AugTwins: Interactive Digital Twins as an Inquiry into Identity and Memory

Anushhka Thakur[*]
Savannah College of Art and Design
athaku20@student.scad.edu

Mateo Larrea[*]
Stanford University
mlarreaf@stanford.edu

Lars Langenbach[*]
Savannah College of Art and Design
llange21@student.scad.edu

Dünya Baradari[*]
MIT Media Lab
dunya@mit.edu

Abstract

AugTwins is a pipeline for creating interactive human digital twins that engage users in natural, face-to-face conversation. By integrating photorealistic embodiment (Unreal Engine MetaHuman), voice cloning, and a hybrid multi-agent memory architecture, the system enables realistic encounters with digital representations of real people. Each twin is grounded in the person's life story, experiences, and knowledge while maintaining conversational continuity across sessions. The architecture combines fast pattern matching for common queries with deep semantic retrieval for biographical details, coordinated through metacognitive monitoring that guides socially appropriate responses. We present two deployed twins demonstrating faithful personality representation and accurate biographical recall. Beyond technical implementation, *AugTwins* invites reflection on how memory and embodiment shape identity, and what it means to extend human presence through computational reconstruction.

1 Vision & Background

You sit across from yourself, not a mirror but a presence. The twin leans forward and says, "Good to see you again. I was thinking about what you told me last time." It remembers.

AugTwins explores the Human Digital Twin (HDT) [1]: a persistent, interactive double that looks, speaks, and behaves like its human counterpart. HDTs may soon attend meetings [2], negotiate on our behalf, filter potential matches [3, 4, 5], or participate in simulated governance [6]. Our focus is subjective and social presence, informed by work on generative agents [7], synthetic personae [8], and AI companions [9]. Meeting your double turns identity into performance and memory into a living, revisable record.

2 Description of the Work and Roles of AI/ML

Overview. The installation presents a conversational, photorealistic digital twin rendered in Unreal Engine (UE) 5.6 [10]. Dialogue is generated by a multi-agent orchestrator (implemented with LangChain [11]) using Groq Llama 3.3, a large language model (LLM) [12, 13], grounded in the person's biographical data through a hybrid multi-layer retrieval system. Visitor speech is transcribed with automatic speech recognition (ASR) using Whisper [14], then voiced with neural text-to-
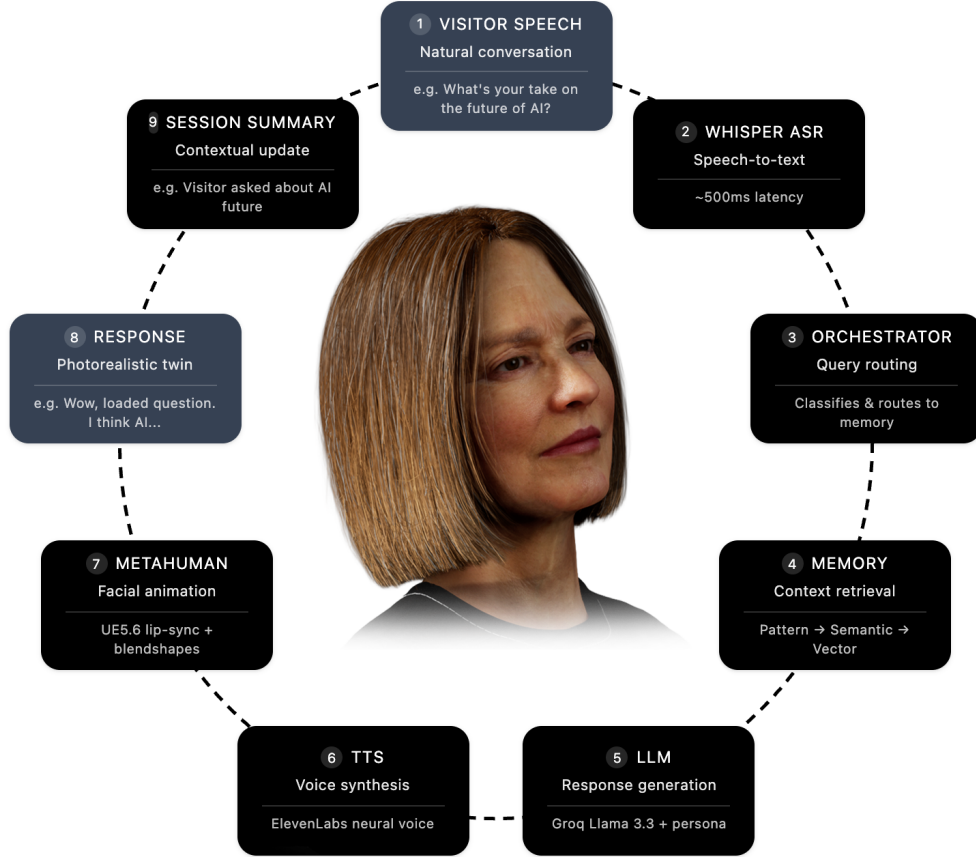
---

