Curator-Supervised AI and Augmented Reality for Museums: A Conversational Assistant for Cultural Heritage Engagement

1st Federico Martusciello
DISIM
University of L'Aquila
L'Aquila, Italy
federico.martusciello@graduate.univaq.it

2nd Henry Muccini
DISIM
University of L'Aquila
L'Aquila, Italy
henry.muccini@univaq.it

***Abstract*—This paper presents a prototype museum application that combines marker- and model-based Augmented Reality with a curator-supervised, AI-powered virtual assistant. Built in Unity with AR Foundation and GPT-4-turbo, the system enables visitors to interact with artifacts via visual markers, accessing accurate, institutionally approved information through a conversational interface. By anchoring AI responses to curator-provided content, the system ensures historical accuracy while supporting voice-enabled, personalized learning. This hybrid framework offers a new model for cultural institutions—one where AI augments, rather than replaces, expert interpretation.**

***Index Terms*—Artificial Intelligence, Augmented Reality, Cultural Heritage, Conversational Systems, Interactive Learning, E-Tourism, E-Culture, E-Education**

I. Introduction

Cultural heritage institutions are increasingly adopting digital technologies to boost public engagement, yet ensuring educational integrity and interpretative accuracy remains a challenge with automated systems. Artificial intelligence (AI) and augmented reality (AR) are reshaping museum experiences by enabling immersive, interactive, and personalized engagement with history and place [1], [2]. While many institutions use these tools to enhance storytelling and accessibility, most AR applications still rely on static content or generic datasets, limiting adaptability and failing to reflect site-specific narratives [6].

To address these gaps, this research presents a context-aware, curator-supervised museum assistant. Developed as a Unity-based mobile application, the system uses AR to detect physical markers and model targets linked to specific artifacts, overlaying them with 3D content and activating a conversational interface. The assistant, powered by GPT-4-turbo, generates responses solely from curator-approved content, ensuring the AI serves as a trusted extension of institutional expertise rather than an autonomous narrator.

II. Proposed system

The application is developed using Unity and employs AR Foundation to support cross-platform AR capabilities (ensuring compatibility with ARCore/ARKit). The system includes an `AR module` for 3D content rendering, a `mobile interface` for text and voice interaction, and an `AI backend` powered by GPT-4-turbo that generates responses from linked, curator-approved datasets. The interaction is mediated by context-sensitive prompts that associate the recognized marker with the relevant artifact metadata and curator-provided content. The full system is illustrated in Fig. 1.

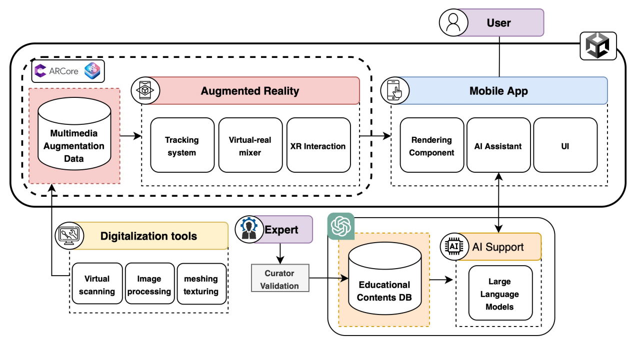

Fig. 1. System architecture of the proposed museum assistant. The mobile app integrates an AR module for marker tracking and augmentation, and an AI module for conversational interaction. Curators upload validated content to an *Educational Contents DB*, which the GPT-4-based assistant uses to generate responses.

Visitors interact with the system through a mobile application that enables AR-based exploration of cultural heritage spaces. The app uses AR to detect physical or model-based markers linked to artifacts, overlaying 3D content and triggering the assistant for text or voice interaction. The assistant may also be queried independently of AR scanning, functioning as a general-purpose tour companion.

III. AI Support module

A. *Curator-Supervised Content Framework*

A core innovation of the system is its reliance on a supervised knowledge base. Unlike chatbots trained on open web data, our assistant is limited to content explicitly curated and approved by domain experts. Using OpenAI's retrieval tools, museum professionals upload structured datasets, such

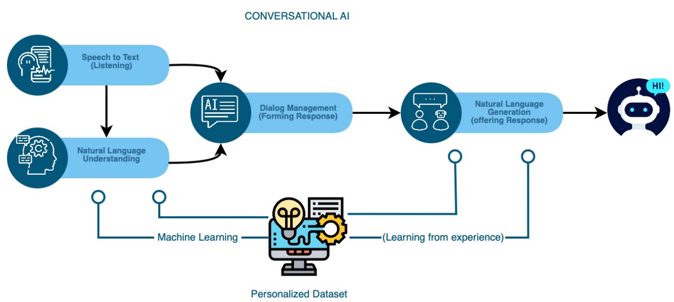

Fig. 2. Workflow of the conversational AI module. Visitor queries (voice or text) undergo STT conversion and natural language understanding, followed by dialogue management using GPT-4 to generate a response from the curator-approved dataset. The answer is then presented via natural language generation (text or spoken output) through the virtual assistant interface.

as JSON knowledge bases, FAQs, metadata, and multilingual narratives, which serve as the assistant's source of information.

