Takeaways from Applying LLM Capabilities to Multiple Conversational Avatars in a VR Pilot Study

Mykola Maslych
University of Central Florida
Orlando, Florida, USA
maslychm@gmail.com

Christian Pumarada
University of Central Florida
Orlando, Florida, USA
cpuma1824@gmail.com

Amirpouya Ghasemaghaei
University of Central Florida
Orlando, Florida, USA
aghaei.ap@ucf.edu

Joseph J. LaViola Jr.
University of Central Florida
Orlando, Florida, USA
jlaviola@ucf.edu

Figure 1: Avatars used in the study with example feedback types: (a) Friend + *state lights* (one active state among Idle, Listening, Thinking, and Speaking), (b) Clerk + *loading bar* (appears above avatar only during Thinking state), (c) Manager + *no feedback* (no processing indication); (d) On-hand UI with user's query and system response in the industrial training demo application.

Abstract

We present a virtual reality (VR) environment featuring conversational avatars powered by a locally-deployed LLM, integrated with automatic speech recognition (ASR), text-to-speech (TTS), and lip-syncing. Through a pilot study, we explored the effects of three types of avatar status indicators during response generation. Our findings reveal design considerations for improving responsiveness and realism in LLM-driven conversational systems. We also detail two system architectures: one using an LLM-based state machine to control avatar behavior and another integrating retrieval-augmented generation (RAG) for context-grounded responses. Together, these contributions offer practical insights to guide future work in developing task-oriented conversational AI in VR environments.

CCS Concepts

• **Human-centered computing** → **Virtual reality**; **Natural language interfaces**; **Interface design prototyping**.

Keywords

Conversational user interface, intelligent virtual agent, large language model, virtual reality, pilot study.

1 Introduction

Increased focus on Large Language Models (LLMs) has led to significant improvements in the quality of generated text, facilitating development of task-specific LLMs. Realism of Non-Playable-Characters (NPCs) in consumer applications has benefited from these advancements [14], and in academia, LLM-powered intelligent virtual agents (IVAs) are being applied to learning [4, 26], health support [20, 25], the development process [16], and companionship [24, 29], among other uses. In this preliminary work, we explore how participants behave while vocally conversing with virtual avatars to inform the development of future systems.

We developed a system for free-form conversations with IVAs powered by a locally-deployed LLM, automatic speech recognition (ASR), and text-to-speech (TTS) through an API. This pipeline was tested in a pilot study where users completed a quest-like scenario by conversing with avatars, whose behavior was controlled
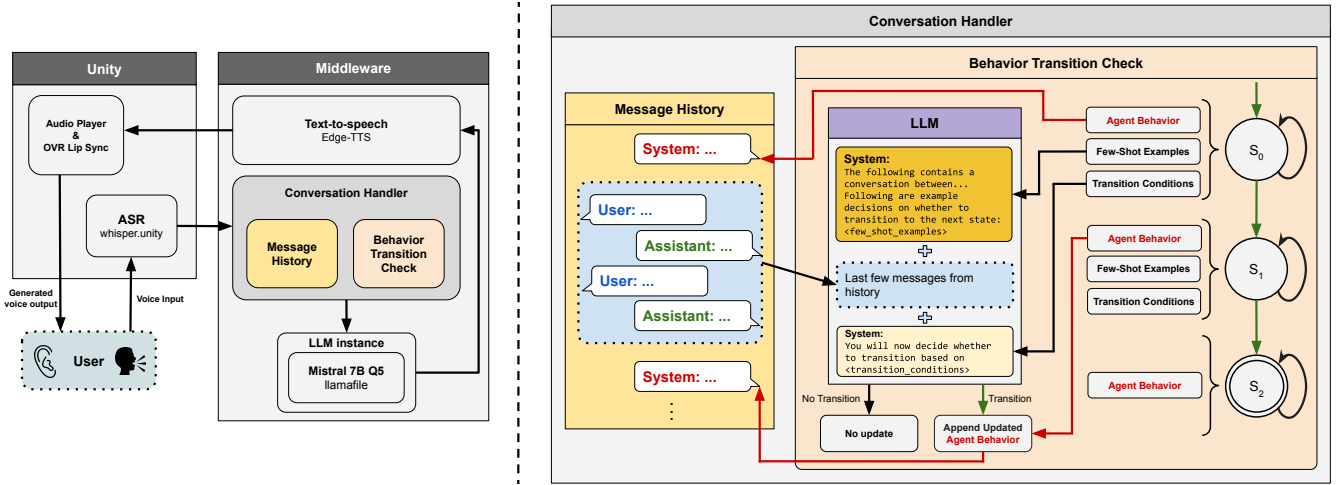
Figure 2: Pipeline for generating responses to user's queries. Left – architecture overview: ASR transcribes user's voice, passing it to Conversation Handler, which uses an LLM to generate a text response that gets voiced by Edge-TTS. Right – Conversation Handler: state management system for controlling agent's behavior. Each state contains agent behavior that gets appended as a system message upon a transition to that state; states with outgoing transitions also contain transition conditions and few-shot examples of transition decisions. Transitions are decided by an LLM, which is instructed to return "transition" / "no transition" responses through system prompts with the last few messages from user-avatar history inserted in-between.

