

PRACTICAL CONTRIBUTION

AI Avatars in Learning Scenarios: Prompting and Characteristics

Judith Zellner[✉], Carina Galler[✉], Daniel Götz[✉], & Bernhard Ertl[✉]

Abstract

This practice contribution focuses on interactive learning with artificial intelligence (AI), specifically learning environments in which learners communicate with AI avatars through natural spoken-language discourse. These scenarios differ from earlier approaches, such as intelligent tutoring systems or pedagogical agents, through their high interactivity, natural language flow, and the capacity of current Large Language Models (LLMs) to support open communication about almost any content. This enables educators to configure and deploy diverse learning scenarios at different difficulty levels, including algebra tutoring, virtual patient simulations in medical education, peer-feedback role-plays in special education, and interactive teachers in STEM education. However, it remains unclear how such scenarios should be configured regarding interaction and appearance. Regarding interaction, the broad affordances of LLMs must be narrowed to specific learning situations with defined goals. Prompting must therefore constrain the LLM so that it provides authentic and motivating dialogue without deviating from the intended learning objectives. Strong responsiveness to learners may increase authenticity and motivation but risk losing instructional focus; strict goal adherence may make the scenario rigid and demotivating. Regarding appearance, the avatar's look, sex, and voice may shape perceptions of personality, credibility, and competence. This contribution first discusses the benefits and characteristics of avatar-based learning scenarios. It then introduces a peer-feedback role-play and an interactive teacher scenario, conceptualizes each educationally, and discusses their AI-avatar implementation, including prompt-specific peculiarities and pitfalls. Finally, it examines effects of avatar appearance on users' perceptions and derives practical conclusions.

Keywords: AI-avatar; instructional design; interactive teaching; role play; dialogic learning; prompting

1 INTRODUCTION

Learners benefit from individualized learning scenarios in contrast to traditional classroom lectures, and therefore several approaches emphasize the beneficial effects of individualized or adaptive learning environments (e.g., Chernikova et al., 2025; Leutner, 2002). A major obstacle for individualizing



This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2026 The Author(s). *Breaking Barriers to Learning and Education* published by Ludwigs-Maximilians-Universität München and Ludwigsburg University of Education in association with the German Hochschulkonferenz Sonderpädagogik des Lernens

learning lies in the personal resources required for individualization because most classroom settings just foresee one teacher for 20 or (many) more students. To overcome these challenges, research has developed several types of computerized individualization over the years, like programmed instruction (e.g., Glaser, 1965), intelligent tutoring systems (e.g., Graesser, 2001), or pedagogical agents (Lester et al., 1997). Such learning systems often proved to be beneficial (see Mousavinasab et al., 2021), but they did not change educational practices, nor did they receive widespread attention. There may be many reasons for this phenomenon, e.g., that the respective systems only focused on a very narrow subject area, that the interaction with these systems was awkward because learners had to communicate by entering written text with their keyboards, and that such pre-produced learning programs were hardly adaptable by teachers for their specific requirements.

Recent technological developments, however, now allow more generic as well as more powerful systems. By *Speech to Text* and *Text to Speech* modules (see Fink et al., 2024), they allow *free* communication (Ertl & Fink, 2025) in a way that learners can ask spoken questions and the system generates a spoken answer. Regarding the systems' knowledge base, *Large Language Models* (OpenAI et al., 2023) like ChatGPT can now support these interactions as they can generate answers from a comprehensive and quite generic knowledge base (see also Ertl & Fink, 2025). These options are now also reflected in the growing numbers of research literature and also meta-analyses on the effects of artificial intelligence (AI) and ChatGPT on learning (e.g., Fortuna et al., 2025; Younas et al., 2025). Furthermore, technology now allows such systems to visualize human-looking avatars as interaction partners (see Fink et al., 2024) that implement lip-sync speaking. Thus, technology facilitates nowadays generating AI-supported avatars as learning partners that communicate as if personal (see Ertl & Fink, 2025) and therefore may implement individualized and adaptive learning scenarios. Figure 1 exemplarily shows such a scenario.

