environment for changes. Using a before-after comparison approach, it compares the current view with the previous snapshot and identifies any significant differences. These differences are then relayed to the LLM as part of the context prompt, allowing the characters to react to environmental changes in their dialogue.

In addition to visual analysis, the system leverages the object tracking module to detect the appearance and disappearance of characters. This information is also included in the context prompt, enabling characters to acknowledge new participants with greetings or comment on the departure of others. For instance, by utilising Vuforia's object tracking status, the system can detect when an object becomes occluded or leaves the scene, triggering appropriate reactions from the remaining characters.

3.2.3 User Interactions. Our system allows users to engage with the AR characters through various actions. These user interactions are captured and incorporated into the context prompt, enabling the characters to respond in a dynamic and personalised manner.

Our current implementation recognises a range of user actions, including physical interactions like touching or picking up an object, which are detected through hand tracking. Additionally, the system can detect changes in the user's proximity, e.g. recognising when the user approaches or moves away from objects. If the AR device supports eye tracking, the system can detect the user's gaze, discerning whether they are simply looking at an object or intently staring at it. This allows for more nuanced and engaging interactions, as characters can react to the user's attention and physical presence.

Furthermore, touch-enabled AR devices, such as smartphones and tablets, allow for indirect interactions with the characters through the touchscreen. Taps and swipes can be interpreted as "pokes" or "strokes," similar to interactions in simulation games, adding another layer of playful engagement. Each of these user actions generates corresponding event logs that are integrated into the context prompt for the LLM, enabling the characters to dynamically acknowledge and respond to the user's behaviour.

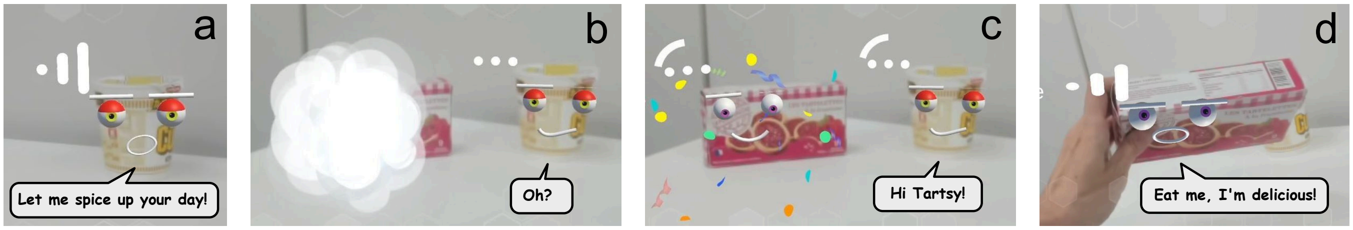

Figure 3: Example scenario of multi-character dialogue and user interaction. (a) The user converses with a character representing a cup of ramen noodle. (b) Upon detecting a new object (a pack of raspberry tartlets), the system displays a cloud animation while generating a corresponding character. (c) The new tartlet character appears and is greeted by the cup noodle character. (d) The tartlet character reacts to being picked up by the user, showcasing context-aware interaction.