by an LLM-based state machine. Further, we created a retrieval-augmented generation (RAG) application, which answers users queries about a digital twin of an industrial machine, generating responses grounded in context extracted from an operation manual. Observing user behavior and collecting system response timings, head gaze directions, and survey responses, provided us with insights into areas of improvement and design of future conversational systems.

Section 2 covers the architecture and implementation of the multi-agent conversational system, which we used in the pilot study described in section 3. In subsection 3.3, we detail the RAG system for a training application, and takeaways with recommendations from working with LLM-based conversational AI are summarized in subsection 4.1.

2 System implementation

Figure 2-Left shows the conversational system architecture. When the system receives an audio input, the Whisper [17] Unity package[1] transcribes it. This transcription is then sent to a middleware server hosted with FastAPI[2], which manages message histories of the avatars. The updated message history is passed to the Mistral 7b LLM [10], locally hosted with llamafile[3], which generates a text response from the avatar's perspective. This text is then passed to the Edge-TTS [4] system, which generates a voice and saves it as an MP3 file. The path of the audio file is returned to Unity, which downloads it and plays it through a directional Audio Source. The OVR Lip Syncing package[5] controls the blendshapes on the

Table 1: Tasks that appeared on the handheld panel interface. Tasks were crossed-out after they were completed.

Task	Appears After
(1) Talk to Friend	System
(2) Buy walnuts from the store	Friend
(3) Bring walnuts to Friend	Clerk
(4) Ask about milk delivery date at the store	Friend
(5) Ask Manager about next shipment date	Clerk
(6) Tell Friend the milk delivery date	Manager

corresponding avatar's face as the audio is played. The avatars were designed and imported from the ReadyPlayerMe package[6], and the environment was designed to fit the avatars' roles. System response time (SRT), measured between the time when participant finished speaking and the avatar started responding, averaged at 3.2 seconds.

2.1 Task Transitions

To determine whether the latest task was completed (see tasks in Table 1) and if a new one had to be issued, we implemented a state machine through an LLM (see Figure 2-Right). Before the system generated avatar's responses, the last few messages between the participant and a current avatar were appended to pre-written system prompts that instructed the LLM to determine whether an event has occured in the conversation (e.g. whether the participant has completed a purchase of walnuts). The LLM outputted a decision in text ("transition" / "no transition"), and in the case of transition, a new system prompt with updated behavior of the avatar was