These curated datasets serve as the sole reference for the assistant, allowing museum professionals to: (1) maintain control over terminological precision and historical interpretation, and (2) update content dynamically in response to new research or exhibit changes. By embedding curatorial authority into the system's knowledge architecture, the assistant acts as a mediated interface for the institutional voice, offering the flexibility of conversational AI without sacrificing trust or educational rigor. Recent studies highlight the need to ground museum chatbots in verified knowledge to prevent factual errors common in unrestricted generative models [4], [5]. Our supervised approach addresses this by ensuring all information stems from vetted sources.

B. Conversational AI Interaction

In addition to content supervision, the system employs a state-of-the-art conversational AI workflow to manage dialogue with visitors. As illustrated in Fig. 2, user speech is first transcribed into text using OpenAI's Whisper model for accurate and multilingual speech-to-text (STT) conversion, and then processed for intent via natural language understanding. The assistant's dialogue management component, powered by the GPT-4 model, formulates a response constrained to the curated knowledge base, and finally a natural language generation step delivers the answer (either as on-screen text or via voice synthesis through text-to-speech (TTS)). This pipeline allows the assistant to handle free-form questions while strictly adhering to the museum-provided content.

IV. Interaction Model and Educational Impact

The user experience is designed to be accessible, intuitive, and educational. When an artifact is recognized via marker or model-based AR tracking, the system overlays a 3D model and activates an interactive assistant. Visitors may pose any relevant questions, to which the assistant—constrained to curator-approved datasets-generates accurate responses in either text or synthesized multilingual voice. While it does not currently offer personalized learning paths, the system supports self-directed exploration within a controlled curatorial framework. Visitors engage freely with content aligned to their immediate interest, ensuring factual and pedagogical consistency across the experience.

By combining natural language interaction with AR augmentation, the system fosters meaningful, artifact-specific inquiry. Prior research confirms that such conversational interfaces enhance engagement, curiosity, and retention in heritage contexts [3]. This impact is reinforced through interpretative coherence, curator empowerment, and visitor engagement. All responses reflect institutional narratives; experts maintain control over content; and conversational interaction strengthens participation and learning.

V. Prototype Evaluation and Future Directions

A functional prototype was developed and tested with archaeological data, confirming the feasibility of the curator-supervised workflow and the assistant's ability to deliver consistent, institutionally approved responses. End-to-end latency was measured across 100 queries on a mid-range Android device over 4G, yielding a mean of 1.4 seconds (SD: 220 ms) and a 95th percentile of 1.8 seconds, including network and rendering overhead. Audio replies, incorporating neural TTS and streaming, added 2–3 seconds on average—perceptible but acceptable in museum contexts where clarity is prioritized. Future work will explore partial TTS streaming and response preloading to reduce delay. AR response time was assessed by measuring the delay from POI detection to full 3D model rendering. The system maintained 30–45 FPS depending on model complexity and lighting, using lightweight OBJ assets and asynchronous loading to prevent visual blocking. Tracking jitter remained below 2° in GPS mode and under 0.5° with visual markers, ensuring stable spatial alignment.

Ongoing development targets scalability, usability, and educational value. A web interface will support curator-managed content with access control. We also plan to integrate an interactive RAG-based dialogue system, enabling the assistant to ask clarifying questions, refine responses, and personalize learning pathways in real time.

References

[1] R. G. Boboc, E. Băutu, F. Gîrbacia, N. Popovici, and D. Popovici, "Augmented Reality in Cultural Heritage: An Overview of the Last Decade of Applications," *Applied Sciences*, vol. 12, no. 19, 2022.

[2] Y. Zhou *et al.*, "A meta-analytic review on incorporating virtual and augmented reality in museum learning," *Computers & Education*, vol. 191, Article 104515, 2022.

[3] S. Chai-Arayalert, S. Puttinaovarat, W. Saetang, *et al.*, "Chatbot-mediated technology to enhance experiences in historical textile museums," *Cogent Arts & Humanities*, vol. 11, no. 1, 2158660, 2024.

[4] P. K. Rachabatuni, F. Principi, P. Mazzanti, and M. Bertini, "Context-aware chatbot using multimodal large language models for cultural heritage," in *Proc. 15th ACM Multimedia Systems Conf. (MMSys)*, 2024.

[5] G. Trichopoulos, M. Konstantakis, G. Caridakis, A. Katifori, and M. Koukouli, "Crafting a Museum Guide Using ChatGPT4," *Big Data and Cognitive Computing*, vol. 7, no. 3, Article 148, 2023.

[6] Bekele, M. K., Pierdicca, R., Frontoni, E., Malinverni, E. S., Gain, J. (2018). A survey of augmented, virtual, and mixed reality for cultural heritage. Journal on Computing and Cultural Heritage (JOCCH), 11(2), 1-36.