[*]Equal contribution.

Figure 1: End-to-end interaction pipeline. Speech is transcribed, routed for retrieval and persona-constrained generation, synthesized to voice, and lip-synced to the MetaHuman avatar; session memory is written back for continuity.

speech (TTS) [15] and lip-synced on a custom MetaHuman avatar [16]. Session summaries extend conversational memory for continuity across visits. See Figure 1 for the end-to-end pipeline.[2]

2.1   Avatar Construction Pipeline

We follow a standard photo-to-MetaHuman workflow: multi-angle photo capture, single-subject head reconstruction, topology transfer to a MetaHuman base, texture refinement, and UE integration with runtime lip sync and a lightweight interaction state machine. Implementation details are depicted in Figure 2.

---

[2]We experimented with alternative stacks (e.g., GPT-4-class models and other ASR systems) and selected this configuration for latency, robustness, and controllability.

Avatar Construction Pipeline
from photographs to a photorealistic digital twin in Unreal Engine

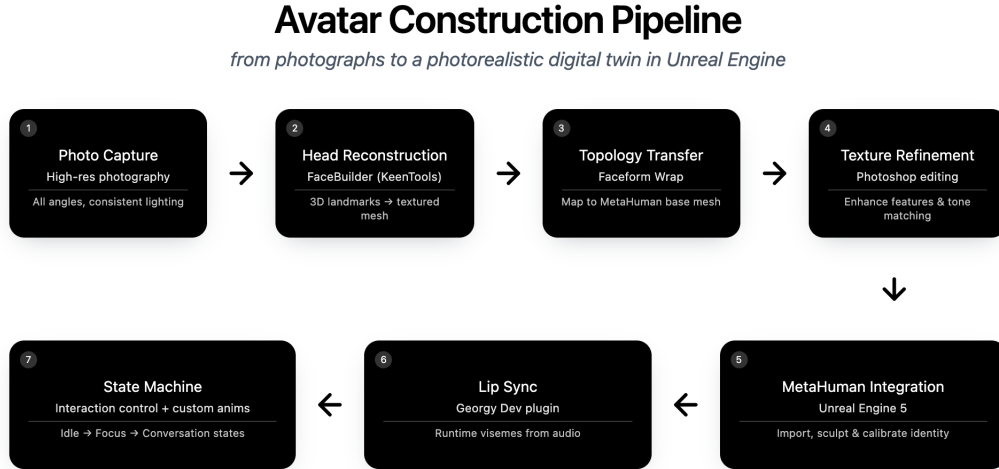| 1 | Photo Capture | High-res photography | All angles, consistent lighting |
2	Head Reconstruction	FaceBuilder (KeenTools)	3D landmarks → textured mesh
3	Topology Transfer	Faceform Wrap	Map to MetaHuman base mesh
4	Texture Refinement	Photoshop editing	Enhance features & tone matching
5	MetaHuman Integration	Unreal Engine 5	Import, sculpt & calibrate identity
6	Lip Sync	Georgy Dev plugin	Runtime visemes from audio
7	State Machine	Interaction control + custom anims	Idle → Focus → Conversation states

Figure 2: Avatar construction pipeline: from photo capture to head reconstruction, topology transfer, texture refinement, and Unreal Engine integration with runtime lip sync and state-based interaction.