---
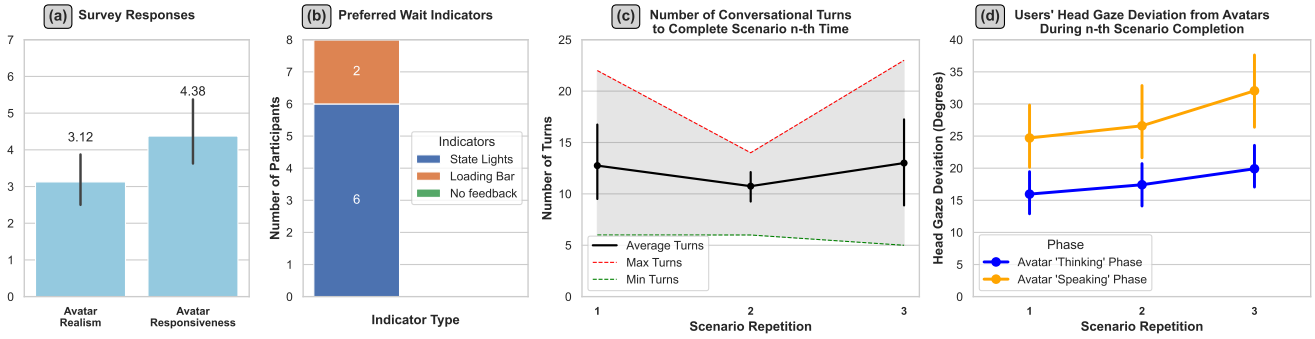
Figure 3: Pilot study results: (a) survey responses about avatar realism and responsiveness; (b) preferred wait feedback types; (c) number of conversational turns required to complete the in-VR scenario n-th time; (d) participant's head gaze deviation angle (from directly looking at the avatar's face) during n-th scenario completion.

appended to its message history, along with displaying the next task to the participant.

3 Pilot Study

The participants were instructed to navigate a virtual environment in a Meta Quest 3 HMD, completing a scenario with a series of tasks (see Table 1) by speaking with three avatars (Friend, Clerk, Manager) at three different locations (Friend's room, store counter, Manager's office). Each avatar was surrounded by an invisible trigger volume (collider), and when participants entered this volume, the avatar turned its head toward the user. While inside the collider, participants activated voice input with "A" button press on a controller, then pressed it again after they finished speaking. After avatars responded in voice, if a previous task was completed (see subsection 2.1), it would appear as strike-through, and a new task was appended to a text UI attached to the participants' left hand. The first task appeared at the application start. When participants navigated to the Friend's room, the Friend asked them to purchase walnuts from a store. The participants then navigated to the store and talked to the Clerk, completing the purchase task through conversation. After participants brought the walnuts back to the Friend, the new task was to return to the store and ask the Clerk about the next milk delivery date. The Clerk told the participants to ask the Manager about the date, and upon completing this, participants returned to the Friend. After informing the Friend about the delivery date, participants took off the HMD and filled out a survey about their experience.
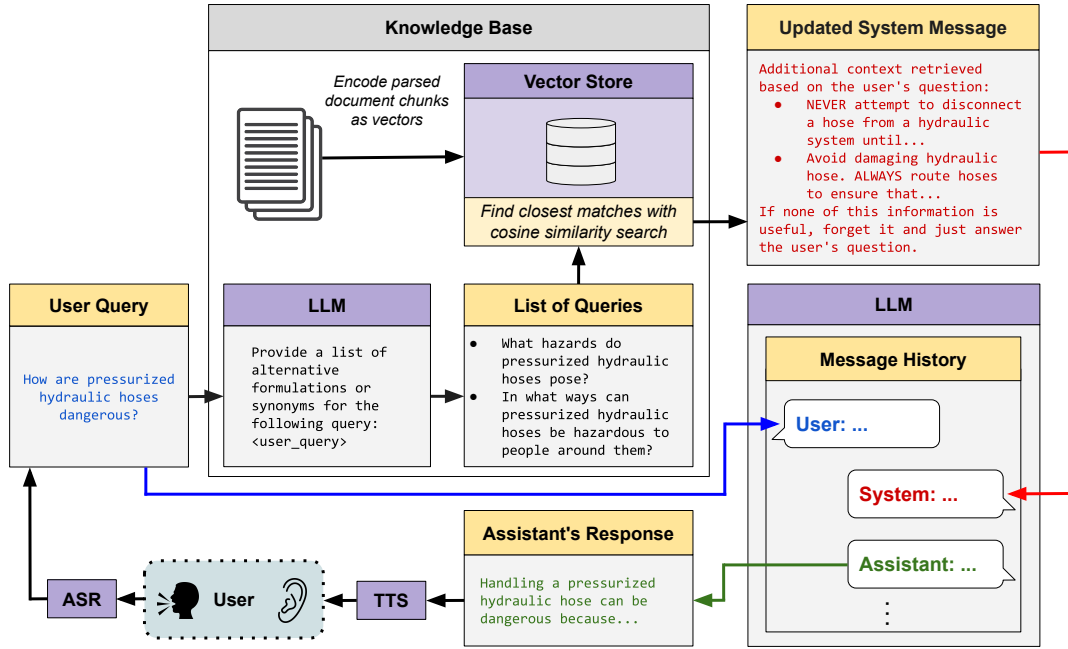
3.1 Conditions

