Immersive Conversations with Digital Einstein: Linking a Physical System and AI

Rafael Wampfler
ETH Zurich, Switzerland
wrafael@inf.ethz.ch

Nikola Kovačević
ETH Zurich, Switzerland
nikolak@inf.ethz.ch

Philine Witzig
ETH Zurich, Switzerland
pwitzig@inf.ethz.ch

Chen Yang
ETH Zurich, Switzerland
yangche@inf.ethz.ch

Markus Gross
ETH Zurich, Switzerland
grossm@inf.ethz.ch

Figure 1: The Digital Einstein Platform. Users interact with a virtual representation of Albert Einstein while sitting in an immersive setup consisting of a screen (1), a table with a lamp (2), decorative elements (3), an armchair (4), and a carpet (5). The user's characteristics and behavior are inferred from a camera at the top of the screen. A microphone hidden in a book on the table captures the user's voice. Speakers in the chair, under the table, and behind the screen create a spatial sound experience.

ABSTRACT

We present Digital Einstein, an immersive platform that enables interactive conversations with a digital representation of Albert Einstein. The physical setup includes a screen, armchair, camera, speakers, microphone, and decorative elements. Utilizing advanced technologies such as natural language processing, machine learning, and dynamic rendering, the platform simulates realistic dialogs with Einstein on various topics about his life and research. The system features two versions: Version 1 uses a dialog tree, motion-captured animations, and pre-recorded speech for fluid, controllable interactions. In Version 2, users can choose between GPT-4o-based and Llama 3-based chatbots for more open and engaging conversations. Data-driven models synthesize Einstein's voice and generate corresponding animations on the fly, while context-specific images are automatically generated to visually enrich the conversation. Digital Einstein offers a glimpse into the future of immersive, personalized dialog experiences with intelligent characters.

SA Emerging Technologies '24, December 03–06, 2024, Tokyo, Japan

ACM ISBN 979-8-4007-1137-4/24/12
https://doi.org/10.1145/3681755.3688939

CCS CONCEPTS

• **Hardware** → **Emerging technologies**; • **Computing methodologies** → **Animation**; **Natural language processing**; • **Human-centered computing** → **Human computer interaction (HCI)**.

KEYWORDS

Conversational Digital Characters, Digital Einstein, Animation, Speech Recognition, Speech Synthesis, Large Language Models

ACM Reference Format:
Rafael Wampfler, Nikola Kovačević, Philine Witzig, Chen Yang, and Markus Gross. 2024. Immersive Conversations with Digital Einstein: Linking a Physical System and AI. In *SIGGRAPH Asia 2024 Emerging Technologies (SA Emerging Technologies '24), December 03–06, 2024, Tokyo, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3681755.3688939

1 INTRODUCTION

Conversational digital characters have gained significant interest, with applications in healthcare [Bin Sawad et al. 2022], education [Pataranutaporn et al. 2021], and entertainment [Casas and Mitchell 2023]. Powered by advancements in natural language processing and machine learning, these characters engage in meaningful dialogs, offering interactive and immersive experiences [Casas and Mitchell 2023]. However, most characters remain confined to the digital realm, limiting interaction depth. To enhance user engagement, researchers are exploring augmented reality setups,

providing a more holistic experience with tangible presence [Weidner et al. 2023]. Despite progress, developing a comprehensive interactive physical setup remains challenging.

This work introduces the Digital Einstein platform, which combines conversational digital characters with an immersive physical setup (see Figure 1). Users engage in interactive dialogs with a digital Albert Einstein, with dynamic expressions and body language, offering a more immersive experience than passive viewing [Helzle and Goetz 2018]. By leveraging natural language processing, speech synthesis, and animation synthesis, Digital Einstein transcends mortality, bringing a historic figure back to life.

2 SYSTEM

The Digital Einstein platform includes a physical setup (see Figure 1) and a modular system that drives user conversations based on input from the platform's hardware (see Figure 2). The modular design allows easy replacement of individual software components. We distinguish between two conversational cores, each providing a slightly different experience. To avoid the uncanny valley, we chose a stylized representation of Albert Einstein.

The Platform. The platform's design is reminiscent of the early 20th century. It consists of an armchair, a carpet, a table, decorative elements (a lamp, books, a Mozart bust, and a pocket watch), a screen with a wooden frame, several speakers, and a media box (a computer and an A/V receiver) hidden behind the screen. While comfortably sitting in the chair, the user can speak to Digital Einstein through a microphone hidden in the top book on the table. Microsoft's Azure speech recognition transcribes the user's speech to text. A hidden camera at the top of the frame tracks people's characteristics and behavior. A spatial sound experience is created by speakers built into the chair (2 speakers), below the table (1), and behind the screen (2).

