

Integration of LLMs with Virtual Character Embodiment

James Hale
University of Southern California
United States of America
jahale@usc.edu

Lindsey Schweitzer
Middlebury College
United States of America
lschweitzer@middlebury.edu

Jonathan Gratch
University of Southern California
United States of America
gratch@ict.usc.edu

CCS Concepts

• **Human-centered computing** → **Collaborative and social computing**.

ACM Reference Format:

James Hale, Lindsey Schweitzer, and Jonathan Gratch. 2024. Integration of LLMs with Virtual Character Embodiment. In *ACM International Conference on Intelligent Virtual Agents (IVA '24)*, September 16–19, 2024, GLASGOW, United Kingdom. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3652988.3696199>

1 Rationale

This work demonstrates a system wherein virtual agents may act in place of human confederates while retaining the benefits (e.g., controllability and interactivity) of scripted AI agents. We develop this system using existing character animation tools [6], and large language models (LLMs) [1] such that we have a virtual agent a human can interact with through natural language. Researchers in the intelligent virtual agent community have used AI and scripted systems previously to investigate IVAs in many contexts and tasks, e.g., negotiation [12], empathetic behavior [2], collaborative talk [14], etc. Researchers could use this system to investigate areas of interest to the IVA community in more natural and human-like interactions. Further, we use this system in an IVA'24 paper, investigating the pitfalls of embodiment relating to gender [5]. Whereas prior work using human-human approaches shows women perform worse in negotiation, the dyadic nature of these interactions makes it challenging to tease apart whether this disparity stems from women's perceptions of their counterparts or increased toughness from men. This platform gains an advantage through increased experimental control, as the agent does not perceive otherwise salient appearance features, such as gender and race. In our work [5], we find men perform relatively better against an embodied masculine agent (presented here) than women, and women perform relatively better against a disembodied agent (chatbot) — we can begin to tease apart the causes of gender effects found in the dyadic literature.

2 System Explanation

We develop a system that allows a human to interact naturally with a virtual agent through natural language while retaining some control for the experiment designer.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

IVA '24, September 16–19, 2024, GLASGOW, United Kingdom

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0625-7/24/09

<https://doi.org/10.1145/3652988.3696199>

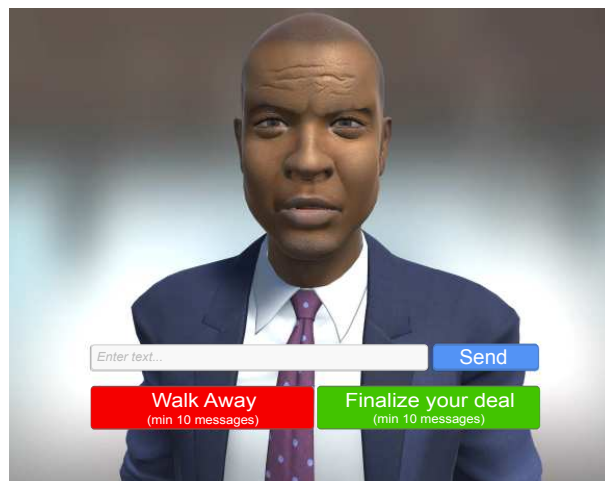


Figure 1: Interaction interface

2.1 Interaction Setting

Our demo takes on a dispute resolution context, where a participant engages in a buyer-seller dispute against this virtual agent. The dispute involves a buyer claiming to have received an incorrect item from a seller's online store, and the seller rejecting that claim — negative reviews from each side escalate the situation. The participant takes the role of a buyer trying to get a refund, apology, and the seller's negative review removed from their profile — they receive instruction from the prompt in Figure 3. Similarly, the virtual agent takes the role of the seller and aims to not give a refund, to get an apology, and to have the buyer's negative review removed. Further, the agent either takes a *professional* or *hostile* demeanor — the following demonstrate fixed opening messages for each condition:

- **Professional:** *Your request for a refund has surprised us, as we believe our product description is quite clear, and we remain committed to our policy. We value your feedback, and we would appreciate the opportunity to address any concerns you may have.*
- **Hostile:** *Your sudden demand for a refund is unwarranted. Our product description is crystal clear, and we stand by our policy. Your behavior is disappointing, and your negative review is unfounded.*

Given this setup, the agent sends an introductory message and the two sides chat in a turn-based manner via the interface in Figure 1.