The participants repeated the scenario three times with three different feedback types: *state lights*, *loading bar*, *no feedback*. The order was counterbalanced using the Balanced Latin Square. The *state lights* (Figure 1-a) highlighted the current interaction stage (Idle = active by default, Listening = audio is being recorded, Thinking = processing, Speaking = avatar is responding). The *loading bar* (Figure 1-b) appeared above the avatar's head from the moment the participant pressed the controller button to stop talking, and until the avatar started responding in voice. The *no feedback* condition (Figure 1-c) did not show the current state of the avatar in any way.

3.2 Results

Eight participants (6 male, 2 female), aged 18 to 24 participated in our pilot study. Participants rated avatar realism and responsiveness, as well as selected their preferred wait feedback type. Additionally, we recorded the number of conversational turns required to complete the scenario, and collected participants' HMD gaze direction (gaze deviation angle from directly looking at the avatar's face) during avatar's response generation (Thinking) and annunciation (Speaking) phases.

3.2.1 Survey Responses. We aggregated the survey data averages into a single plot (Figure 3-a) since we found no differences between the three wait feedback conditions. Avatar realism scores were quite low at 3.12 out of 7, which can be explained by the lack of body animations besides lip syncing and avatars turning their head towards the participants. Future studies should include idle and responsive animations, as well as facial expressions to improve realism. Avatar responsiveness was rated more positively than realism (at 4.38 out of 7), still, in future work we will try mitigating the delay caused by SRT (3.2 seconds) through voice and gesture fill-ins, as prior work indicated that such fill-ins can reduce the perceived response latency in related contexts [11]. While realism and responsiveness were not affected by the wait feedback type, most participants preferred *state lights* (6 out of 8) and the *loading bar* (2 out of 8). Some participants commented that presence of any kind of system processing indication gave them the confidence that the avatar heard them, as compared to no indication at all. This is supported by the fact that no participants selected *no feedback* condition as their preferred one.

3.2.2 Objective Metrics. The average number of conversational turns (see plot in Figure 3-c) required to complete the in-VR portion for the first time was ≈13, and lowered to ≈11 during the second play-through (since participants have learned what to say to the avatars in order to progress in the scenario). However, during the third run, the average increased to ≈14 turns, because some participants experimented with the system, testing its limits by saying things unrelated to task completion. Plotting participants' head gaze deviation revealed that participants looked at avatars less and less over the course of scenario repetitions (see Figure 3-d). This

Knowledge Base

Encode parsed document chunks as vectors

Vector Store

Find closest matches with cosine similarity search

Updated System Message

```
Additional context retrieved
based on the user's question:
● NEVER attempt to disconnect
  a hose from a hydraulic
  system until...
● Avoid damaging hydraulic
  hose. ALWAYS route hoses
  to ensure that...
If none of this information is
useful, forget it and just answer
the user's question.
```

User Query

```
How are pressurized
hydraulic hoses
dangerous?
```

LLM

```
Provide a list of
alternative
formulations or
synonyms for the
following query:
<user_query>
```

List of Queries

```
● What hazards do
  pressurized hydraulic
  hoses pose?
● In what ways can
  pressurized hydraulic
  hoses be hazardous to
  people around them?
```

LLM

Message History

User: ...

