

TLIC PAPER

AI-Powered Avatars as Expert Mentors: Knowledge Transfer and Professional Development within Multimodal Learning Environments

Fernando Salvetti¹  ,
Barbara Bertagni¹,
Roberto Pieraccini²

¹e-REAL Labs at Logosnet,
Turin, Italy – Lugano,
Switzerland – Houston, TX,
USA – New York, NY, USA

²Scientific Advisor,
New York, NY, USA

salvetti@logosnet.org

ABSTRACT

The growing demand for personalized, scalable, and high-quality learning experiences has accelerated the adoption of AI-powered avatars in professional training, higher education, and organizational learning. These avatars—also referred to as Embodied Conversational Agents (ECAs)—simulate expert mentors who provide adaptive guidance, enabling learners to acquire complex cognitive, behavioral, and communicative skills in realistic scenarios. By leveraging advances in natural language processing, affective computing, and multimodal interaction design, avatars embody a new paradigm of virtual coaching that blends scalability with human-like social presence. This paper presents Mia, an intelligent avatar developed within the e-REAL ecosystem, designed to function as an *expert mentor* across multiple domains, including healthcare, legal practice, corporate leadership, and education. Unlike generic digital assistants, Mia operates within a human-centric framework that prioritizes empathy, trust, and inclusivity. The avatar's adaptive capabilities are powered by a layered architecture combining retrieval-augmented generation (RAG), a flexible orchestration system, and real-time speech and gesture synthesis. This architecture ensures reliable knowledge transfer while minimizing hallucinations and aligning with ethical and regulatory standards such as the EU AI Act and GDPR. The introduction of avatars like Mia raises important questions about how embodiment, proxemics, and non-verbal communication influence learning outcomes, particularly in immersive environments where presence and credibility strongly affect engagement. Research on social robotics and immersive virtual environments shows that avatars capable of managing turn-taking, gaze, and emotional tone foster stronger identification and learning retention compared to text-based interfaces. By combining these insights with the flexibility of generative AI, virtual mentors can deliver personalized, repeatable, and unbiased coaching at scale. We argue that avatars such as Mia exemplify the convergence of pedagogy, cognitive science, and technological innovation, pointing toward a future where expert mentoring is no longer bound by time, location, or human availability. In this vision, AI-powered avatars do not replace human educators but amplify their reach, ensuring that high-quality expertise is available across contexts, cultures, and languages.

Salvetti, F., Bertagni, B., Pieraccini, R. (2026). AI-Powered Avatars as Expert Mentors: Knowledge Transfer and Professional Development within Multimodal Learning Environments. *International Journal of Advanced Corporate Learning (ijAC)*, 19(1), pp. 63–72. <https://doi.org/10.3991/ijac.v19i1.58759>

This article is an expanded version of a paper presented at The Learning Ideas Conference, held in New York, NY, USA, June 11–13, 2025. Article submitted 2025-09-21. Revision uploaded 2025-10-15. Final acceptance 2025-11-18.

© 2026 by the authors of this article. Published under CC-BY.

KEYWORDS

AI-powered avatars, expert mentoring, virtual coaching, human-centered AI, embodied conversational agents, knowledge transfer, multimodal learning

1 INTRODUCTION: AN INTELLIGENT AVATAR AS AN EXPERT MENTOR

The rapid development of artificial intelligence (AI) and immersive technologies is changing how knowledge is shared, practiced, and remembered in professional and educational settings. Traditional mentoring and expert coaching, while useful, face limitations due to human availability, scalability, and expense. At the same time, the increasing complexity of healthcare systems, legal practices, corporate leadership, and academic training requires ongoing learning opportunities that are both tailored and repeatable [1–2].

Against this backdrop, the development of AI-powered avatars—also known as Embodied Conversational Agents (ECAs)—introduces new possibilities for delivering high-quality expertise at scale. These avatars simulate the presence of a human mentor, offering learners contextualized feedback, adaptive dialogue, and emotional resonance within immersive environments [3]. Unlike static e-learning modules or text-based chatbots, avatars integrate voice, gesture, gaze, and affective responses, thus fostering a stronger sense of social presence and credibility, both of which are key factors in effective learning [4–5].