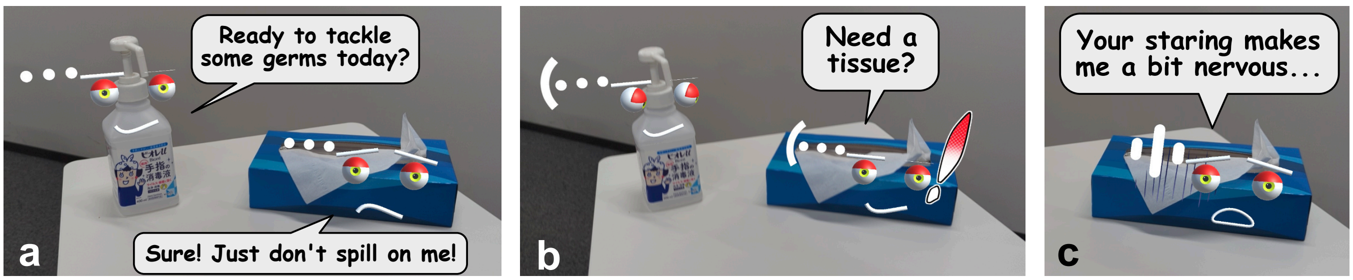

Figure 4: Disinfectant and tissue characters engaging in dialogue based on their associated objects and user interactions. (a) Upon detection, the characters identify the function of their object and spontaneously initiate a conversation about hygiene. (b) The tissue character reacts to being tapped by the user on the phone screen with an exclamation mark and a corresponding verbal response. (c) Using the device's eye-tracking capabilities, the tissue box detects and responds to the user's gaze.

```
"Tartsy":"<Joy>Hi! My name's Tartsy. I hope you're having a sweet time!@"
"Curry Cup":"<Joy>Hi Tartsy! I'm Curry Cup. I feel this is going to be a spicy day!@"
```

Figure 5: Example dialogue script with mood tags generated by the LLM for characters representing a pack of tartlets and a cup of instant curry ramen.

3.3 Character animations

Visual appeal and engaging animations are crucial for creating a compelling user experience with virtual characters. While our long-term goal is to automatically generate character designs and animations through generative AI based on objects' visual features, our current prototype utilises a set of three pre-designed face styles. These styles are randomly assigned to objects upon detection.

To provide visual feedback during the initial object analysis and personality generation, a cloud animation is displayed, subtly suggesting that the system is "conjuring" a character for the object (Figure 3b). Each face style includes four distinct facial expressions that reflect different emotional states (Figure 1d): joy, anger, sorrow, and pleasure. The appropriate expression is dynamically determined based on mood tags embedded within the dialogue script generated by the LLM, as shown in Figure 5.

Furthermore, an exclamation mark symbol pops up above the character's head whenever the user physically touches the object or interacts with it through the touchscreen (Figure 4b). This visual cue, reminiscent of common tropes in cartoons and video games, provides a playful acknowledgment of the user's interaction.

3.4 System Deployment

We develop our prototype system on Unity, which streams LLM requests to OpenAI's servers using a wrapper for its API. We deploy the system on two different AR platforms: a Microsoft HoloLens 2 and a Google Pixel 7 phone.

In our current implementation, each device supports a unique set of user interactions. The HoloLens, with its integrated hand and eye tracking, enables natural and intuitive interactions, such as physically touching objects and making eye contact with characters. Conversely, the phone primarily relies on its touchscreen for interaction, allowing users to indirectly "touch" characters through taps and gestures.

Although custom hand and eye tracking features could also be implemented on the phone, the head-mounted HoloLens offers a key advantage: users can freely interact with objects using both hands, unencumbered by the need to hold a device. In contrast, phone-based interaction requires the user to hold the device, potentially limiting the range and complexity of physical interactions.

4 Preliminary Evaluation

4.1 Experiment Design

To gain initial insights into the feasibility of our concept and users' experience with our prototype, we conducted a preliminary qualitative evaluation with 10 participants (7 male, 3 female, mean age = 36.2 years, SD = 9.2) recruited from our institution. All participants were familiar with and regularly used LLMs such as ChatGPT.

The study was conducted in an office environment using both the HoloLens and the Pixel 7, allowing us to explore the system's usability across different devices. With the HoloLens, we used 6 DoF tracking with two pre-registered objects (a cup noodle and a pack of raspberry tarts) to demonstrate precise character anchoring and interaction. For the phone, we used the 3 DoF tracking method, which allows for spontaneous interaction with a wider range of objects. Participants were invited to select personal objects or objects from their surroundings for this part of the study.

After a brief introduction to the system's features, participants engaged in conversations with the AR characters or observe dialogues between them. Participants used the HoloLens for 10 minutes to interact with the two pre-registered objects, followed by 15 minutes with the phone where they interacted with their personal objects. At the end of the experiment, we conducted semi-structured interviews to gather detailed feedback on their experiences and suggestions for improvement. We further inquired about the types of objects in their homes they would be interested in conversing with if they had access to such an application. Finally, we discussed desired features for this type of AR experience.

4.2 Results

All participants expressed enjoyment and interest in the concept, as shown by their smiles and laughter during the interactions. Three participants explicitly stated their desire to have such an application on their phones. For two participants, the concept evoked associations with Disney-like animation and films like Toy Story.

