Use of Embodied Conversational Agents to Engage Visitors in Art Exhibits

José Fernando R. Ferreira Neto[1], Daniel M. Colares[1], Daniel Ribeiro[1], Yuri Nekan S. Fontes[1], Vladia Pinheiro[1]

[1]Fundação Edson Queiroz – Universidade de Fortaleza (UNIFOR)
Av. Washington Soares 1321 – 60.811-905 – Fortaleza – CE – Brazil

josefernando@unifor.br, danielmcolares@gmail.com

daniel.ribeiro0833@gmail.com, yurinekan@gmail.com, vladiacelia@unifor.br

Abstract. *The rapid advancement of Artificial Intelligence presents new opportunities and challenges across numerous fields. This paper explores the application of Embodied Conversational Agents in art galleries as a means to enhance visitor engagement. We propose a general system architecture for developing such an agent and describe the implementation of an interactive exhibit based on the painting "Cabeça de Mulato" by Candido Portinari. Preliminary results from logged visitor interactions indicate that ECAs can indeed foster meaningful engagement in cultural settings.*

1. Introduction

In recent years, Artificial Intelligence (AI) has transformed many industries. In the creative industry, specifically, in the arts community, its role has been a little more controversial, yet it undeniably enables new cultural experiences that are reshaping how audiences engage with exhibitions and galleries. Driven by continual advances and growing accessibility of technologies such as Augmented Reality (AR) and Generative AI (GenAI), powered by the remarkable natural language understanding and generation capabilities of Large Language Models (LLMs), such as GPT-4 [OpenAI and et al. 2024], Llama 3 [Meta 2024], and DeepSeek [DeepSeek-AI and et al. 2025], these experiences are increasingly immersive and participatory, turning passive observers into active participants while opening new avenues for learning across diverse topics.

The motivation to this work arose when the University of Fortaleza's art gallery, preparing the 'Centelhas em Movimento' exhibit in 2024, realized that a key work named 'Cabeça de Mulato' (Figure 1) was on display at the Venice Biennale and thus unavailable locally, starting a search for a distinctive attraction. 'Cabeça de Mulato' is a famous existing painting by the infamous Brazilian artist Cândido Portinari. After researching ways to boost engagement and leveraging the lab's prior experience, the team proposed an interactive digital twin: a "talking painting" that could converse in natural language about the artwork, its collector, the exhibit, and related topics. With approvals secured, the project moved forward with enthusiastic support.

One objective of this study is to investigate how interactive technologies can enrich visitors' experiences in art galleries by increasing engagement and fostering cultural curiosity. To this end, we present an architecture for an Embodied Conversational

Agent (ECA) [Cassell et al. 2000] that integrates a LLM for natural language understanding (NLU) and personalized response generation, and text-to-speech (TTS) for spoken feedback. Following this design, we developed **Mulato**, a digital twin of Portinari's painting—and conducted an interactive experiment during the 'Centelhas em Movimento' exhibition. With participants' consent, conversations were logged and later analyzed to understand how visitors interacted with the agent.

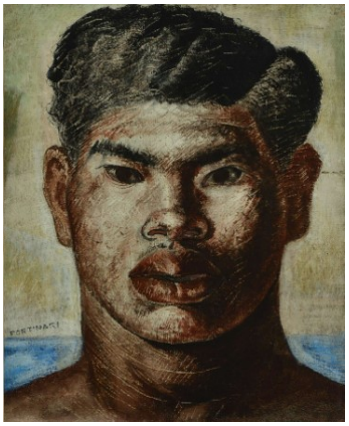

Figure 1. Original painting: "Cabeça de Mulato"

2. Related Works

Regarding the use of AI to generate representation and historical and fictional characters, we find the work of Deep Nostalgia [Kidd and Nieto McAvoy 2023, Pomella 2021], licensed by MyHeritage from D-ID, which is a technology using Deep Learning to animate faces in historical photos and create high-quality realistic videos. This work is mostly used to animate photographs of family and friends, but it has also been used to animate images of historical figures. However interesting, this does not achieve the pursued interaction sought by the team.

Several groups now enable text-based conversations with figures from history, fiction, and pop culture, e.g., Hello History.ai [History 2024] and CharacterAI [CharacterAI 2024]. Hutson et al. [Hutson et al. 2024] analyze this "digital resurrection," showing that a customized ChatGPT can revitalize historical figures (e.g., Mary Sibley) for education and engagement. PeopleAI [PeopleAI 2024] similarly facilitates dialogues with notable figures, even staging multi-party debates "around a dinner table." While these efforts introduced conversationality, they remained largely text-centric; our work sought a more immersive, embodied experience. A Brazilian initiative, Euvatar [Flávia Peres 2024], identified just before the 'Centelhas em Movimento' exhibit opening, deploys real-time 3D digital humans in public venues; a demo featuring Machado de Assis conversing with visitors [Torres 2024] exemplifies an Embodied Conversational Agent (ECA) with a visible avatar rather than a simple chat interface.