2.2 Architecture

Figure 2 illustrates the platform's architecture, where a user inputs natural language, and the virtual agent responds in natural language

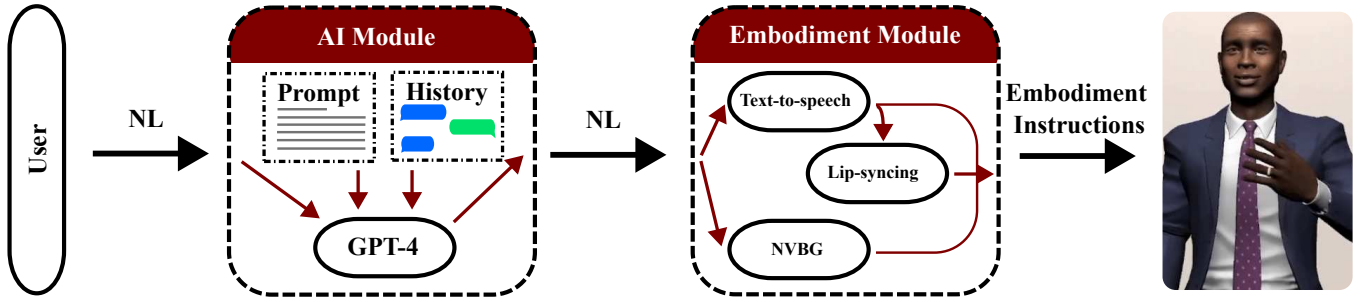


Figure 2: Diagram of our architecture, where a user communicates in natural language (NL) with our agent

Buyer role play instructions

Your terminally ill nephew is a huge Kobe Bryant fan so you purchased him a replica of Bryant's last Basketball Championship jersey for \$75. The website clearly indicated the purchase was for Bryant's jersey. When the jersey arrived, it was for a different player you never heard of. You request the correct jersey be sent.

The Seller responds: "The website clearly indicated this was for a Los Angeles Lakers jersey, not for a specific player. All sales are final." You see they now removed mention of Kobe Bryant from their website but you know they are lying. To protect other customers, you post a negative review warning about the Seller's deceptive behavior.

The Seller posted a negative review about you, calling you a "SMARTASS, SLANDERER and a FRAUD." You have dozens of transactions on this site and have a near-spotless reputation. Now you worry others won't sell to you.

Below are issues you could negotiate to resolve this dispute:

- **Refund** *The buyer could demand a full or partial refund for the item's price.*
- **Keep the negative review of seller** *The Seller might demand the buyer take down their negative review.*
- **Seller removes negative review** *The buyer could demand the seller remove the bad review they made.*
- **Receive formal apology** *The buyer could demand the seller update their bad review with an apology clarifying that they made a mistake and apologize for how they initially responded.*

Figure 3: Negotiation setting and overview of issues

with corresponding verbal and non-verbal behaviors. Two modules comprise this architecture — an *AI Module* and an *Embodiment Module*.

2.2.1 AI Module. This module handles the virtual agent's reasoning, natural language understanding, and natural language generation. As we take the user's natural language as input and output natural language to the next module, we use OpenAI's GPT-4 [1] (an LLM) to drive this module with the default temperature setting of 1. We instruct the LLM with dialog history and a prompt, constructed similarly to the block of text the user reads before the interaction

in Figure 3; we add further text to drive the agent's behavior, such as telling it to act professionally or angrily, or informing of its preferences. Despite our simple approach of using a raw LLM, a researcher may swap out this module with another that fits their experiment needs. For example, much research exists on creating negotiating agents, and one may desire an implementation that allows higher controllability [10, 11]; more human-like behavior [3, 9]; or somewhere in-between.

2.2.2 Embodiment Module. This module controls the embodiment of the virtual agent and complements the output from the *AI Module*. Here, we handle three primary tasks: generating the agent's gestures, speech, and lip-syncing. Given the semantic output of the prior module, we use the Non-verbal Behavior Generation (NVBG) tool developed by Lee and Marsella [8] to generate appropriate non-verbal behaviors for the agent, such as hand gestures, nodding, etc. Amazon's Polly handles the text-to-speech — allowing for controlling pitch, tone, and intonation — while FaceFX [4]'s viseme scheduler creates lip-syncing based on the resulting audio. For the character animation, Hartholt et al. [7]'s RIDE handles the character animation, which allows the execution of various behaviors — e.g., facial expressions, gesturing, and back-channeling — informed by the non-verbal behavior and lip-syncing generation. Lastly, given the agent can either act *professionally* or with *hostility*, we force the agent to display either a *happy* or *angry* expression prior to each utterance respectively.