This paper presents Mia, an intelligent avatar conceived as an *expert mentor*. Mia embodies the principles of human-centered AI, emphasizing empathy, inclusivity, and trust while providing learners with structured opportunities to practice communication, decision-making, and problem-solving. The avatar leverages generative AI to produce nuanced responses while remaining anchored to validated knowledge through retrieval-augmented generation (RAG) pipelines. This ensures that learners receive accurate, relevant, and context-sensitive guidance without the risks of hallucinations or misinformation [6].

Mia's design reflects a broader pedagogical ambition: to bridge the gap between theoretical knowledge and professional practice. For instance, in healthcare, Mia can simulate challenging patient interactions, offering students the opportunity to practice empathy and conflict resolution. In legal training, the avatar can act as a virtual witness or client, allowing learners to rehearse questioning strategies. In corporate contexts, Mia can function as a feedback coach in leadership or negotiation scenarios, while in education, she serves as a facilitator for intercultural communication and collaborative problem-solving.

By blending AI-driven adaptability with embodied interaction, avatars like Mia extend the reach of expert mentoring. They enable deliberate practice—repetition with feedback—while maintaining learner engagement and reducing subjective bias. Far from replacing human mentors, they provide a scalable complement, ensuring that expert guidance remains available across time zones, languages, and institutional boundaries.

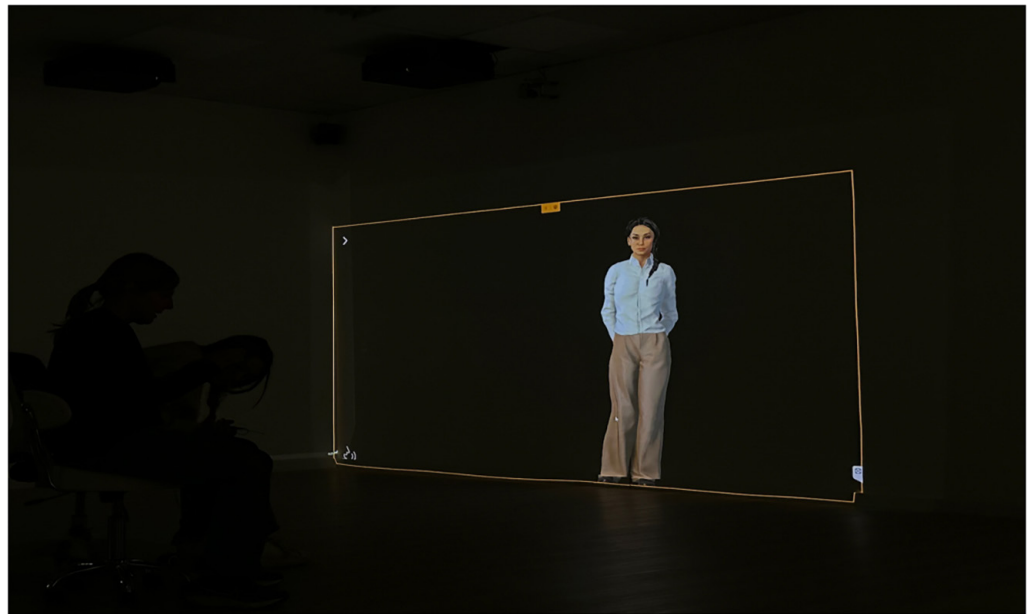


Fig. 1. Mia Williams: The AI-driven digital human, or avatar, designed and developed by the e-REAL Labs. In this case Mia is a 3D human-sized “persona”, projected on an interactive wall



Fig. 2. The same AI-driven avatar, Mia Williams, displayed on a personal computer’s screen

2 HUMAN-CENTRIC INTERACTION

For an avatar to act as an effective expert mentor, it must be more than a repository of information: it must engage learners through human-like interaction that respects the principles of cognition, perception, and social communication. Research in embodied cognition demonstrates that knowledge is not only processed intellectually but also constructed through bodily experience, gestures, and spatial orientation. Accordingly, avatars must be designed to leverage non-verbal communication cues—such as gaze, facial expression, intonation, and proxemics—to foster immersion and credibility [7–8].

The concept of social presence is central to this process. When learners perceive avatars as socially aware interlocutors, their sense of engagement and trust increases, leading to stronger motivation and retention of knowledge. Studies on immersive environments confirm that even subtle adjustments in eye contact, posture, and interpersonal distance can significantly influence how learners evaluate the credibility and authority of virtual agents [9–10]. Mia, for example, employs adaptive gaze and gesture synchronization, creating the impression of responsiveness and attentiveness, qualities associated with effective human mentoring.

