

Kongque: An Interactive Conversational Virtual Agent Based on LA Chinatown Oral and Written Histories

Spencer Lin
Estuary
San Francisco, USA
spencer@estuary-ai.com

Isabel Fowlkes
School of Cinematic Arts
University of Southern California
Los Angeles, USA
ifowlkes@usc.edu

Karen Shieh
Estuary
San Francisco, USA
karen@estuary-ai.com

Stanley Lin
School of Cinematic Arts
University of Southern California
Los Angeles, USA
lin.stanley2000@gmail.com

Sharon Mozgai
Institute for Creative Technologies
University of Southern California
Los Angeles, USA
mozgai@ict.usc.edu

Scott Fisher
School of Cinematic Arts
University of Southern California
Los Angeles, USA
scott.fisher@usc.edu



Figure 1: Memory Space, a geolocated WebAR experience with IVAs based on oral & written histories of Old LA Chinatown

Abstract

Kongque is an Intelligent Virtual Agent developed for Memory Space: ChinatownAR, a mobile augmented reality experience that provides interactive access to the oral histories, archival materials, and community narratives of Los Angeles' original Chinatown. Kongque guides visitors through this lost community by engaging visitors in open-ended spoken dialogue grounded in historical source materials rather than scripted narration. This demonstration allows attendees to converse with Kongque on their own mobile devices in geolocated augmented reality and explore how embodied conversational agents can support engaging, character-driven access to community history. The demo also highlights the use of mobile AR and conversational AI to connect archival knowledge with contemporary audiences.

ACM Reference Format:

Spencer Lin, Isabel Fowlkes, Karen Shieh, Stanley Lin, Sharon Mozgai, and Scott Fisher. 2026. Kongque: An Interactive Conversational Virtual Agent Based on LA Chinatown Oral and Written Histories. In *ACM International Conference on Intelligent Virtual Agents (IVA 2026)*, September 07–11, 2026, Puebla, Mexico. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3806774.3832332>



This work is licensed under a Creative Commons Attribution 4.0 International License. *IVA 2026, Puebla, Mexico*

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2647-7/26/09
<https://doi.org/10.1145/3806774.3832332>

1 Introduction

Intelligent virtual agents (IVAs) have demonstrated the potential to provide interactive access to oral testimony and historical narratives [1, 9], and have been deployed increasingly in cultural heritage settings to support storytelling, engagement, and learning [6, 8]. Advances in large language models and generative AI now make it possible to move beyond scripted responses and fixed installations, opening new possibilities for generative IVAs grounded in community archival sources in place-based mobile AR experiences [5]. Los Angeles' original Chinatown, displaced in the 1930s to build what is now Union Station, is one such community that had much of its history preserved through oral histories and archival materials. We introduce Kongque, depicted in Figure 1, an intelligent virtual agent grounded in those histories, deployed in web augmented reality (WebAR) at the site where that community once stood to make the past visible through interactive open-ended dialogue. Inspired by the peacock motif found throughout Old Chinatown's architecture and cultural landscape, Kongque greets visitors at Union Station, answers questions, and reflects on the history of the place she inhabits. For this demonstration, attendees can interact with the system firsthand by scanning a QR code on their own device.

2 System Overview

2.1 Data Source

Memory Space: ChinatownAR is grounded in archival photography from The Huntington Library, historical census records, GIS mapping data, oral histories from former residents and descendants,

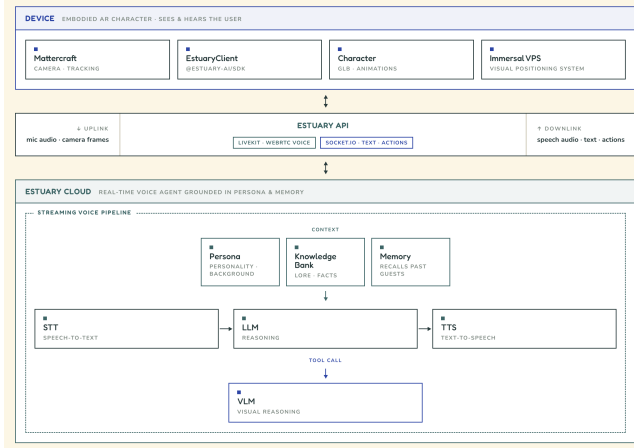


Figure 2: System diagram of server-client architecture

and community archival materials gathered through the Chinatown History Project. These sources were integrated to reconstruct the people, places, and lived experiences of Los Angeles’ original Chinatown prior to its demolition for Union Station.

2.2 Conversational System

Kongque is powered by Estuary [3, 4], a platform for multimodal conversational voice AI characters. The overview of the system is illustrated in Figure 2. Kongque’s answers are grounded through Retrieval-Augmented Generation (RAG) against a knowledge bank of oral and written histories of Old Chinatown. The RAG system is optimized for character voice interactions, achieving sub-millisecond latency through predictive chunk loading. It is inspired by VoiceAgentRAG [7] as well as utilizing Cache-Augmented Generation (CAG) for smaller corpora when possible. Multimodal interactions are powered by Gemini 2.5 Flash, a Vision Language model (VLM). To accommodate an international audience and ensure that characters accurately pronounce terms, we used Soniox V4 for multilingual code-switching Automatic Speech Recognition (ASR) and ElevenLabs 2.5 Flash for Text-to-Speech (TTS).