In this context, we conceptualized and implemented several learning scenarios, like a tutoring scenario in algebra, a virtual patient simulation in medical education, a peer-feedback role-play in special education, or an interactive teacher in STEM education (see Ertl & Fink, 2025; Fink et al., 2024; Fink et al., 2025; Zellner et al., 2025a). While these scenarios technically run on quite similar platforms, e.g., with GPTAvatar (Robinson, 2023; avatar-research.com or avatar-lernen.de; Fink et al., 2024; Fink et al., 2025), their characteristics and the avatar's behavior must be defined by the respective prompting (see e.g., Fink et al., 2025). This prompt pre-structures the learning environment and introduces the specific learning scenario. It shapes the GPT's responses towards an authentic and motivating dialogue without losing track of the learning goals. We will exemplify our experiences regarding scenario creation and prompting in the next chapter by discussing two exemplary learning scenarios. Besides the behavior implemented in the prompt, also the visualization of the context and the appearance of the



FIGURE 1: Example of an AI supported Avatar learning scenario

avatar may affect learners' perception of its credibility: quite obviously, an avatar for a discussion with a pupil should look quite different from one for a discussion with a medical doctor or a teacher. Yet also minor differences may affect perception and credibility – an aspect that we will discuss two sections ahead.

2 INTERACTIVE AVATAR-BASED LEARNING SCENARIOS

In the following, we will discuss our experiences with prompting AI-based avatar-supported learning scenarios for a scenario of an individual role-play and an interactive teacher. For each of the two scenarios, we will first describe the scenario, the role of the avatar in the learning task, and characteristics of the prompting and reflect on our experiences with the respective implementations.

2.1 Individual role play

The individual role-play scenario addresses a core challenge of special education: enabling future teachers to make adaptive, data-based diagnostic and instructional decisions for learners with heterogeneous needs (Gebhardt, 2024). In this scenario, prospective teacher students engage in a one-to-one dialogue with an AI teacher avatar that simulates a diagnostic conversation about a student with reading difficulties. The interaction follows the logic of the diagnostic process step by step, including information gathering, interpretation of data, and the derivation of case-specific support measures. The pedagogical focus lies on translating diagnostic information into well-justified, feasible, and individualized instructional decisions.

This scenario is highly relevant because special education professionals are expected to support learners in overcoming diverse learning barriers within inclusive educational systems. Inclusive schooling requires adaptive instruction, individualized support systems, and professionally grounded, data-informed decisions (Fuchs & Fuchs, 2001; Wilcox et al., 2021). These competencies require repeated practice in authentic contexts. Although such practices are addressed across multiple university courses, traditional paper-based formats and group work often reach their limits when complex cases require active processing, iterative decision-making, and individualized reasoning.

Dialog-based interaction with an AI avatar opens new didactic possibilities for cognitive activation (see Zellner et al., 2025a; Zellner et al., 2025b). Unlike plenary or group discussions, in which only a few students tend to participate actively, the individual role-play ensures that every participant engages directly with the case. Each student conducts a full diagnostic conversation, explores alternative decision paths, and poses individual questions to the avatar. The avatar functions as a learning partner rather than an examiner, creating a protected learning environment in which mistakes are allowed and productive. This kind of individual role-play thus supports engagement with diagnostic reasoning, encourages experimentation with different solution strategies, and fosters reflective, practice-oriented learning in an ethically safe and motivating setting (Chernikova et al., 2025).

2.1.1 *Role of the avatar in the learning task*

The learning goal of the individual role-play scenario was to foster applied diagnostic knowledge through intensive cognitive activation in an individualized learning setting. By engaging in a one-to-one role-play with an AI avatar, participants were required to actively apply theoretical knowledge to a concrete case, justify their reasoning, and reflect on diagnostic and instructional decisions. The scenario was designed to promote deep processing rather than reproductive knowledge by requiring learners to articulate, revise, and substantiate their decisions throughout the interaction.