Equally important is the role of empathy and affective computing in human–avatar interaction. By recognizing emotional signals from learners—such as frustration, hesitation, or enthusiasm—avatars can adapt their communication strategies, offering reassurance or scaffolding where needed. This contributes to what scholars describe as a relational dimension of learning, where emotional resonance is as critical as cognitive accuracy [11].

Moreover, interaction must be inclusive. Human-centric design requires accessibility features such as multilingual dialogue, simplified explanations, adaptive pacing, and compatibility with assistive technologies, ensuring that avatars do not reproduce social or cognitive barriers but instead broaden participation. This inclusivity reflects broader ethical frameworks in AI design, aligning with standards such as the EU AI Act and GDPR, which emphasize transparency, accountability, and fairness [12].

Human-centric interaction is not a peripheral aspect of avatar design but its core foundation. By combining embodied cognition, proxemics, and emotional intelligence, avatars like Mia foster environments where learners feel both guided and respected. This transforms avatars from mere digital assistants into credible expert mentors, capable of supporting skill development across diverse contexts and populations.

A representative video is available here: <https://youtu.be/MqIuiyyjMg>. In this case, Mia is performing as a healthcare expert and answering questions about cognitive biases related to the sickle cell disease, utilizing a knowledge base provided by the School of Medicine from Washington University (Missouri, USA). In addition to the healthcare application showcased in the video, Mia Williams has been successfully deployed in multiple domains. For example, she supports corporate training programs by coaching employees on decision-making and leadership skills; in academic environments, she guides students through complex subjects such as engineering and management; and in customer service simulations, she helps develop communication and conflict resolution abilities. These diverse use cases demonstrate Mia’s versatility and the broad applicability of her emotionally and socially intelligent interactions.

3 TECHNOLOGICAL FRAMEWORK

3.1 Mia’s AI-powered architecture: An in-depth exploration

Mia’s architecture is a sophisticated integration of advanced artificial intelligence, leveraging fine-tuned large language models (LLMs) such as GPT-4, Claude, or Llama within a scalable and configurable framework (patent pending). Several factors, including cost efficiency influence the choice of LLM, required performance metrics, and the deployment platform—whether it is cloud-based or local. Different models demonstrate varying performance levels in areas such as response quality, computational efficiency, and latency. However, the evaluation and benchmarking of different models are beyond the scope of this discussion.

3.2 Core functionalities of Mia's architecture

Several key components drive Mia's real-time adaptability and responsiveness:

- *Retrieval-Augmented Generation (RAG)*: By integrating RAG, Mia ensures precise, context-aware responses. This approach allows her to retrieve relevant information dynamically, significantly enhancing the depth and accuracy of her interactions.
- *Computer Vision and Audio Processing*: Utilizing microphone arrays and webcam-based tracking, Mia maintains eye contact, follows users' movements, and interprets body language. This capability enhances realism and engagement, making interactions more immersive and natural.
- *Emotionally Intelligent Interactions*: Mia incorporates emotional analysis and contextual awareness, enabling her to adjust tone, expressions, and feedback dynamically. This emotional intelligence fosters more profound and meaningful interactions, particularly in educational and customer service applications.

Multimedia Integration: To support complex explanations, Mia seamlessly presents videos, images, and interactive sketches. This capability is particularly valuable in educational environments, where visual aids can significantly improve comprehension. In the below picture, taken at the e-REAL Labs during a testing session, Mia is displaying a video on request, featuring Prof. Jenny Rudolph from the Harvard Center for Medical Simulation. In the video (available here: https://youtu.be/QfhrTIwN2_M), Prof. Rudolph shares her initial experiences using e-REAL avatars for coaching and training. It's important to highlight that she enthusiastically collaborates with us, embracing the potential of e-REAL technology to enhance learning and professional development.



Fig. 3. Mia Williams, as an advanced agentic application, can display images, videos, graphs, and infographics in real-time. Whether it's illustrating complex data, visualizing trends, or presenting multimedia content, Mia enhances interactive experiences by providing contextual visual support during conversations. This capability makes her an invaluable tool for training, education, and professional decision-making

3.3 System architecture overview

A reference architectural diagram (Figure 4) provides a visual representation of Mia's system, detailing the interactions between core components. Below is a breakdown of its primary modules and functionalities:

3.4 The orchestrator: The heart of the system

The orchestrator is the central and most critical module within Mia's architecture. Its primary responsibilities include:

- *State Management*: Maintaining the interaction flow, tracking conversational context, and synchronizing responses with animations.
- *User Input Handling*: Processing voice or textual input to determine the appropriate response strategy.
- *Decision-Making*: Determining whether to retrieve responses from the RAG system or invoke specific functions to complete tasks.
- *Integration with Unity™*: The orchestrator is implemented within the Unity framework, ensuring seamless synchronization between conversation states and avatar animations.