Participants selected a diverse range of objects for conversation, including shoes, glasses, a mouse, a microphone, various beverages, chocolate bars, a rucksack, scissors, books, a gamepad, gloves, earphones, and pen holders. With one exception, all objects were correctly recognised by the VLM, leading to appropriate personality assignments and relevant conversations. Due to the short and playful nature of the study, most conversations remained casual and involved brief exchanges between the user and the characters. Occasional performance issues with the HoloLens and latency also impacted the conversational flow.

Regarding AR devices, most participants appreciated the freedom of using both hands and directly interacting with objects afforded by the HoloLens. However, the quick and simple interaction provided by tapping on the phone's touchscreen was also seen as a positive. Two participants noted fatigue from holding up the smartphone for extended periods. Additionally, in the 3 DoF tracking condition, some instances of misalignment between the character's face and the object were observed, particularly when the viewing angle changed. This is a limitation of this tracking method.

When asked about the types of objects they would be interested in interacting with in their homes, three participants mentioned plants, and six participants expressed a desire to converse with objects with practical functionality, such as appliances or devices that could explain their usage or provide status updates. This aligns with utilitarian communication scenarios for smart IoT devices, which our system could potentially support. While our current implementation relies on external general LLMs, we envision future integration with customised models provided by manufacturers, which would integrate deeper knowledge of objects and access to sensor data. This could enable more informative and useful interactions, presented through engaging AR characters instead of traditional UIs or audio-only voice assistants.

Interestingly, two participants preferred observing dialogues between characters rather than actively participating in conversations. One participant envisioned comedy scenarios involving seemingly unrelated objects humorously discussing commonalities and differences. Another suggested applying the system to books, where users could engage in conversations with the book's characters via AR avatars. These ideas highlight the potential for entertainment and creative storytelling. Another participant suggested that ongoing conversations between objects could create a sense of presence and companionship, particularly for individuals living alone. In such scenarios, our system could foster a sense of connection and attachment to personal objects through daily interaction and shared experiences.

5 Conclusion and Future Work

This paper introduced a novel AI-driven system that transforms everyday objects into interactive AR characters with distinct personalities, capable of engaging in dynamic dialogues and responding to context. Our preliminary evaluation highlights the potential of this approach to enrich human-object interaction and create more engaging experiences.

In future work, we would like to extend this concept beyond household objects to include larger structures and outdoor environments, allowing users to converse with buildings, vehicles, or landmarks. Furthermore, we believe there is significant potential for creating AI-based characters for comedy shows in the vein of "Annoying Orange" [1], where anthropomorphised food items and objects have humorous exchanges for entertainment purposes. To enhance user agency and personalisation, we plan to explore mechanisms that would give users more control over characters' personalities and designs. This could involve offering a selection of personality profiles and face styles for each object. Finally, we plan to conduct more extensive user studies in natural home environments to gain deeper insights into the long-term impact of interacting with AI-powered AR characters. This will help us refine our system and explore its potential to foster companionship, enhance emotional well-being, and create new relationships with the objects that surround us.

References

[1] Dane Boedigheimer. [n. d.]. The Annoying Orange. https://www.youtube.com/channel/UCi-5OZ2tYuwMLIcEyOsbdRA. Accessed: 2024-12-03.

[2] Mustafa Doga Dogan, Eric J Gonzalez, Karan Ahuja, Ruofei Du, Andrea Colaço, Johnny Lee, Mar Gonzalez-Franco, and David Kim. 2024. Augmented Object Intelligence with XR-Objects. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 19, 15 pages. doi:10.1145/3654777.3676379

