

How Embodiment Changes AI Tutoring: Comparing Student Perceptions of Text and AI Human Avatar-Based Chatbot Systems

Benedikt **Brünner** Michael Nestler Lukas Pucher Martin Ebner

2026-07-28

HCII 2026 | FLAAIT | TU Graz

ed-tech.at

Slides & License

These slides may be used under the terms of the CC license CC BY 4.0 International (<https://creativecommons.org/licenses/by/4.0/>).

Please use the links and references as sources for individual illustrations, if applicable.

Slides available at <http://doi.org/10.3217/xnhnv-kka76>



Benedikt Brünner, Michael Nestler,
Lukas Pucher, Martin Ebner
ed-tech.at



Outline

1 Motivation

2 Study

3 Results

4 Implications

Embodiment adds social cues, but not pedagogy by itself

- Text-based generative AI tutors are mainly valued for **explanations, clarification, and practice support**.
- Learner interaction is typically dominated by **information seeking**; goal setting and self-evaluation remain rare.
- A live human-like avatar adds **voice, facial expression, and gestures**.
- These cues may increase social presence, but also raise expectations about natural dialogue, empathy, and motivation.
- The open question is whether embodiment changes only the **experience** or also the **learning behaviour**.

Prior text-based deployments: (Brünner & Ebner, 2025; Brünner et al., 2025)

Research Questions

RQ1 – Perception

How do students perceive the communicative qualities and overall usefulness of an avatar-based RAG tutor compared with a text-based system?

RQ2 – Self-regulated learning

Does embodied, multimodal interaction foster different forms of self-regulated learning behaviour in chatbot usage data?

Outline

1 Motivation

2 Study

3 Results

4 Implications

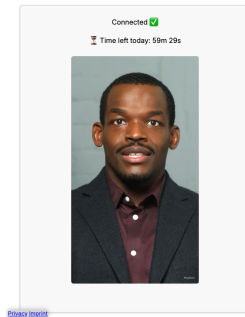
A field deployment connected perceptions with behaviour

- **Design-Based Research** deployment in two authentic higher-education contexts:
 - introductory computer science,
 - introductory Python programming.
- Voluntary use of an avatar-based RAG tutor.
- **Questionnaires:** users and non-users, with items comparable to the prior text-based deployment.
- **Interaction logs:** 341 anonymised user messages, coded with an adapted SRL process-action framework and manually reviewed.

SRL framework: (Lai, 2024)

The tutor combined RAG with a live human avatar

- Course-specific knowledge base with Retrieval-Augmented Generation.
- Real-time avatar with synchronised speech, facial expression, and gestures.
- Text or voice input with visible voice transcripts.
- Answers remained available as text; the full chat history could be reviewed and copied.



If desired, you can activate voice control for the communication with the avatar . To mute the avatar, you can press  at any time.

[15:56:29] Session started 

[15:56:29] Anthony: Hi, I am Anthony. How are you today and how can I help you?

[15:56:50] Me: What are the goals for the MOOC?

[15:56:53] Anthony: The goals of the MOOC "Informatik Fit" are to provide participants with essential knowledge and skills in computer science and programming. Specifically, it aims to:

1. Introduce fundamental concepts of computer science and programming.
2. Teach participants how to approach problem-solving using computational thinking.
3. Familiarize students with various programming languages, particularly Python.
4. Encourage hands-on practice through exercises and projects to solidify understanding.

If you have any specific aspects of the MOOC you're curious about, feel free to ask! 

New Session:  

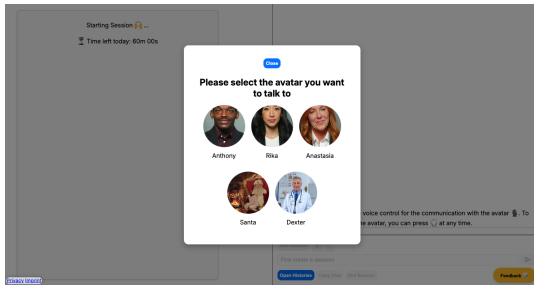
Enter Message

[Open History](#) [Copy Chat](#) [End Session](#)

[Feedback](#) 

Main tutoring interface: live avatar and parallel text chat.