Mia's animated avatar responds dynamically based on the stage of the interaction. Animation states include:

- *Idle*: The avatar exhibits subtle, natural motions, enhancing its lifelike presence.
- *Attention*: Demonstrating active listening behavior when a user is speaking.
- *Explaining*: Using gestures to complement verbal explanations.
- *Responding*: Providing direct answers to user inquiries.

3.5 Task execution and external querying

Tasks within Mia's system are implemented as C# processes, which can be invoked by the orchestrator as needed. These tasks typically involve:

- Querying external resources, such as weather services, databases, or APIs, to fetch real-time information.
- Emotional analysis, which augments the avatar's social intelligence by analyzing the user's emotional state and adjusting responses accordingly.

3.6 Interaction modalities: Voice and text

Mia supports both chat-based and voice-based interfaces. While textual interactions are viable in various scenarios, this discussion primarily focuses on the voice interface, which provides a richer user experience. The voice interaction pipeline comprises:

- *Automatic Speech Recognition (ASR)*: A state-of-the-art ASR system transcribes spoken input with high accuracy.
- *Emotion Detection*: A dedicated system classifies the user's emotions into predefined states based on audio and visual cues.

- *Text-to-Speech (TTS)*: Mia's TTS engine generates highly natural and intelligible speech output. The voice can be customized based on parameters such as gender, age, and accent, ensuring adaptability to different user preferences and contexts.

3.7 Knowledge base and retrieval-augmented generation

Mia's knowledge base is composed of curated documents that serve as input for the RAG system. The process of retrieving and generating responses follows these steps:

1. *Document Chunking*: Input documents are segmented into smaller, manageable chunks and indexed in a vector database.
2. *Vector Search*: When a user asks a question, the system retrieves the most relevant chunks by computing vector distances.
3. *Rescoring and Ranking*: The retrieved chunks are ranked using a rescoring function to ensure optimal selection.
4. *LLM Processing*: The highest-ranking chunks are sent to an LLM, which synthesizes an accurate, contextually appropriate response.
5. *Fine-Tuning with LoRA*: To enhance response accuracy, the LLM is fine-tuned based on the curated documents, using Low-Rank Adaptation (LoRA), allowing efficient customization without retraining the entire model.

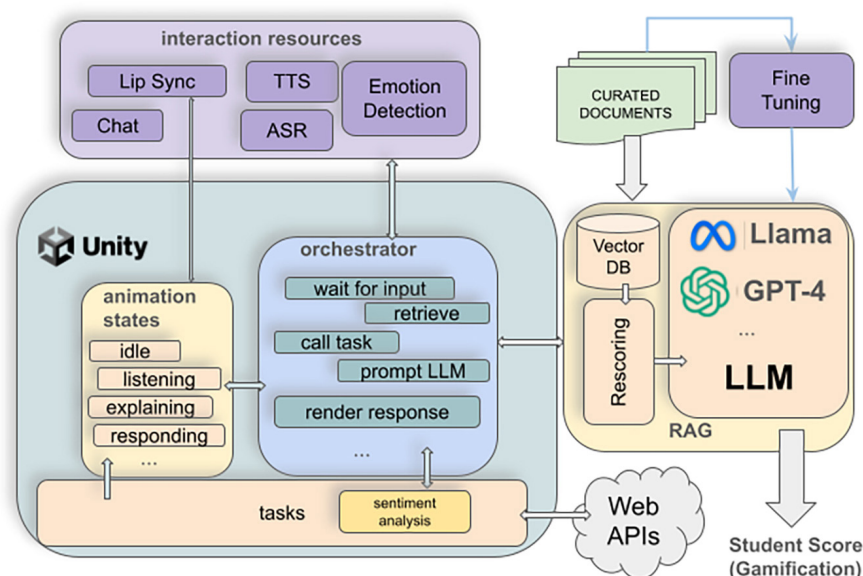


Fig. 4. Reference architecture of Mia

3.8 Educational applications: AI-powered adaptive feedback

Beyond conversational interactions, Mia plays a crucial role in educational environments by evaluating student interactions. The LLM assigns interaction scores based on criteria such as:

- **Engagement Level**: Tracking participation and responsiveness.
- **Adaptive Feedback**: Providing tailored explanations or additional resources to enhance learning outcomes.