Digital Einstein 1.0. Motion-captured facial and body animations are combined with pre-recorded speech from an artist mimicking Einstein's German accent and movements. Responses are generated from a dialog tree in Rasa, covering topics like Einstein's theories, academic journeys, and relationships. The algorithm introduces randomness, ensuring the conversation evolves in unexpected ways. For example, if Einstein observes an interlocutor's attention drifting (inferred through the camera), he will promptly inquire if the interlocutor is disengaged or uninterested.

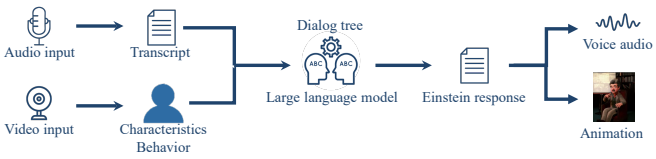

Figure 2: System Overview. The user's transcribed speech and the user's characteristics and behavior inferred from the camera are fed to a dialog tree or LLM-based chatbot. Einstein's voice is synthesized from the produced response, and facial and body animations are generated accordingly.

Digital Einstein 2.0. To enhance conversation flexibility, users can choose between a GPT-4o-based chatbot [Achiam et al. 2023] and a Llama 3 8B-based chatbot [Touvron et al. 2023]. The GPT-4o chatbot follows dialog tree topics via prompt engineering, while the Llama 3 chatbot is fine-tuned with synthetic Einstein conversations generated by GPT-4. The chatbots are prompted with the user's characteristics and behavior inferred from the camera. Einstein's text responses are synthesized into speech using a Microsoft Azure neural model, fine-tuned on over 2,000 recordings of an artist imitating Einstein's voice. Based on the generated speech signal, the facial animations are generated using a data-driven model. Body animations are blended from motion-captured sequences based on the avatar's state (idle, listening, speaking). Lastly, we regularly generate and display an image illustrating the conversation topic, using GPT-4o to create a prompt and Midjourney for image creation.

3 EXPERIENCE

Our platform allows visitors to interact with a digital Albert Einstein in an immersive setup. When a visitor sits down (detected by the camera), the conversation begins automatically with Einstein's greeting. In version 1.0, Einstein guides the user through an entertaining conversation about his life and research, with added fun elements like guessing games. In version 2.0, users can freely ask questions on topics that interest them. Visitors can end the conversation at any time by standing up or verbally finishing, allowing the next visitor to engage with Digital Einstein.

4 CONCLUSION & FUTURE WORK

The Digital Einstein platform combines advanced AI with a physical setup, allowing users to engage in dynamic conversations with a digital Albert Einstein. Using natural language processing, speech synthesis, and dynamic animation, it offers an engaging dialog experience. This platform highlights the potential of digital immortality, preserving and reviving the legacies of influential figures for future generations. Future developments will expand conversational capabilities, integrate more historical figures, and enhance interaction through improved animation and emotional responsiveness.

REFERENCES

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

Abdullah Bin Sawad, Bhuva Narayan, Ahlam Alnefaie, Ashwaq Maqbool, Indra Mckie, Jemma Smith, Berkan Yuksel, Deepak Puthal, Mukesh Prasad, and A. Baki Kocaballi. 2022. A systematic review on healthcare artificial intelligent conversational agents for chronic conditions. *Sensors* 22, 7 (2022), 2625.

Llogari Casas and Kenny Mitchell. 2023. Intermediated reality with an AI 3D printed character. In *ACM SIGGRAPH 2023 Real-Time Live!* 1–2.

Volker Helzle and Kai Goetz. 2018. Digital Albert Einstein, a case study. In *ACM SIGGRAPH 2018 Talks*. 1–2.

Pat Pataranutaporn, Valdemar Danry, Joanne Leong, Parinya Punpongsanon, Dan Novy, Pattie Maes, and Misha Sra. 2021. AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence* 3, 12 (2021), 1013–1022.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

Florian Weidner, Gerd Boettcher, Stephanie Arevalo Arboleda, Chenyao Diao, Luljeta Sinani, Christian Kunert, Christoph Gerhardt, Wolfgang Broll, and Alexander Raake. 2023. A Systematic Review on the Visualization of Avatars and Agents in AR & VR displayed using Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* (2023).