These works collectively demonstrate a growing trend toward leveraging AI to animate and revitalize cultural narratives, providing a strong foundation for our approach to embedding conversational agents in art exhibits.

3. An Architecture for Embodied Conversational Agents (ECA)

Figure 2 presents the architecture of the Embodied Conversational Agent (ECA) proposed in this work. It integrates a Large Language Model (LLM) for natural language understanding and personalized response generation, supported by a knowledge base via Retrieval-Augmented Generation (RAG), and a voice-generation module that delivers spoken feedback through text-to-speech (TTS).

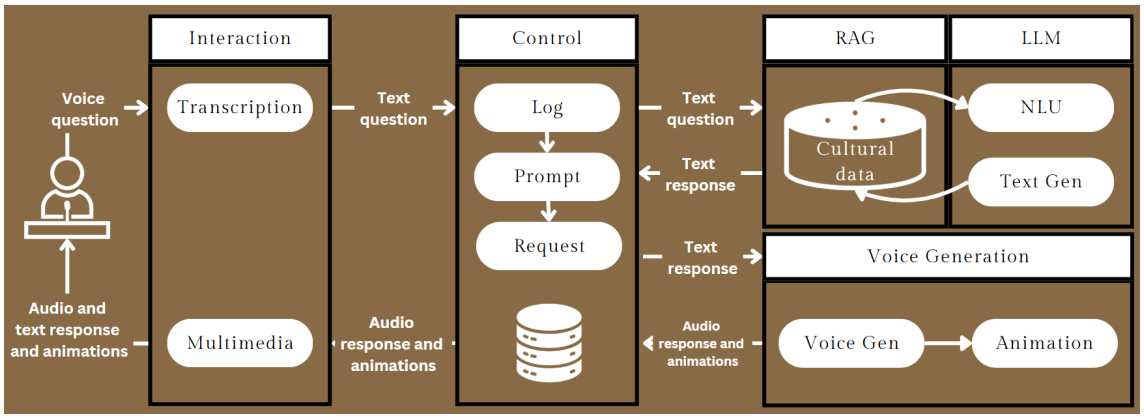

Figure 2. Embodied Conversational Agents (ECA)'s System Architecture

The cycle starts in **Interaction** module when a user input a voice prompt to the agent, or **Mulato** in our case. So there needs to be a microphone and an interface through which the person can engage in conversation. In this scenario, a web client was developed for Chrome in which the user can visualize the animated character, the prompted message, and also receive the audio response generated by the system. This web client is also responsible for the speech-to-text (STT) conversion, which will enable both the prompted message and response to be recorded in an onscreen chat, so that the user can track his or her conversation with the agent. Depending on the implementation method for prompting the agent, we suggest a push-to-talk feature for the mic. Clear instructions are also advised to guarantee intuitiveness and usability, a simple and effective "Press the mic button to speak" was necessary, as we chose to not have a mouse or keyboard.

In the **Control** module, the user's and system's prompt are logged in a database for further analysis in a back-end application. This API is also responsible for requesting external services through their own APIs, and logging success statuses for maintenance purposes. So practically speaking, this middleware is what collects the response from the LLM and the lipsync data to be played by the Client.

The main module is the **LLM**, which is responsible for the NLU and response generation. Currently, there are many different and very reliable options in the market, so any will do. However, if the LLM doesn't perform well, the ECA can become dull and unattractive. In this work, the team decided to go with the LLM that best performed in our tests, GPT-4 [OpenAI and et al. 2024]. Through OpenAI's platform, we created a personalized assistant using a technique called **Retrieval-Augmented Generation (RAG)**, which provided a set of additional instructions and additional cultural data, which was not accessible or revealed to the LLM during training, such as other artists displayed in the same exhibit as Mulato, and the name of the man who served as model for the original

painting. The prompt engineering followed the framework called CO-STAR [Teo 2023], which stands for: Context, Objective, Style, Tone, Audience, and Response. This allowed us to customize each aspect of the agents performance, making sure the agent doesn't speak too much, or speaks rudely to visitors, for example.

Finally, after the text response is generated, the **Voice Generation** module renders it as audio. To enhance realism, we synchronize the avatar's lip movements with the spoken output. In our implementation, Azure Cognitive Services Text-to-Speech (TTS) produces both the MP3 audio and the lip-sync (viseme) timing data needed to animate the character during playback. When the response finishes, the interaction loop restarts and the user can ask another question.