In conclusion, Mia's architecture represents a cutting-edge fusion of AI, real-time interactivity, and advanced multimedia capabilities. By integrating RAG, emotional analysis, computer vision, and robust LLM-based processing, she delivers engaging and intelligent interactions tailored to various applications, from education to customer engagement. With ongoing advancements in AI and machine learning, Mia's capabilities will continue to evolve, setting new standards for virtual assistants and interactive AI systems.

4 FUTURE PROSPECTS AND INNOVATION

The evolution of AI-powered avatars such as Mia points toward a future in which expert mentoring and virtual coaching become increasingly accessible, adaptive, and embedded within diverse professional ecosystems. Several trajectories of innovation can already be identified, each linked to technological advances, pedagogical frameworks, and ethical imperatives.

First, ethical and regulatory frameworks will play a decisive role. As avatars become more sophisticated, ensuring compliance with data protection standards (GDPR, HIPAA, FERPA) and alignment with the EU AI Act will be crucial to maintaining trust. Transparency in how avatars generate responses, the sources they rely on, and the way user data is processed will become baseline expectations for learners and institutions.

Second, avatars will need to demonstrate cultural and linguistic adaptability. While Mia already operates across more than forty languages, future iterations will integrate localized cultural scripts, ensuring that mentoring resonates with diverse norms of politeness, authority, and communication style. This adaptability is especially relevant for healthcare and legal training, where misunderstanding cultural cues can have serious consequences.

Third, integration with emerging XR and metaverse environments is likely to expand. In these ecosystems, avatars will not only speak and gesture but also inhabit persistent shared spaces, interacting with learners, instructors, and other avatars. Early research indicates that such immersive multi-user settings enhance collaboration, social presence, and retention, but they also raise concerns about interoperability, equity of access, and cognitive overload. Designing calm and inclusive XR environments—where technology amplifies rather than distracts from human interaction—will be critical to success.

Fourth, the fusion of avatars with advanced analytics opens new frontiers for personalized learning. By tracking tone, pacing, and conversational strategies, avatars can provide learners with detailed performance dashboards, highlighting strengths and areas for improvement. Predictive analytics could identify at-risk learners and recommend targeted practice scenarios, aligning avatar-based mentoring with competency-based education models and institutional assessment frameworks.

Finally, scalability and sustainability remain central challenges. Avatars must function effectively across different platforms—mobile, desktop, immersive rooms, holographic installations—while minimizing infrastructure costs and environmental impact. Cloud-native modular architectures, combined with energy-efficient AI models, will enable institutions to expand access without compromising sustainability goals.

In summary, the future of AI-powered avatars lies in the balance between innovation and responsibility. As Mia and similar systems evolve, they will need

to combine pedagogical depth, technological robustness, cultural sensitivity, and ethical transparency, ensuring that virtual coaching not only scales expertise but also preserves the human values of mentoring and learning.

5 CONCLUSIONS

The development of intelligent avatars such as Mia illustrates how AI-powered virtual mentoring can transform knowledge transfer, professional training, and lifelong learning. By combining generative AI, retrieval-augmented knowledge systems, multimodal interaction, and human-centric design, avatars provide scalable, adaptive, and inclusive solutions for contexts ranging from healthcare and law to corporate leadership and higher education.

The evidence presented suggests that avatars are most effective when designed not as replacements for human mentors but as complements that amplify reach and availability. Human educators retain a central role in fostering reflection, ethical judgment, and creativity, while avatars ensure repeatable, unbiased, and accessible practice opportunities. This synergy reflects a new paradigm of hybrid mentorship, where expertise is distributed across human and artificial agents working in concert.

Future directions will require careful attention to ethical, cultural, and technological challenges. Ensuring transparency of AI decision-making, protecting learner privacy, and designing inclusive interactions will be critical to securing trust. Equally important will be the cultivation of cultural adaptability, allowing avatars to respond effectively to different communication norms, languages, and professional contexts.

As institutions adopt avatars at scale, the emphasis must shift from technological novelty to pedagogical value and long-term sustainability. This involves embedding avatars within curricular frameworks, aligning them with competency-based assessment, and designing them as part of inclusive learning ecosystems. Only then can avatars like Mia deliver on their promise of equitable access to expertise.