2.2 Hybrid Multi-Agent Memory Architecture

A coordinator classifies queries and composes results from fast and deep memory tiers (Figure 3). Concretely, we use (i) a *pattern* tier for greetings and etiquette, (ii) a *semantic cache* for frequent question–answer (Q&A) pairs, (iii) curated *declarative* facts stored as JSON, (iv) *hybrid vector/lexical* search over documents and transcripts using the Qdrant vector database and FAISS with BM25 lexical retrieval, and (v) longer-form *procedural* memory for session summaries.

2.3 System Components

- **Orchestrator.** Classifies intent (identity, research, conversational, general) and supervises retrieval and persona-constrained generation by a large language model (LLM) [12, 13].

- **ASR & Voice.** Whisper provides robust streaming automatic speech recognition (ASR) [14]; ElevenLabs provides expressive low-latency text-to-speech (TTS) [15].

- **Avatar.** Custom MetaHuman head in UE 5.6 with runtime lip sync and state-based attention and gaze control [16, 10].

- **Memory hygiene.** Session summaries are deduplicated into both the semantic cache and procedural store to maintain continuity without drift.

Roles of Artificial Intelligence and Machine Learning (AI/ML). AI/ML models drive multi-agent orchestration, memory retrieval, language generation, ASR, TTS, and real-time facial animation, sustaining a coherent digital twin grounded in the person's life story with social presence.

3 Addressing the Theme of Humanity

AugTwins explores identity as shaped by continuity of memory and coherent narrative rather than biological presence. By grounding responses in the person's life story, recalling past encounters, maintaining stable conversational style, and following social norms, the twin invites relationships that feel authentic and grounded in shared history.

As fidelity rises, privacy, consent, parasocial attachment, and impersonation risks become central. We handle provenance, consent boundaries, and deferral when uncertain. The piece invites reflection on what aspects of presence are replicable, what remains uniquely human, and what protections are needed as these qualities are reproduced by machines.

AugTwin Cognitive Architecture

Multi-Agent System with Dual-Speed Cognition & Metacognitive Awareness

USER QUERY

↓

MULTI-AGENT ORCHESTRATOR
Routes queries & coordinates parallel memory retrieval

―― FAST

SLOW ――

<1ms	5-10ms	10-50ms	50-100ms	200-800ms
PATTERN MATCH	SEMANTIC CACHE	JSON LOOKUP	VECTOR SEARCH	MEM0 CLOUD

↓ ↓ ↓ ↓ ↓

Episodic Memory	Semantic Cache	Declarative KB	Vector DB	Procedural Mem0
e.g. ~200 messages	e.g. 8 queries	e.g. 31 people	Qdrant+Mods	Unlimited
Conversational context	Reusable queries	MentalModel JSON	1536-dim vectors	Graph Memory API
L1	L2	L3	L4	L5

RESULT AGGREGATION + DEDUPLICATION
Combines tiers + Graph context + Fusion detection