Learners retained choice and transparency



Optional character selection before the session.

- Learners selected among predefined appearances and voices.
- GDPR-compliant consent was collected; no audio recordings were stored.
- Source code, knowledge base, and data-processing scripts were released openly.
- A daily session limit controlled computational and API costs.

Outline

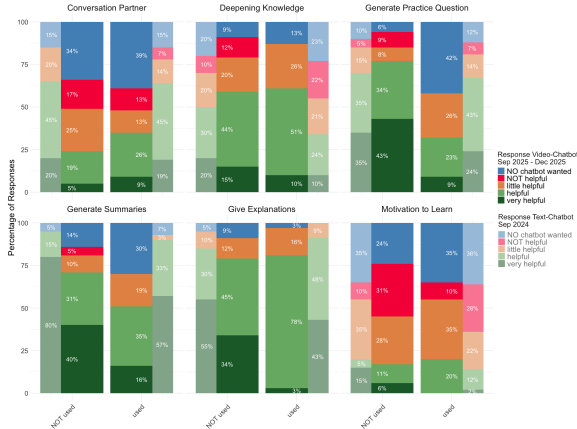
1 Motivation

2 Study

3 Results

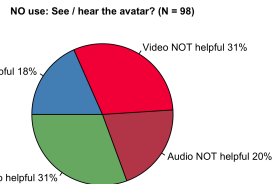
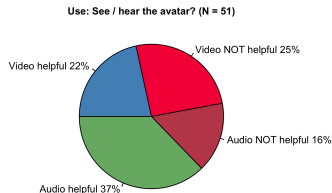
4 Implications

Learning-support ratings stayed close to text-only patterns



- Explanations and practice support remained the strongest perceived functions.
- Motivation support remained comparatively weak.
- The exception was the **communication-partner** dimension: active avatar users rated it more positively than non-users.

Audio mattered more than the visual avatar



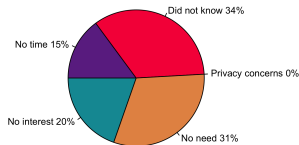
Users: perceived value of video and audio ($N = 51$).

Non-users: perceived value of video and audio ($N = 98$).

- Users rated audio more positively than video: **37% versus 22%**.
- Non-users showed the same pattern: **31% versus 18%**.
- Visual embodiment was evaluated more critically than speech.

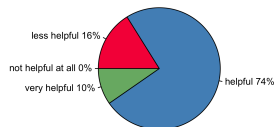
Adoption was positive; non-use was mostly situational

NO use: Why didn't you use the avatar? (N = 67)



Main reasons for non-use ($N = 67$, multiple selections).

Use: Was the avatar helpful overall? (N = 31)

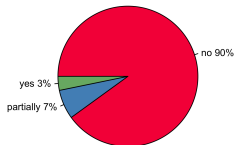


Overall helpfulness among active users ($N = 31$).

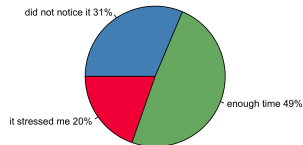
- Non-use reflected **lack of awareness** and **no perceived need**, not privacy concerns.
- **74%** of users rated the tutor helpful and **10%** very helpful.
- No user selected “not helpful at all”.

The prototype was stable, but the time cap added pressure

Use: Were there any technical problems? (N = 31)



Use: Was time-limit per day sufficient? (N = 31)

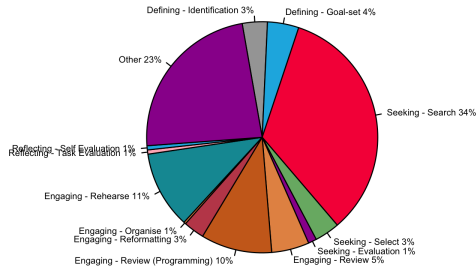


Reported technical problems among users ($N = 31$).

Perception of the daily time limit ($N = 31$).

- **90%** reported no technical problems.
- **49%** considered the daily limit sufficient and **31%** did not notice it.
- **20%** experienced the limit as stressful; **88%** supported offering the tutor again.