System: ...

Assistant: ...

Assistant's Response

```
Handling a pressurized
hydraulic hose can be
dangerous because...
```

ASR — **User** — **TTS**

Figure 4: Pipeline for the RAG-enhanced system architecture for answering user's queries about a specific application and machine. After user's speech is transcribed with ASR, alternative formulations of their query are used to retrieve closest matches of text chunks from a machine's manual. This additional context is provided to the LLM as an appended system message.

indicates that in user studies with conversational AI, repeating the same scenario multiple times under varied conditions leads to learning effect (memorization) and lower engagement for some participants, while in others, it leads to undesired experimentation instead of focusing on the completion. Such behaviors add noise to the data and can make detecting differences between conditions more difficult. For a successful user study involving conversational AI with multiple conditions, distinct yet comparable scenarios must be present and counterbalancing carefully applied.

3.3 RAG Application for Industrial Training

We adapted our conversational system to build a demo for an industrial VR training application, where the user could ask questions about a static digital twin of a hydraulic press machine. Unlike system responses for entertainment purposes, responses for safety training must be more precise, so we set the LLM generation temperature to zero [19] and added a RAG component [27]. Figure 4 shows the architecture of the RAG-enhanced system. At application start, a PDF manual for a hydraulic press is parsed into text chunks and encoded as embeddings using the Sentence Transformer library [18]. User queries are reformulated by an LLM, embedded using the same library, and matched to relevant text chunks through cosine similarity search. Before generating a response, a system message with these text chunks is appended to history. In addition to audible output, the text of the latest query and answer was shown on the UI handheld by the user (see Figure 1-d).

We demoed this interface during informal showcases, gathering feedback for system improvements and new features. Users

appreciated the ability to inspect the 3D machine representation but suggested additional interactivity, such as touch or pointing functionality for targeted queries about machine parts, and a stored per-component message history for revisiting prior queries. Incorporating this feedback, we plan to combine the pipelines in Figure 2 and Figure 4, applying them to training [15] and museum exploration studies [2, 26].

4 Discussion

This section reflects on lessons learned during system implementation and evaluation, proposing actionable recommendations and future directions to improve task-oriented conversational AI systems.

4.1 Lessons Learned

4.1.1 Leveraging Open-Source and Free Software. An advantage of developing conversational systems today is the availability of reliable open-source and free software. The modern capabilities of these tools, especially in terms of speed and quality, make it possible to create complex, high-performance systems without costly licensing fees. Every component in our system – from environment assets and avatars, to generative text and audio models – was built using tools that are either open-source or free to use. While paid APIs often produce higher-quality output, they come with their own limitations of potential downtime, higher latency, and recurring costs. For projects where hardware capabilities allow, we recommend exploring locally-deployed alternatives. These not only

reduce dependency on external services but also enable greater control over the system's responsiveness and reliability. As consumer hardware improves and demand for conversational applications grows, we anticipate further advancements in open-source tools, creating a rich ecosystem with plenty of fast and quality options to choose from.

4.1.2 Avatars. Avatars are central to creating an immersive experience, and our current implementation revealed areas for improvement. Participants noted that the avatars appeared too cartoony, which diminished realism. We recommend using higher-fidelity models such as from the Rocketbox [8] or VALID [5] avatar libraries, and ensuring avatars turn toward the user based on proximity, as this feature was well-received. Future work will incorporate idle animations, such as subtle movements, to enhance realism and engagement further.