2.3 Virtual Agents

Several IVAs were built including Kongque and Peter Soohoo and David Lee, volumetrically captured humans seen in in Figures 1 and 3 respectively. They will use the same conversational and knowledge system. Kongque is a conversational anthropomorphic peacock who serves to tell the story of historic Chinatown. In bringing the history of this displaced community to life, the character’s primary role is to evoke the atmosphere of Chinatown’s Peacock Alley, fostering empathy and helping visitors imagine the experiences of those who once lived there. Her personality and character design efforts were pointed at this intent across three layers: a character specification for personality and delivery, a background context document to place character background and perspective, and a knowledge bank for historical accuracy. Kongque is embodied as a peacock because of its venerable symbolism in Chinese folklore and thematic appropriateness to Peacock Alley, an Old Chinatown



Figure 3: Talking volumetric capture of Peter Soohoo

street named for its wandering wild peacocks. This street is among those that were destroyed in the construction of Union Station.

We elected to use web augmented reality (WebAR) using Mattercraft [10] for better accessibility. The peacock is spatially localized by a Visual Positioning System (VPS) constructed in Immersal [2], using an interior scan of LA Union Station.

3 Demo Experience

Memory Space is launched by scanning a QR code. At Union Station, Kongque appears upon the recognition of VPS anchors, where she will pace around and examine the station. A non-VPS build will be used for the conference so Kongque can appear at the venue. Kongque recognizes the presence of visitors and proactively asks if they would like to learn about the history of Old Chinatown and her home, Peacock Alley. Users may engage in open-ended voice conversation. Kongque can both lead the conversation and follow the visitor’s interests through fluid turn-taking and user-initiated interruptions. Other IVAs such as David and Peter can be demoed as well. A video of our demo is linked.¹

4 Future Work

This demonstration establishes a foundation for future investigation of conversational IVAs in mobile AR cultural heritage experiences. We plan to use the same methodology to build additional IVAs with their own unique personalities and knowledge of Chinatown histories. Future work will also examine visitor engagement, social presence, perceived authenticity, and learning outcomes, as well as how open-ended dialogue with Kongque influences exploration of archival materials and connections to the history of Los Angeles’ original Chinatown. Memory Space is envisioned to be a reference design for other museums and organizations to build similar WebAR IVA experiences.

¹<https://youtu.be/KDN6jgxC31c>

References

- [1] Ron Artstein, Alesia Gainer, Kallirroi Georgila, Anton Leuski, Ari Shapiro, and David Traum. 2016. New dimensions in testimony demonstration. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. 32–36.
- [2] Immersal. 2026. Immersal: Accurate Device Positioning for Navigation & Visualization. <https://immersal.com/>. Part of Hexagon. Accessed: 2026-06-22.
- [3] Spencer Lin, Miru Jun, Basem Rizk, Karen Shieh, Scott S. Fisher, and Sharon A. Mozgai. 2025. Optimizing SIA Development: A Case Study in User-Centered Design for Estuary, a Multimodal Socially Interactive Agent Framework. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 688, 9 pages. <https://doi.org/10.1145/3706599.3707399>
- [4] Spencer Lin, Basem Rizk, Miru Jun, Andy Artze, Caitlin Sullivan, Sharon Mozgai, and Scott Fisher. 2024. Estuary: A Framework For Building Multimodal Low-Latency Real-Time Socially Interactive Agents. In *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents (GLASGOW, United Kingdom) (IVA '24)*. Association for Computing Machinery, New York, NY, USA, Article 50, 3 pages. <https://doi.org/10.1145/3652988.3696198>
- [5] Andrea Loretto, Valentine Bernasconi, Pasquale Cascarano, Alessio Trofai, Mengting Lai, Luca Vitale, Andrea Bortolotti, and Gustavo Marfia. 2026. Investigating the role of conversational agents in augmented cultural heritage tours. *Journal of Big Data* 13, 1 (2026), 84.
- [6] Octavian M Machidon, Mihai Duguleana, and Marcello Carrozzino. 2018. Virtual humans in cultural heritage ICT applications: A review. *Journal of Cultural Heritage* 33 (2018), 249–260.
- [7] Jieli Qiu, Jianguo Zhang, Zixiang Chen, Liangwei Yang, Ming Zhu, Juntao Tan, Haolin Chen, Wenting Zhao, Rithesh Murthy, Roshan Ram, et al. 2026. VoiceAgentRAG: Solving the RAG Latency Bottleneck in Real-Time Voice Agents Using Dual-Agent Architectures. *arXiv preprint arXiv:2603.02206* (2026).
- [8] Stella Sylaiou and Christos Fidas. 2022. Virtual humans in museums and cultural heritage sites. *Applied Sciences* 12, 19 (2022), 9913.
- [9] David Traum, Andrew Jones, Kia Hays, Heather Maio, Oleg Alexander, Ron Artstein, Paul Debevec, Alesia Gainer, Kallirroi Georgila, Kathleen Haase, et al. 2015. New Dimensions in Testimony: Digitally preserving a Holocaust survivor’s interactive storytelling. In *International Conference on Interactive Digital Storytelling*. Springer, 269–281.
- [10] Zappar Ltd. 2026. Mattercraft: Build Rich 3D Experiences for the Web. <https://zap.works/mattercraft/>. Zapworks. Accessed: 2026-06-22.