The learning scenario centered on a diagnostic conversation about a fictional student with reading difficulties (“Derek”). The AI avatar initiated the dialogue by stating, “I need your help with my student

Derek,” thereby positioning participants in the role of special education teachers. Together with the avatar, they worked through the diagnostic process in a structured manner. This process included the collection of general and case-specific information, the interpretation of available data-based results, and the derivation of individualized and feasible support measures. Throughout the conversation, participants were able to ask follow-up questions, request clarification, or inquire about the next appropriate diagnostic step. This design supported a systematic engagement with the diagnostic cycle while maintaining a realistic, problem-oriented conversational flow.

The role of the avatar was defined through content-specific prompts and structured question–answer patterns with conditional if–then linkages. The avatar primarily acted as a learning buddy who initiated reflection and provided information when requested. In one isolated instance, a brief role shift occurred at the end of a conversation when the avatar responded, “That’s great, thank you for your support. Now I know how to proceed,” followed by, “May I help you with anything else?” This momentary change did not affect the overall interaction quality. Across sessions, conversations were characterized by balanced turn-taking and reciprocal dialogue between the avatar and the participants.

2.1.2 Prompting

The avatar’s role was clearly defined from the outset. Participants were explicitly positioned as teachers, which implicitly framed the interaction in a pedagogically supportive and professional tone. This role attribution prompted responses characterized by collegiality, encouragement, and didactic clarity, such as acknowledging suggestions, asking reflective follow-up questions, and maintaining a constructive conversational style.

The avatar was controlled through a detailed text-based prompt that precisely specified its role, conversational behavior, and procedural structure. The prompt defined the avatar as Mr. Huber, a teacher at an inclusive elementary school, seeking advice on how to support a student named Derek with reading difficulties. The interaction followed a decision-tree structure aligned with the diagnostic process and divided into four numbered steps: case description, diagnostic information, data interpretation, and support measures. Each step included explicit linguistic instructions, for example: “Start with: ‘I’d like to tell you about a student named Derek who struggles in reading. I need your help to decide how to support him.’” The avatar was instructed to wait for the participant’s response before proceeding, to use short and clear sentences, and to refrain from offering solutions unless explicitly requested.

Conditional if-then rules structured the dialogue flow. For example, if participants requested a diagnostic test, the avatar provided predefined test results; if no further diagnostics were requested, the avatar proceeded to the next step. This rule-based logic ensured coherence and comparability across interactions while still allowing individual decision paths. To ensure subject-matter accuracy, all relevant background knowledge was embedded in the prompt, including standardized test results and developmental models for theory-based interpretation.

2.1.3 Reflection

Despite this shared knowledge base, participants differed substantially in depth of reasoning and use of technical terminology, indicating that the prompt supported differentiated engagement without constraining individual levels of expertise. Figure 2 shows examples for wording uttered by a student with thorough and data-based reasoning and use of appropriate terminology (Figure 2a) and a student with rather superficial reasoning and less appropriate terminology use (Figure 2b).

What worked well. The avatar created a high level of authenticity and engagement by simulating realistic professional dialogue. Its adaptive responses and consistent role behavior supported learners in applying diagnostic reasoning processes. Students appreciated the objectivity of the interaction, the