4.1.3 Scenario Design. Designing effective scenarios is crucial for user studies involving conversational AI, especially with multiple factors. We recommend using distinct but comparable scenarios to minimize bias from confounding variables and applying careful counterbalancing to account for order effects. Repeating the same scenario under different conditions, as in our pilot, introduced unintended behaviors like memorization or experimentation, which reduced engagement and added noise to the data [28]. Despite these issues, the quest-like, task-oriented approach proved effective overall, guiding participants naturally through interactions with virtual characters.

4.2 Future Work

4.2.1 Gesture Recognition Integration. Humans intuitively interpret nonverbal language such as gestures, and effectively use it to communicate in virtual social and collaborative settings [7, 22]. While our current implementation does not give virtual avatars the ability to see users' gestures, in future work we plan to employ a continuous (real-time) gesture recognizer such as Machete [21, 23] or OO-dMVMT [3], and appending a recognized gesture class to the message history of the nearest agent. An alternative recognition approach could involve passing screenshots from the virtual avatar's point-of-view to a visual language model (VLM), prompting it to classify gestures of an embodied human. Gestures could also trigger microphone input instead of pressing a dedicated controller button or pointing at the avatar [13], reducing reliance on manual inputs. By making avatars more perceptive to nonverbal cues, this approach could improve the naturalness of interactions and create a more dynamic user experience [1].

4.2.2 Response Delay Mitigation. Generating speech responses is computationally intensive, requiring sequential processing through ASR, LLM and TTS systems. In our current architecture, the TTS engine relies on receiving the complete response text before generating audio, resulting in an average SRT of 3.2 seconds. In future work, we will explore token streaming to enable overlapping processing, allowing audio for subsequent sentences to be generated while earlier ones are still played. Given the inherent latency, it would be useful to derive design recommendations to improve system usability. A promising direction is to mitigate perceived delays through conversational fillers, such as gesture or voice utterances,

while responses being are generated. Prior work has shown that fillers can reduce perceived latency, but experiments were limited to pre-scripted interactions with fixed delays [11, 12] or Wizard-of-Oz setups [6, 9]. A useful experiment would apply capabilities and speed of modern models, investigating the effect of conversational fillers on perceived latency and learning outcomes in free-form conversations with IVAs.

5 Conclusion

In this work, we demonstrated the use of LLMs for conversational avatars in VR, exploring design considerations for response feedback and realism. By detailing two system architectures and incorporating user feedback, we provide practical insights to guide future development of task-oriented conversational AI systems. Additionally, we outlined promising directions for future work, including possible approaches to gesture recognition integration and response delay mitigation through token streaming and conversational fillers. These advancements aim to enhance the naturalness and efficiency of interactions, paving the way for more immersive and responsive virtual environments.

References

[1] Deepali Aneja, Rens Hoegen, Daniel McDuff, and Mary Czerwinski. 2021. Understanding Conversational and Expressive Style in a Multimodal Embodied Conversational Agent. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3411764.3445708

[2] Rojin Bayat, Elios De Maio, Jacopo Fiorenza, Massimo Migliorini, and Fabrizio Lamberti. 2024. Exploring Methodologies to Create a Unified VR User-Experience in the Field of Virtual Museum Experiences. In *2024 IEEE Gaming, Entertainment, and Media Conference (GEM)*. IEEE, Turin, Italy, 1–4. https://doi.org/10.1109/GEM61861.2024.10585452 ISSN: 2766-6530.

[3] Federico Cunico, Federico Girella, Andrea Avogaro, Marco Emporio, Andrea Giachetti, and Marco Cristani. 2023. OO-dMVMT: A Deep Multi-view Multi-task Classification Framework for Real-time 3D Hand Gesture Classification and Segmentation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, Vancouver, BC, Canada, 2745–2754. https://doi.org/10.1109/CVPRW59228.2023.00275

[4] Rahul R. Divekar*, Jaimie Drozdal*, Samuel Chabot*, Yalun Zhou, Hui Su, Yue Chen, Houming Zhu, James A. Hendler, and Jonas Braasch. 2022. Foreign language acquisition via artificial intelligence and extended reality: design and evaluation. *Computer Assisted Language Learning* 35, 9 (Dec 2022), 2332–2360. https://doi.org/10.1080/09588221.2021.1879162

