

Lumière: A Holographic Polaroid Camera for Creating Multiplatform Object-Grounded Intelligent Virtual Agents

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

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

Tiger Jewell-Alibhai
tigesky@gmail.com
Cambridge, United States

Sharon Mozgai
mozgai@ict.usc.edu
University of Southern California
Los Angeles, United States

Scott Fisher
scott.fisher@usc.edu
University of Southern California
Los Angeles, United States

Stanley Lin
lin.stanley2000@gmail.com
University of Southern California
Los Angeles, United States



Figure 1: Lumière: a custom Polaroid-style camera (left) with an integrated Pepper's Ghost display (second left), WebAR interaction with a generated intelligent virtual agent (second right), and a printed character card artifact (right).

Abstract

Lumière is a holographic camera that creates persistent, object-grounded intelligent virtual agents. Lumière transforms captured images of everyday objects into embodied, conversational agents with unique personalities, voices, and 3D avatars based on a vision language model's interpretation of the image. All agents and avatars are created generatively without any pre-authoring and appear through a transparent Pepper's Ghost display. Estuary provides a persistent backend runtime that maintains each agent's identity and memory across multiple platforms including web, mobile, extended reality devices. The demo showcases end-to-end agent creation, conversational interaction, and multiplatform migration of the agent via share token through near-field communication from Lumière onto mobile web augmented reality.

ACM Reference Format:

Spencer Lin, Karen Shieh, Tiger Jewell-Alibhai, Sharon Mozgai, Scott Fisher, and Stanley Lin. 2026. Lumière: A Holographic Polaroid Camera for Creating Multiplatform Object-Grounded Intelligent Virtual Agents. In *ACM International Conference on Intelligent Virtual Agents (IVA '26)*, September

7–11, 2026, PUEBLA, Mexico. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 Introduction

Intelligent Virtual Agents (IVA) are increasingly able to converse and remember interactions in ways that support richer forms of social interaction. However, many agents still require predefined avatars or manually authored personas and are confined to one device. Lumière explores a different model: everyday physical objects as the origin points for persistent, embodied, conversational agents. Everyday possessions often function as extensions of personal identity and are associated with important memories, relationships, and life experiences [4, 5]. Creating IVAs from personal belongings presents a unique method of leveraging existing sentimental value and is further grounded in the user's environment when displayed in Lumière as a Pepper's Ghost illusion [1].

Lumière leverages the Estuary [2, 3] agent runtime for image-to-agent creation and multiplatform migration of agents. This allows a character created on Lumière to carry its identity and memories across web, mobile extended reality devices, or as a printed Polaroid-style artifact with an embedded near-field communication (NFC) tag that brings the agent into a web augmented reality (WebAR) experience built on Mattercraft [6].

2 System Overview

An overview of the system is illustrated in Figure 3.

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 '26, September 7–11, 2026, PUEBLA, Mexico
© 2026 Copyright held by the owner/author(s).
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

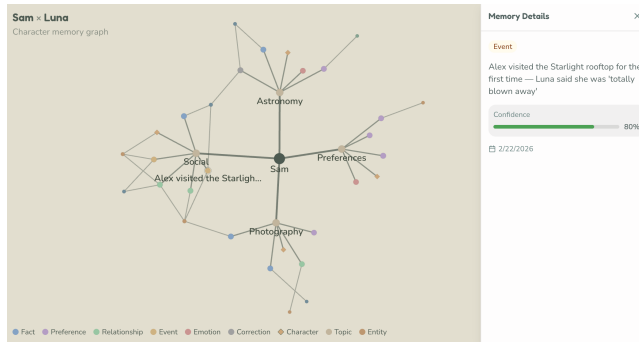


Figure 2: Example memory graph showcasing how agents organize and carry memories across platforms

2.1 Object-Grounded Agent Generation

A captured image is sent to Estuary’s agent generation pipeline where a Vision Language Model (VLM) and generative 3D model then creates an agent with a unique personality, voice, backstory, and 3D avatar based on the semantic details of the photo as illustrated in in Figure 1. Environmental factors such as lighting and background can contribute to the agent’s persona as well. For example, stark lighting from a single desk lamp can evoke the anxious persona of a nervous stage technician. The image constrains and inspires the agent, but the result is a new agent rather than a literal reconstruction or scan. This design approach shifts the focus from object digitization to character creation, allowing personally meaningful artifacts to serve as grounding sources for persistent intelligent virtual agents and allows everyday objects to become expressive characters capable of conversation, storytelling, and social interaction rather than static digital reproductions.