↓

LLM GENERATION (GROQ/OPENAI)
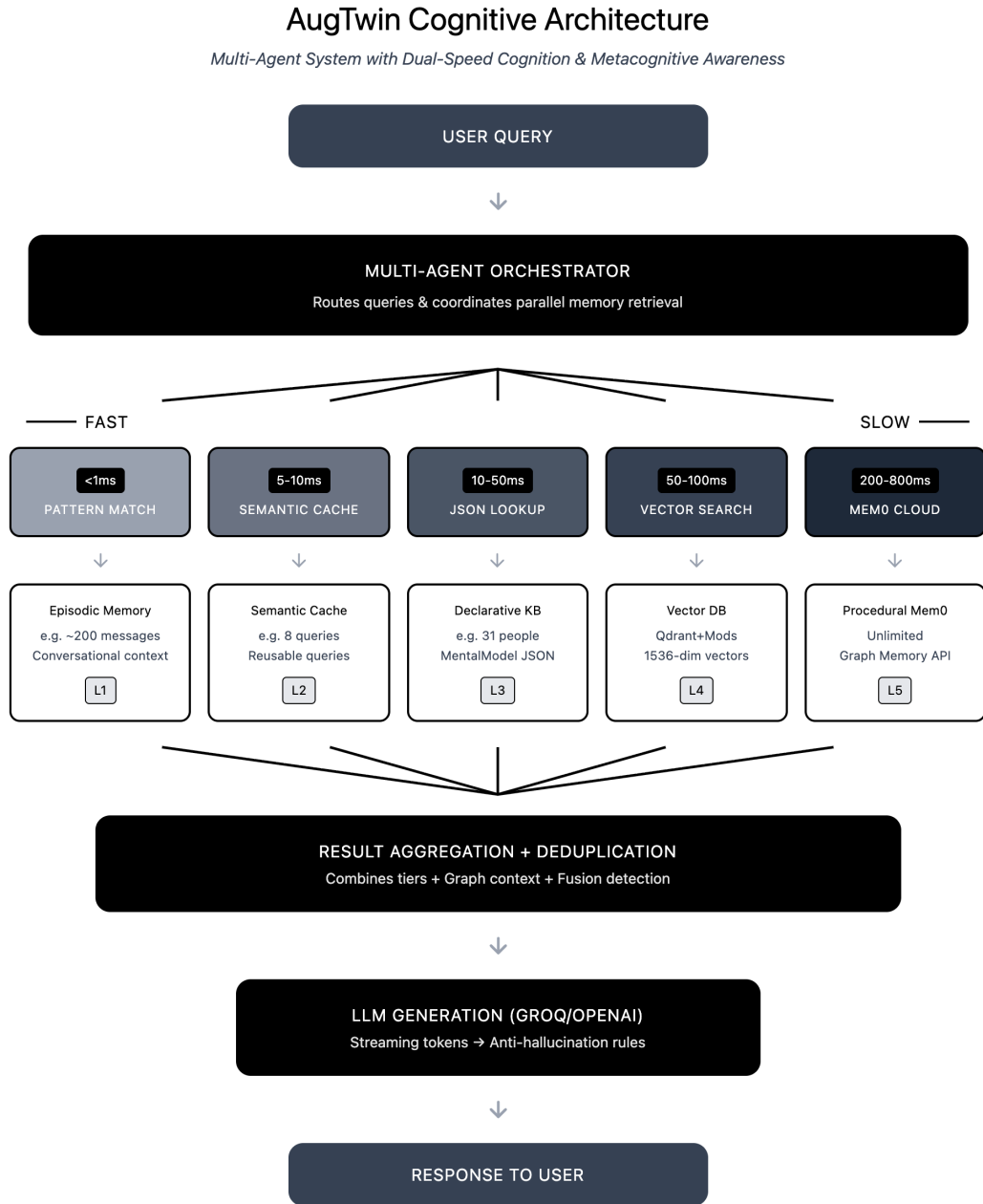Streaming tokens → Anti-hallucination rules

↓

RESPONSE TO USER

Figure 3: Hybrid memory architecture: a coordinator routes to pattern, semantic, declarative, hybrid search, and procedural tiers, aggregates with metacognitive checks, then triggers persona-constrained generation.