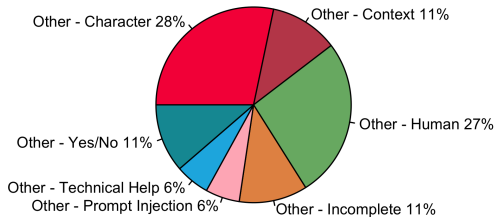
Embodiment did not shift learners toward deeper SRL



SRL action-process distribution in 341 user messages.

- **Seeking–Search: 34%** – still the dominant action.
- **Engaging–Rehearse: 11%.**
- **Programming review: 10%** – a distinct course-specific mode.
- Task evaluation and self-evaluation: only **1% each**.
- Embodiment alone did not trigger more goal setting or reflection.

Embodiment also attracted peripheral dialogue



Breakdown of the “Other” category ($N = 80$).

- “Other” interactions increased to **23%**, compared with **11%** in the text-only deployment.
- Character prompts (**28%**) and human-likeness checks (**27%**) made up most of this category.
- The remainder included incomplete input, prompt injection, and technical help: social presence created motives beyond learning.

Outline

1 Motivation

2 Study

3 Results

4 Implications

The two research questions point to different effects

■ **RQ1 – Perception:**

- Embodiment improved the tutor's perceived role as a communication partner.
- Core instructional usefulness stayed broadly comparable to the text-only deployment.
- Audio contributed more perceived value than the visual avatar.

■ **RQ2 – SRL behaviour:**

- Information seeking continued to dominate interaction.
- Metacognitive actions remained rare.
- Embodiment added peripheral dialogue rather than deeper regulation.

Future tutors need scaffolding, control, and role clarity

- **Prompt forethought:** ask learners to define a goal before requesting help.
- **Support monitoring:** use self-check questions and ask learners to explain or test their reasoning.
- **Trigger reflection:** end sessions with summaries, confidence checks, and next-step planning.
- **Separate programming support:** frame code review as iterative learning rather than answer generation.
- **Offer modality control:** let learners select text, audio, or avatar presentation.
- **Manage expectations:** communicate capabilities, limitations, privacy, and the tutor's pedagogical role during onboarding.

Limitations shape the next comparison

Limitations

- Voluntary participation and self-selection.
- Short deployment period.
- No direct learning-outcome measure.
- Contextual differences from the prior text-based study.
- Dependence on one avatar technology provider.

Next research steps

- Controlled text versus audio-only comparisons.
- Adaptive motivational and SRL scaffolds.
- Longitudinal studies across courses.
- Learning outcomes alongside perception and logs.
- Effects of modality control on trust, agency, and AI literacy.

Take-away: social presence is not self-regulation

- A live avatar can make an AI tutor feel more like a **communication partner**.
- It does not automatically change the tutor's dominant role as an **information source**.
- Visual embodiment can also redirect attention toward the character and away from learning.
- The next design step is therefore not “more human-like”, but **more pedagogically explicit**.

References I

- Brünner, B., & Ebner, M. (2025). **InfoFit and Beyond: AI Chatbots as EdTech Tools for Self-Regulated Learning in MOOCs.** In B. K. Smith & M. Borge (Eds.), *Learning and Collaboration Technologies* (pp. 45–60, Vol. 15807). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-93567-1_4
- Brünner, B., Ebner, M., & Schön, S. (2025). **Exploring genAI Chatbots in MOOCs: Analyzing Student Interactions and Self-Regulated Learning Behaviors.** In E. Hamonic & R. Sharrock (Eds.), *Digital education: Shaping sustainable lifelong learning for all in the era of ai. emoocs 2025* (pp. 14–24, Vol. 15733). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-00056-9_2
- Lai, J. W. (2024). **Adapting self-regulated learning in an age of generative artificial intelligence chatbots.** *Future Internet*, 16(6), 218. <https://doi.org/10.3390/fi16060218>

Questions & Discussion



- **Benedikt Brünner**
 - bruenner@tugraz.at
 - Graz University of Technology
 - <https://ed-tech.at>
-
- Co-authors: Michael Nestler, Lukas Pucher, and Martin Ebner.
 - Conducted within the FutureDEAL initiative and the doctoral programme “Bildungsinnovation braucht Bildungsforschung”.