[3] Irene Lopez Garcia, Ephraim Schott, Marcel Gohsen, Volker Bernhard, Benno Stein, and Bernd Froehlich. 2024. Speaking with Objects: Conversational Agents' Embodiment in Virtual Museums . In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE Computer Society, Los Alamitos, CA, USA, 279–288. doi:10.1109/ISMAR62088.2024.00042
[4] Google. 2024. MediaPipe. https://developers.google.com/mediapipe. Accessed: January 20, 2025.
[5] Valerie K. Jones, Michael Hanus, Changmin Yan, Marcia Y. Shade, Julie Blaskewicz Boron, and Rafael Maschieri Bicudo. 2021. Reducing Loneliness Among Aging Adults: The Roles of Personal Voice Assistants and Anthropomorphic Interactions. *Frontiers in Public Health* 9 (2021). doi:10.3389/fpubh.2021.750736
[6] Evan King, Haoxiang Yu, Sangsu Lee, and Christine Julien. 2024. Sasha: Creative Goal-Oriented Reasoning in Smart Homes with Large Language Models. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 1, Article 12 (March 2024), 38 pages. doi:10.1145/3643505
[7] Mango Madness Studio. 2024. Date Everything. https://store.steampowered.com/app/2201320/Date_Everything/. Accessed: January 20, 2025.
[8] Emanuela Marchetti, Sophie Grimme, Eva Hornecker, Avgi Kollakidou, and Philipp Graf. 2022. Pet-Robot or Appliance? Care Home Residents with Dementia Respond to a Zoomorphic Floor Washing Robot. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 512, 21 pages. doi:10.1145/3491102.3517463
[9] Keijiroh Nagano, Maya Georgieva, Harpreet Sareen, and Clarinda Mac Low. 2023. The Talk: Speculative Conversation with Everyday Objects. In *ACM SIGGRAPH 2023 Posters* (Los Angeles, CA, USA) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 12, 2 pages. doi:10.1145/3588028.3603643
[10] Hirotaka Osawa, Jun Mukai, and Michita Imai. 2007. Anthropomorphization Framework for Human-Object Communication. *Journal of Advanced Computational Intelligence and Intelligent Informatics* 11, 8 (2007), 1007–1014. doi:10.20965/jaciii.2007.p1007
[11] Anuradha Reddy, A. Baki Kocaballi, Iohanna Nicenboim, Marie Louise Juul Søndergaard, Maria Luce Lupetti, Cayla Key, Chris Speed, Dan Lockton, Elisa Giaccardi, Francisca Grommé, Holly Robbins, Namrata Primlani, Paulina Yurman, Shanti Sumartojo, Thao Phan, Viktor Bedö, and Yolande Strengers. 2021. Making Everyday Things Talk: Speculative Conversations into the Future of Voice Interfaces at Home. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 23, 16 pages. doi:10.1145/3411763.3450390
[12] Jens Reinhardt, Luca Hillen, and Katrin Wolf. 2020. 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* (Sydney NSW, Australia) *(TEI '20)*. Association for Computing Machinery, New York, NY, USA, 299–310. doi:10.1145/3374920.3374956
[13] Kristen Rogers. 2024. Why do people give feelings to inanimate objects? What experts say. *CNN* (7 September 2024). https://edition.cnn.com/2024/09/07/health/empathize-inanimate-objects-anthropomorphize-wellness/index.html Accessed December 3, 2024.
[14] Shane Saunderson and Isabella Kuo. 2022. Why We See Humanity in Everyday Objects. (6 April 2022). https://shanesaunderson.com/2022/04/06/why-we-see-humanity-in-everyday-objects/ Accessed on December 03, 2024.
[15] Nicola Slawson. 2021. Sucks to be him! How Henry the vacuum cleaner became an accidental design icon. *The Guardian* (24 July 2021). https://www.theguardian.com/lifeandstyle/2021/jul/24/how-henry-vacuum-cleaner-became-accidental-design-icon Life and style section.
[16] Karen Stollznow. 2014. *Talking Objects*. Palgrave Macmillan UK, London, 207–216. doi:10.1057/9781137404862_21
[17] Ja-Young Sung, Lan Guo, Rebecca E. Grinter, and Henrik I. Christensen. 2007. "My Roomba Is Rambo": Intimate Home Appliances. In *UbiComp 2007: Ubiquitous Computing*, John Krumm, Gregory D. Abowd, Aruna Seneviratne, and Thomas Strang (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 145–162.
[18] Bowen Wen, Jonathan Tremblay, Valts Blukis, Stephen Tyree, Thomas Müller, Alex Evans, Dieter Fox, Jan Kautz, and Stan Birchfield. 2023. BundleSDF: Neural 6-DoF Tracking and 3D Reconstruction of Unknown Objects. In *CVPR*.
[19] Jiarui Zhu, Radha Kumaran, Chengyuan Xu, and Tobias Höllerer. 2023. Free-form Conversation with Human and Symbolic Avatars in Mixed Reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 751–760. doi:10.1109/ISMAR59233.2023.00090