In conclusion, AI-driven avatars exemplify the convergence of cognitive science, immersive design, and responsible AI innovation. They represent not a distant vision but a rapidly maturing reality, one in which learners worldwide can engage with credible, empathetic, and adaptive expert mentors. By bridging the gap between knowledge and practice, avatars contribute to shaping a future where high-quality mentoring is no longer a privilege but a universal educational right.

6 REFERENCES

- [1] <https://www.e-real.net>
- [2] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse, "The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2003, pp. 529–536. <https://doi.org/10.1145/642611.642703>
- [3] S. Andrist, X. Z. Tan, M. Gleicher, and B. Mutlu, "Conversational gaze aversion for humanlike robots," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI '14)*, Association for Computing Machinery, New York, NY, USA, 2014, pp. 25–32. <https://doi.org/10.1145/2559636.2559666>

- [4] J. Bailenson, A. Beall, J. Loomis, J. Blascovich, and M. Turk, “Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments,” *Presence*, vol. 13, pp. 428–441, 2004. <https://doi.org/10.1162/1054746041944803>
- [5] C. Justine, S. Joseph, P. Scott, and C. Elizabeth, *Embodied Conversational Agents*. Cambridge, MA: The MIT Press, 2000. <https://doi.org/10.7551/mitpress/2697.001.0001>
- [6] B. Lugrin, C. Pelachaud, and D. Traum, Eds., *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, 1st ed. New York, NY: Association for Computing Machinery, 2022. <https://doi.org/10.1145/3563659>
- [7] W. Mattias and M. Virgolin, “Conversational agents: Theory and applications,” in *Handbook on Computer Learning and Intelligence: Volume 2: Deep Learning, Intelligent Control and Evolutionary Computation*, World Scientific, 2022, pp. 497–544.
- [8] S. Fernando, B. Barbara, G. Roxane, and M. Rebecca, “Effective extended reality: A mixed-reality simulation demonstration with intelligent avatars, digitized and holographic tools,” in *Innovative Approaches to Technology-Enhanced Learning for the Workplace and Higher Education: Proceedings of the Learning Ideas Conference 2022*, D. Guralnick, M. Auer, and A. Poce, Eds. Cham, Switzerland: Springer, 2023. https://doi.org/10.1007/978-3-031-21569-8_35
- [9] S. Fernando, B. Barbara, and C. Ianna, “Intelligent digital humans for bias-free recruitment interviews: A diversity & inclusion training program,” in *Creative Approaches to Technology-Enhanced Learning for the Workplace and Higher Education: Proceedings of The Learning Ideas Conference 2023*, D. Guralnick, M. Auer, and A. Poce, Eds. Cham, Switzerland: Springer, 2023. https://doi.org/10.1007/978-3-031-41637-8_36
- [10] S. Fernando and B. Barbara, *Learning 4.0. Advanced Simulation, Immersive Experiences and Artificial Intelligence, Flipped Classrooms, Mentoring and Coaching*. Milan, Italy: Angeli, 2018.
- [11] S. Fernando, B. Barbara, C. Ianna, G. Roxane, R. Jenny, and M. Rebecca, “Feedback reimaged: Generative AI and conversational avatars as your new training partners,” in *Creative Approaches to Technology-Enhanced Learning for the Workplace and Higher Education: Proceedings of The Learning Ideas Conference 2024*, D. Guralnick, M. Auer, and A. Poce, Eds. Cham, Switzerland: Springer, 2024. https://doi.org/10.1007/978-3-031-73427-4_13
- [12] K. Louisa, F. Dennis, and S. Annekathrin, “A comparison of the affectiva iMotions facial expression analysis software with EMG for identifying facial expressions of emotion,” *Front. Psychol.*, vol. 11, p. 329, 2020. <https://doi.org/10.3389/fpsyg.2020.00329>

7 AUTHORS

Fernando Salvetti is with the e-REAL Labs at Logosnet, 10014 Turin, Italy; 6900 Lugano, Switzerland; Houston, TX 77008, USA; New York, NY 10013, USA (E-mail: salvetti@logosnet.org).

Barbara Bertagni is with the e-REAL Labs at Logosnet, 10014 Turin, Italy; 6900 Lugano, Switzerland; Houston, TX 77008, USA; New York, NY 10013, USA.

Roberto Pieraccini is with the Scientific Advisor, New York, NY 10019, USA.