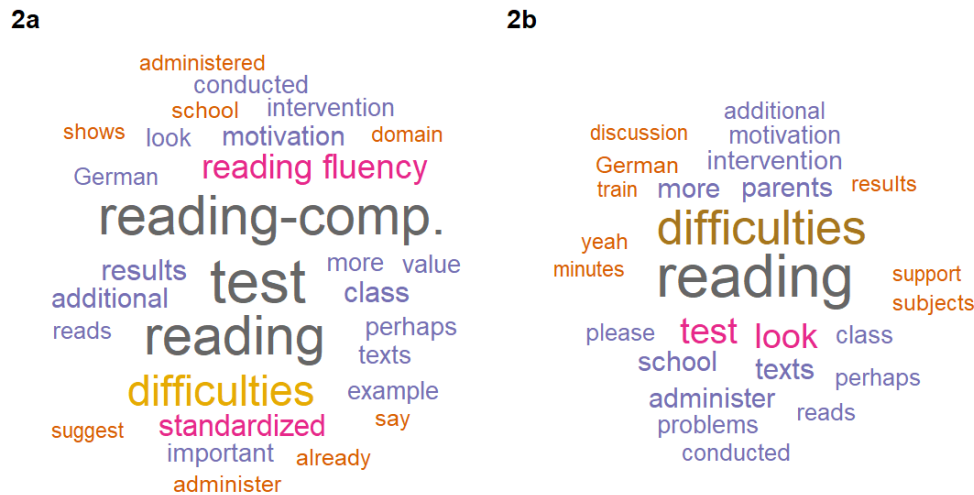


FIGURE 2: Word cloud of a student using appropriate terminology (2a) and a student without (2b).

Note. Terms were translated for presentation purposes while preserving original frequencies. Reading comprehension abbreviated as reading-comp. for display purposes.

possibility to work independently, and the standardized yet flexible structure of the conversation. Technical access was intuitive, and the format proved scalable for multiple simultaneous users.

What was challenging. Some participants, especially novices, experienced high cognitive load due to the open, spontaneous dialogue format. Without prior practice, the avatar conversation led to a higher dropout rate and occasional misunderstandings. Technical challenges such as microphone or browser issues disrupted flow in some cases. Moreover, overly general or rigid prompts sometimes caused the avatar to deviate from its role or to provide premature solutions. The pedagogical depth of the responses has not yet been systematically analyzed qualitatively, indicating the need for further in-depth evaluation in future studies.

Detailed descriptions of the respective implementation contexts and avatar design principles of the individual role-play scenario are provided in complementary publications (see Zellner et al., 2025a; Zellner et al., 2025b; Zellner, 2026).

2.2 Interactive Teacher

The interactive teacher scenario focused on concepts and processes that are relevant for wastewater treatment. The setting extends an immersive virtual field trip scenario (IVFT; Fink et al., 2026) that was designed for 10th graders towards an interactive learning setting. Students found themselves in a 360° environment of a wastewater treatment plant (see Figure 3). During the learning process, students visited different stations of the plant by following the flow of the water. At each station, students could freely rotate their view in the 360° sphere and inspect the sphere as they would have been able when physically standing at this point in the real plant. They furthermore learned about the core concepts and processes at the respective stations that, however, relate to different science domains like physics, chemistry, and biology. Instead of just receiving pre-recorded explanations at each station of the IVFT, students can now interact with the interactive teacher, ask questions, and receive special support. Because we assumed that students had heterogeneous and partially quite low prior knowledge for at least some of the stations. This means that the learning environment had to adapt to the learners' prior knowledge. This adaptation may take place through students asking open questions – or via a dedicated environment that provides prompting for special educational needs.