4. MULATO Agent: experiment and results

To optimize the experience, we ran extensive pre-launch tests in the lab and on-site at the gallery [Ferreira 2025]. These iterations refined the avatar's appearance, response length, latency between question and answer, the clarity of on-screen instructions, and the overall interface, with contributions from both the lab and the Culture Arts department. Because these were internal performance tests, we do not treat them as validation due to potential bias; formal usability studies with naive participants remain necessary. Even so, logs from interactions with external visitors offer initial evidence of engagement. At the March 13, 2024 opening (Figure 3), **Mulato** drew a diverse audience, and several visitors said they would return. Field observations surfaced usability issues, such as high ambient noise, users mistaking on-screen messages for the mic button and touching the screen, and frequent failures to toggle the microphone despite prompts. Younger visitors probed the agent's knowledge boundaries, which Mulato generally handled well, including apologizing when unable to answer, as designed.

Figure 3. Mulato ECA Interactive Prototype on Display

The experiment was on display for 34 weeks, from April to December in 2024, during which time he received and answered 4,580 questions. An average of 6.57 tokens were used in each question and 19.19 tokens in each answer. During that time the exhibit

recorded a total of 34,810 visitants. We estimate a total of 4,074 different interactions with the experiment, which represents almost 12% of the total visitants. This number may not be enough to indicate the attractiveness of the experiment. Some factors may have contributed to this result, such as the placement of the experiment in the exhibit and some group interactions.

In order to try and determine the **Mulato's** quality and attractiveness, we analyzed all logged text interactions. Because the "plug-and-play" setup and time constraints prevented per-user session IDs, we segmented sessions using a 3-minute idle gap. On average, we had 1.12 questions per user, with a duration of 2m40s per interaction. An abandonment analysis was conducted to determine how often users engaged in only a single-turn interaction with the Mulato conversational agent. Among 4,074 total interactions, 3,610 (approximately 88%) consisted of just one user question. Few users interacted with more than 4 questions. This high abandonment suggests limited follow-up—potentially due to satisfaction with the first answer, limited time/interest, gallery conditions (noise/connectivity), or responses not engaging enough to prompt continued dialogue.

Analyzing the top 20 bigrams and trigrams (Figure 4), we found the main topics addressed by users were: about the painting itself and it's relevance; the artist and his personal life; the character in the painting and how it was created; location of the real painting and where to see more; general topics and exploration.

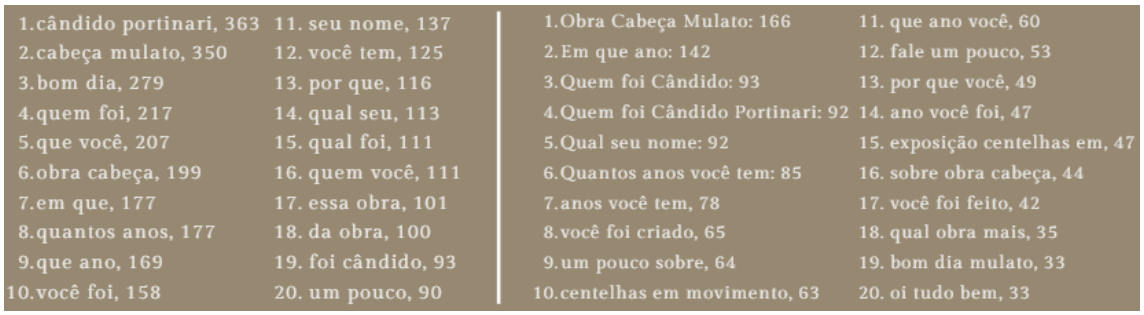

Figure 4. Top 20 Bigrams and Trigrams extracted from User's interactions with Mulato ECA.

To identify the main discussion topics in user interactions with the virtual character **Mulato**, we applied an unsupervised clustering approach. First, user questions were vectorized using the TF-IDF method to capture their semantic content. Then, we used the K-Means algorithm to group similar questions into thematic clusters. Each cluster was analyzed by extracting the most frequent keywords, which helped us label and interpret the predominant themes within each group (Figure 5).

To evaluate the emotional tone of user interactions with the Mulato conversational agent, we performed a sentiment analysis on all user-submitted questions. Using the TextBlob library, each question was analyzed for its polarity score, which ranges from -1 (very negative) to +1 (very positive). Based on this score, questions were classified into three categories: Positive, Neutral, and Negative. Figure 6 presents a graph where the majority of questions were classified as Neutral (4,477), indicating a predominantly factual or inquisitive tone, with only a small part falling into Positive (81) or Negative (22) sentiment categories.