Author Biographies

Anushhka Thakur is an immersive designer and developer who creates interactive experiences using artificial intelligence (AI), extended reality (XR), and game design. Her projects include Crime Scene Assistant and an AI-guided Natural History Museum virtual reality (VR) experience.

Mateo Larrea is a researcher and Master's student at Stanford's Center for Computer Research in Music and Acoustics (CCRMA) working at the intersection of human–computer interaction (HCI), auditory neuroscience, and generative systems. He created LoveSims (CHI '25) and Bio-Graphic Notations (Massachusetts Institute of Technology, MIT Media Lab, 2021).

Lars Langenbach is an immersive reality artist whose work blends virtual and physical environments through VR, AR, and interactive installations.

Dünya Baradari is a scientist and engineer at the MIT Media Lab focused on human digital twins, personal analytics, and neuroadaptive brain–computer interfaces.

Project repository: github.com/AugTwins
3-minute video preview: AugTwins NeurIPS

References

[1] Martin Wolfgang Lauer-Schmaltz, Philip Cash, John Paulin Hansen, and Anja Maier. Towards the human digital twin: Definition and design—a survey. *arXiv*, February 2024. URL https://arxiv.org/abs/2402.07922.

[2] Joanne Leong, John Tang, Edward Cutrell, Šaša Junuzovic, Gregory Paul Baribault, and Kori Inkpen. Dittos: Personalized, embodied agents that participate in meetings when you are unavailable. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2):1–28, November 2024. doi: 10.1145/3687033. URL https://doi.org/10.1145/3687033.

[3] Dünya Baradari, Tejaswi Polimetla, and Pattie Maes. Data-driven AI avatars for valuation in dating scenarios. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–13, April 2025. doi: 10.1145/3706599.3719833. URL https://doi.org/10.1145/3706599.3719833.

[4] Mateo Larrea, Xingyi Zhang, and Xuyang Zhu. Lovesims: Exploring 'what-if' scenarios for relationship insights and compatibility. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–7, April 2025. doi: 10.1145/3706599.3720011. URL https://doi.org/10.1145/3706599.3720011.

[5] Leonard Foner. Yenta: A multi-agent, referral-based matchmaking system. In *Proceedings of the First International Conference on Autonomous Agents (AGENTS '97)*, pages 301–307, February 1997. doi: 10.1145/267658.267732. URL https://doi.org/10.1145/267658.267732.

[6] Joon Sung Park, Carolyn Q. Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S. Bernstein. Generative agent simulations of 1,000 people. *arXiv*, November 2024. URL https://arxiv.org/abs/2411.10109.

[7] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*, number Article 2, pages 1–22, October 2023. ISBN 9798400701320. doi: 10.1145/3586183.3606763. URL https://doi.org/10.1145/3586183.3606763.

[8] Mirjana Prpa, Giovanni Maria Troiano, Matthew Wood, and Yvonne Coady. Challenges and opportunities of LLM-based synthetic personae and data in HCI. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–5, May 2024. doi: 10.1145/3613905.3636293. URL https://doi.org/10.1145/3613905.3636293.

[9] Baris Demiralay. The rise of AI companions: Psychological effects of artificially intelligent relationships, 2024. Preprint/working paper.

[10] Epic Games. Unreal engine 5.6. https://www.unrealengine.com/, 2025. Accessed: 2025-10-29.

[11] LangChain. Langchain. https://python.langchain.com, 2025. Accessed: 2025-10-29.

[12] Groq inference. https://groq.com, 2025. Accessed: 2025-10-29.

[13] Llama 3.3 model card. https://ai.meta.com/llama, 2025. Accessed: 2025-10-29.

[14] Whisper speech recognition. https://github.com/openai/whisper, 2023. Accessed: 2025-10-29.

[15] ElevenLabs. Professional voice cloning. https://elevenlabs.io, 2025. Accessed: 2025-10-29.

[16] Epic Games. Metahuman creator. https://www.unrealengine.com/en-US/metahuman, 2025. Accessed: 2025-10-29.