2.2.3 Speech Only Communication. We augment the typing interface with speech-to-text using OpenAI's Whisper [13] — allowing purely verbal communication while retaining the possibility of communicating through text.

3 Video Materials

One can find a short video demonstration of our system here.

Acknowledgments

This work is supported by the U.S. Government including the Air Force Office of Scientific Research, under grant FA9550-23-1-0320, and the Army Research Office under Cooperative Agreement Number W911NF-20-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government

is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Floren-
cia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal
Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*
(2023).
- [2] Mukesh Barange, Sandratra Rasendrasoa, Maël Bouabdelli, Julien Saunier, and
Alexandre Pauchet. 2022. Impact of adaptive multimodal empathic behavior on
the user interaction. In *Proceedings of the 22nd ACM International Conference on*
Intelligent Virtual Agents. 1–8.
- [3] Kushal Chawla, Ian Wu, Yu Rong, Gale M Lucas, and Jonathan Gratch. 2023. Be
selfish, but wisely: Investigating the impact of agent personality in mixed-motive
human-agent interactions. *arXiv preprint arXiv:2310.14404* (2023).
- [4] FaceFX. 2024. FaceFX Software. <http://www.facefx.com/> Accessed: 2024-07-11.
- [5] James Hale, Lindsey Schweitzer, and Jonathan Gratch. 2024. Pitfalls of em-
bodiment in human-agent experiment design. In *Proceedings of the 24th ACM*
International Conference on Intelligent Virtual Agents. 1–9.
- [6] Arno Hartholt, Kyle McCullough, Ed Fast, Andrew Leeds, Sharon Mozgai, Tim
Aris, Volkan Ustun, Andrew Gordon, and Chris McGroarty. 2020. Rapid Proto-
typing for Simulation and Training with the Rapid Integration & Development
Environment (RIDE).
- [7] Arno Hartholt, Kyle McCullough, Ed Fast, Andrew Leeds, Sharon Mozgai, Tim
Aris, Volkan Ustun, and Andrew S. Gordon. 2021. Rapid Prototyping for Sim-
ulation and Training with the Rapid Integration & Development Environment
(RIDE). In *Interservice/Industry Training, Simulation and Education Conference*
(IITSEC).
- [8] Jina Lee and Stacy Marsella. 2006. Nonverbal behavior generator for embodied
conversational agents. In *International Workshop on Intelligent Virtual Agents*.
Springer, 243–255.
- [9] Mike Lewis, Denis Yarats, Yann N Dauphin, Devi Parikh, and Dhruv Batra. 2017.
Deal or no deal? end-to-end learning for negotiation dialogues. *arXiv preprint*
arXiv:1706.05125 (2017).
- [10] Eleanor Lin, James Hale, and Jonathan Gratch. 2023. Toward a better understand-
ing of the emotional dynamics of negotiation with large language models. In
Proceedings of the Twenty-fourth International Symposium on Theory, Algorith-
mic Foundations, and Protocol Design for Mobile Networks and Mobile Computing.
545–550.
- [11] Johnathan Mell and Jonathan Gratch. 2016. IAGO: Interactive Arbitration Guide
Online (Demonstration).
- [12] Johnathan Mell, Gale M Lucas, and Jonathan Gratch. 2020. Varied magnitude
favor exchange in human-agent negotiation. In *Proceedings of the 20th ACM*
International Conference on Intelligent Virtual Agents. 1–8.
- [13] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and
Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision.
In *International Conference on Machine Learning*, PMLR, 28492–28518.
- [14] Joseph B Wiggins, Toni V Earle-Randell, Dolly Bounajim, Yingbo Ma, Ju-
lianna Martinez Ruiz, Ruohan Liu, Mehmet Celepkolu, Maya Israel, Eric Wiebe,
Collin F Lynch, et al. 2022. Building the dream team: children’s reactions to
virtual agents that model collaborative talk. In *Proceedings of the 22nd ACM*
International Conference on Intelligent Virtual Agents. 1–8.