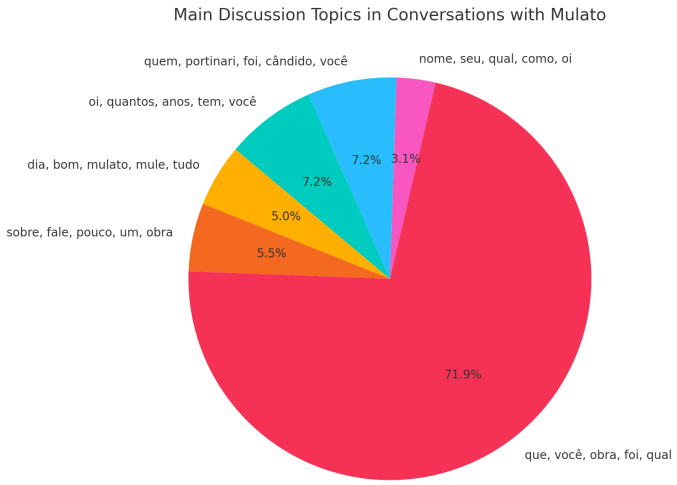

Figure 5. Topic Clusters extracted from User's interactions with Mulato ECA.

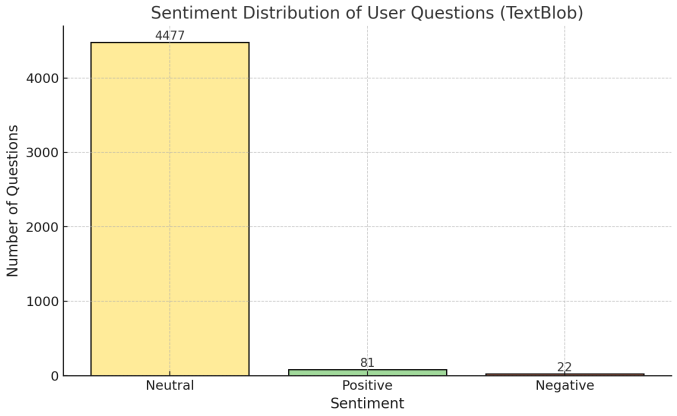

Figure 6. Sentiment analysis from User's interactions with Mulato ECA.

5. Conclusion

GenAI technology is successful at capturing attention and engaging people with art and culture. User observations and conversations presented in this paper revealed curiosity and satisfaction with the interactions, enriching their experience in the exhibit. Further testing using consolidated and structured scales for usability and utility can be applied to capture objective and subjective perceptions of the users. The data collected is very rich and can be further analyzed to determine sentiment towards ECAs. Future work can integrate other senses to the ECA as to further enhance user experience, such as sight for example. Future iterations could also incorporate adaptive conversational memory, enabling the agent to build context over multiple exchanges, thereby fostering deeper dialogue and repeat engagement.

6. Acknowledgments

To Vortex Lab and the Culture and Arts department at the University of Fortaleza (UNIFOR) for collaborating and funding this project.

References

Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. (2000). *Embodied Conversational Agents*. The MIT Press.

CharacterAI (2024). Characterai. https://character.ai/. Accessed: 2024-05-30.

DeepSeek-AI and et al. (2025). Deepseek-v3 technical report.

Ferreira, F. (2025). Demo do mulato para dr. igor queiroz. https://youtube.com/shorts/2f3U1AvG7Ts?feature=share. Accessed: 2025-06-16.

Flávia Peres (2024). Euvatar. https://www.euvatar.com.br/. Accessed: 2024-05-30.

History, H. (2024). Hello history. https://www.hellohistory.ai/. Accessed: 2024-05-30.

Hutson, J., Huffman, P., and Ratican, J. (2024). Digital resurrection of historical figures: A case study on mary sibley through customized chatgpt. *Metaverse*, 4(2).

Kidd, J. and Nieto McAvoy, E. (2023). Deep nostalgia: Remediated memory, algorithmic nostalgia and technological ambivalence. *Convergence*, 29(3):620–640.

Meta (2024). Introducing llama 3.1: Our most capable models to date.

OpenAI and et al., J. A. (2024). Gpt-4 technical report.

PeopleAI (2024). Peopleai. https://peopleai.app/. Accessed: 2024-07-03.

Pomella, A. (2021). Deep nostalgia. *Maradagàl: pensiero lettere arti*, III(7):16–19. Casalini id: 5078080.

Teo, S. (2023). How i won singapore's gpt-4 prompt engineering competition. https://towardsdatascience.com/how-i-won-singapores-gpt-4-prompt-engineering-competition-34c195a93d4 Accessed: 2025-08-14.

Torres, B. (2024). Inteligência artificial "ressuscita" machado de assis, que comenta possível traição de capitu; veja a resposta. Accessed: 2024-07-03.