2.2 Multiplatform Agent Runtime

Estuary maintains the agent’s identity and memory across multiple platforms. This allows the same agent to appear in the holographic device, a mobile or web interface, AR or VR, or even be associated with a printed artifact through a physical near-field communication (NFC) tag as seen in Figure 1.

Memory is stored in a vector database for fast lookup and visualization as seen in Figure 2. Extraction and de-duplication occurs at the end of each conversational turn using a Large Language Model (LLM). An embedding model produces embeddings for efficient storage and access in the vector database. Rather than pushing all memories into the context window, memory is accessed as needed through fast vector searches each turn or skipped if the memory is already loaded. Voice conversations often change topics, so memories are shuffled in and out of the context window as needed, reducing context bloat and latency.

As a demonstration of this multiplatform architecture, Lumière creates an instance of an agent and stores a share token to an NFC tag. Upon tapping an NFC-enabled smartphone, Estuary redirects the user to a WebAR conversational experience that dynamically loads the agent’s avatar, personality and voice configuration, and memories.

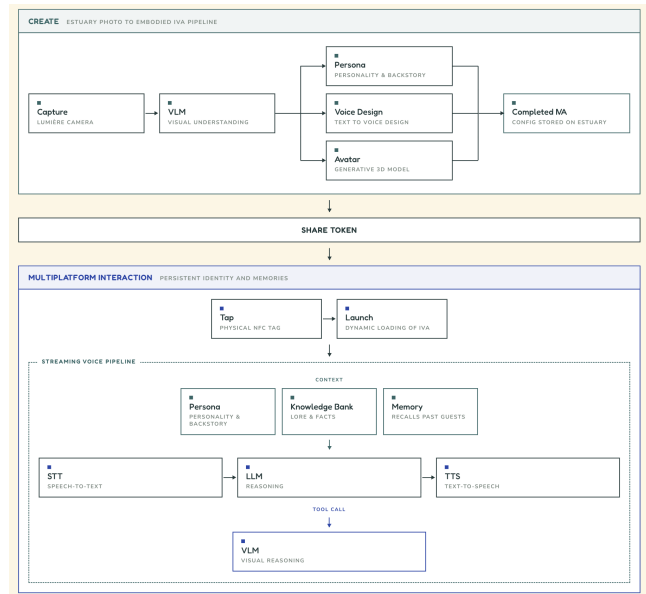


Figure 3: System diagram encompassing creation and multiplatform interaction

3 Demonstration Experience

Guests begin by selecting an everyday object and scanning it with Lumière. The system then generates a character from the scan and displays it inside the Pepper’s Ghost viewer. Guests can speak with the character through the device while seeing the surrounding room through the display, making the character appear in relation to the same environment where it was created. Guests can then print a Polaroid-style card of the character and tap its embedded NFC tag to open and interact with the same agent in a WebXR experience on a phone, laptop, or tablet. A video of our demo is linked.¹

4 Conclusion

Lumière explores two novel concepts: how personal belongings can be used as a basis for bespoke IVA creation that leverages existing sentimental value and how agents can migrate across multiple platforms while retaining their identity and memories. Lumière’s physical design and multiplatform migration method via printed artifacts presents a familiar, tangible means of creating and sharing IVAs.

References

- [1] Beth A. Kattelman. 2013. *Spectres and Spectators: The Poly-Technologies of the Pepper’s Ghost Illusion*. Palgrave Macmillan UK, London, 198–213. https://doi.org/10.1057/9781137319678_11
- [2] 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>

¹https://youtu.be/IWpCc4ks_mE

- [3] Spencer Lin, Basem Rizk, Miru Jun, Andy Artze, Caitlín 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>
- [4] Elise Van den Hoven, Daniel Orth, and Annemarie Zijlema. 2021. Possessions and memories. *Current opinion in psychology* 39 (2021), 94–99.
- [5] S Christian Wheeler and Christopher J Bechler. 2021. Objects and self-identity. *Current Opinion in Psychology* 39 (2021), 6–11.
- [6] Zappar Ltd. 2026. Mattercraft: Build Rich 3D Experiences for the Web. <https://zap.works/mattercraft/>. Zapworks. Accessed: 2026-06-22.

5 Demo Logistics

We'll need a table, WiFi connection, and a small area (about 5 square feet) in front of the table for guests to gather and interact with Lumière. We will bring the Lumière device to the demo session.