[5] Tiffany D. Do, Steve Zelenty, Mar Gonzalez-Franco, and Ryan P. McMahan. 2023. VALID: a perceptually validated Virtual Avatar Library for Inclusion and Diversity. *Frontiers in Virtual Reality* 4 (Nov. 2023), 1–15. https://doi.org/10.3389/frvir.2023.1248915

[6] Kotaro Funakoshi, Kazuki Kobayashi, Mikio Nakano, Seiji Yamada, Yasuhiko Kitamura, and Hiroshi Tsujino. 2008. Smoothing human-robot speech interactions by using a blinking-light as subtle expression. In *Proceedings of the 10th international conference on Multimodal interfaces (ICMI '08)*. Association for Computing Machinery, New York, NY, USA, 293–296. https://doi.org/10.1145/1452392.1452452

[7] Ryan Khushan Ghamandi, Ravi Kiran Kattoju, Yahya Hmaiti, Mykola Maslych, Eugene Matthew Taranta, Ryan P. McMahan, and Joseph LaViola. 2024. Unlocking Understanding: An Investigation of Multimodal Communication in Virtual Reality Collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 602, 16 pages. https://doi.org/10.1145/3613904.3642491

[8] Mar Gonzalez-Franco, Eyal Ofek, Ye Pan, Angus Antley, Anthony Steed, Bernhard Spanlang, Antonella Maselli, Domna Banakou, Nuria Pelechano, Sergio Orts-Escolano, Veronica Orvalho, Laura Trutoiu, Markus Wojcik, Maria V. Sanchez-Vives, Jeremy Bailenson, Mel Slater, and Jaron Lanier. 2020. The Rocketbox Library and the Utility of Freely Available Rigged Avatars. *Frontiers in Virtual Reality* 1 (2020), 20. https://doi.org/10.3389/frvir.2020.561558

[9] Yuin Jeong, Juho Lee, and Younah Kang. 2019. Exploring Effects of Conversational Fillers on User Perception of Conversational Agents. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*.

Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312913

[10] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7B. http://arxiv.org/abs/2310.06825 arXiv:2310.06825 [cs].

[11] Junyeong Kum and Myungho Lee. 2022. Can Gestural Filler Reduce User-Perceived Latency in Conversation with Digital Humans? Applied Sciences 12, 21 (Jan. 2022), 10972. https://doi.org/10.3390/app122110972 Number: 21 Publisher: Multidisciplinary Digital Publishing Institute.

[12] Soledad López Gambino, Sina Zarrieß, and David Schlangen. 2017. Beyond On-hold Messages: Conversational Time-buying in Task-oriented Dialogue. In Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue, Kristiina Jokinen, Manfred Stede, David DeVault, and Annie Louis (Eds.). Association for Computational Linguistics, Saarbrücken, Germany, 241–246. https://doi.org/10.18653/v1/W17-5529

[13] Mykola Maslych, Difeng Yu, Amirpouya Ghasemaghaei, Yahya Hmaiti, Esteban Segarra Martinez, Dominic Simon, Eugene Matthew Taranta, Joanna Bergström, and Joseph J. LaViola Jr. 2024. From Research to Practice: Survey and Taxonomy of Object Selection in Consumer VR Applications. In 2024 IEEE International Symposium on Mixed and Augmented Reality (Seattle, WA, USA) (ISMAR '24). IEEE, Piscataway, NJ, USA, 990–999. https://doi.org/10.1109/ISMAR62088.2024.00115

[14] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (San Francisco, CA, USA) (UIST '23). Association for Computing Machinery, New York, NY, USA, Article 2, 22 pages. https://doi.org/10.1145/3586183.3606763

[15] Gustav Bøg Petersen, Aske Mottelson, and Guido Makransky. 2021. Pedagogical Agents in Educational VR: An in the Wild Study. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3411764.3445760

[16] Hua Xuan Qin, Shan Jin, Ze Gao, Mingming Fan, and Pan Hui. 2024. CharacterMeet: Supporting Creative Writers' Entire Story Character Construction Processes Through Conversation with LLM-Powered Chatbot Avatars. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3613904.3642105

[17] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust Speech Recognition via Large-Scale Weak Supervision. arXiv:2212.04356 [eess.AS] https://arxiv.org/abs/2212.04356

[18] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Hong Kong, China, 3982–3992. https://arxiv.org/abs/1908.10084

[19] Matthew Renze and Erhan Guven. 2024. The Effect of Sampling Temperature on Problem Solving in Large Language Models. arXiv:2402.05201 [cs.CL] https://arxiv.org/abs/2402.05201

[20] Amir Bani Saeed, Zahra Moussavi, and Bruce Hardy. 2024. Developing an Avatar in Virtual Reality for Mental Health Treatment. CMBES Proceedings 46 (June 2024), 1–1. https://proceedings.cmbes.ca/index.php/proceedings/article/view/1107

[21] Eugene Matthew Taranta, Mykola Maslych, Ryan Ghamandi, and Joseph LaViola. 2022. The Voight-Kampff Machine for Automatic Custom Gesture Rejection Threshold Selection. 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 556, 15 pages. https://doi.org/10.1145/3491102.3502000

[22] Eugene M. Taranta, Corey R. Pittman, Jack P. Oakley, Mykola Maslych, Mehran Maghoumi, and Joseph J. LaViola. 2020. Moving Toward an Ecologically Valid Data Collection Protocol for 2D Gestures In Video Games. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376417

[23] Eugene M. Taranta II, Corey R. Pittman, Mehran Maghoumi, Mykola Maslych, Yasmine M. Moolenaar, and Joseph J. Laviola Jr. 2021. Machete: Easy, Efficient, and Precise Continuous Custom Gesture Segmentation. ACM Transactions on Computer-Human Interaction (TOCHI) 28, 1, Article 5 (Jan. 2021), 46 pages. https://doi.org/10.1145/3428068

[24] Hongyu Wan, Jinda Zhang, Abdulaziz Arif Suria, Bingsheng Yao, Dakuo Wang, Yvonne Coady, and Mirjana Prpa. 2024. Building LLM-based AI Agents in Social Virtual Reality. In Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24). Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3613905.3651026

[25] Jieyu Wang, Li Zhang, Dingfang Kang, and Katherina G. Pattit. 2024. Designing Conversational Agents for Student Wellbeing. International Journal of Advanced Computer Science and Applications 15, 10 (2024), 43–52. https://doi.org/10.14569/IJACSA.2024.0151006

[26] Zhan Wang, Lin-Ping Yuan, Liangwei Wang, Bingchuan Jiang, and Wei Zeng. 2024. VirtuWander: Enhancing Multi-modal Interaction for Virtual Tour Guidance through Large Language Models. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24). Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3613904.3642235

[27] Shangyu Wu, Ying Xiong, Yufei Cui, Haolun Wu, Can Chen, Ye Yuan, Lianming Huang, Xue Liu, Tei-Wei Kuo, Nan Guan, and Chun Jason Xue. 2024. Retrieval-Augmented Generation for Natural Language Processing: A Survey. arXiv:2407.13193 [cs.CL] https://arxiv.org/abs/2407.13193

[28] Difeng Yu, Qiushi Zhou, Benjamin Tag, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. Engaging Participants during Selection Studies in Virtual Reality. In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, Atlanta, GA, USA, 500–509. https://doi.org/10.1109/VR46266.2020.00071

[29] Jiarui Zhu, Radha Kumaran, Chengyuan Xu, and Tobias Höllerer. 2023. Freeform Conversation with Human and Symbolic Avatars in Mixed Reality. In 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, Sydney, Australia, 751–760. https://doi.org/10.1109/ISMAR59233.2023.00090 ISSN: 2473-0726.