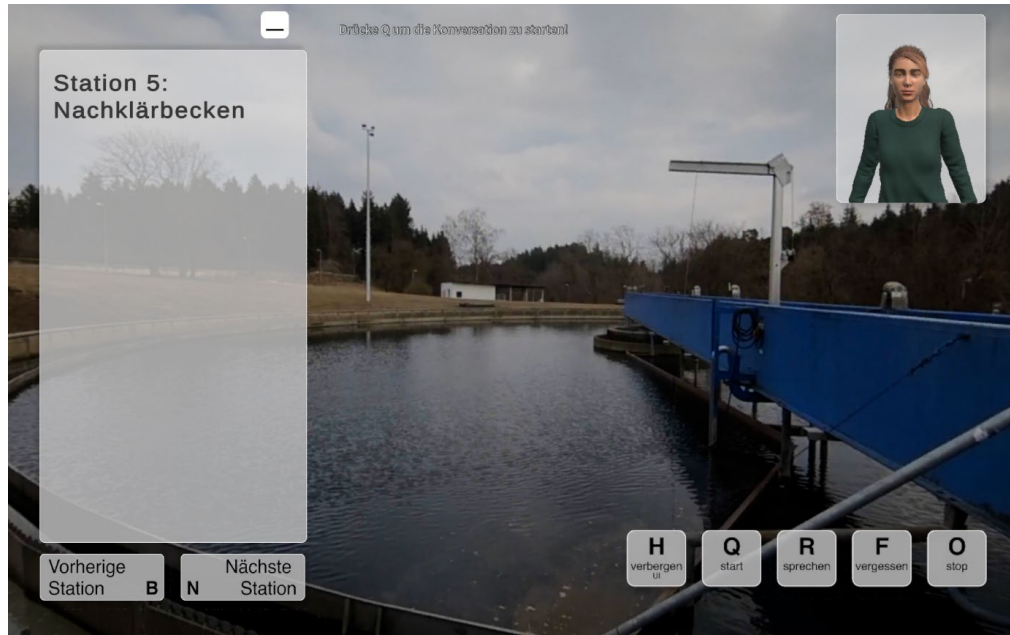


FIGURE 3: Screenshot of the learning environment at the secondary clarifier of the wastewater treatment plant.

The didactic rationale for these adaptive processes is to support a learner developing its understanding about a specific content area collaboratively with a teacher avatar. Learners worked interactively with the teacher on the core elements (concepts and processes) at the respective station. This design is grounded in the *ICAP framework*, which conceptualizes learning as a function of observable cognitive engagement (Chi & Wylie, 2014). Consequently, our teacher was designed to support learners' interactive construction of knowledge about core elements of the stations: instead of explaining the elements, the interactive teacher provided background information about them to the student and then encouraged the learner to conclude reasons or characteristics for the respective elements. In the case that the learner was able to elaborate the requested elements appropriately, the interactive teacher continued with the next elements, while in the other case, the teacher gave scaffolds to support the student's thinking over the respective elements. The interactive teacher was therefore designed to elicit students' reasoning rather than provide explanations.

2.2.1 Role of the avatar in the learning task

To meet these goals, the interactive teacher (called Alex) was designed to support and maintain a learner's thinking process without taking over. Direct explanations were reduced to a minimum. Instead, Alex externalized important background information for each station and elicited learners' conclusions. This interactive construction process was scaffolded by Alex's stepwise support and gradual hints. For the sedimentation processes, e.g., Alex may give an example of a glass bottle filled with sand and water and ask the learner what would happen if someone shook the bottle heavily and then put it on a table. Based on that, Alex elicits the learner's thinking that the sand may sink down when the water gets quiet. In a follow-up question, Alex may further elicit the learner's reasoning that a wastewater treatment plant would need an area with very slow water movement to allow sand and other heavy particles to sink down and sediment.

Furthermore, Alex directs learners' attention toward salient visual features of the 360° scenes and encourages them to connect observations to underlying principles. In the aeration tank, for example, Alex directs attention towards the bubbles plumes on the water and gives a hint by explaining that they consist of air and oxygen.

2.2.2 Prompting

Alex relies on station-specific content anchors, about three for each station, that are embedded in the system prompt and that define the learning goals for a station (e.g., “*Your questions should prompt the user to actively work out which personal behaviors can reduce unnecessary load on the grit and grease trap*”). Alex provides background information on these learning goals but invites the learner to provide self-generated elaborations to meet the learning goals. Alex uses open questions to elicit explanations and asks targeted follow-up questions that require learners to connect scene cues with underlying mechanisms (e.g., “*We are standing at the aeration tank. Describe the activity in the tank in an objective manner (air injection, mixing). Mention that this is the main biological treatment step*”). Learner errors serve as productive entry points: rather than correcting directly, Alex offers short directional hints that redirect attention to relevant perceptual features or previously articulated statements. Thus, in case the learner’s elaborations do not meet the learning goal yet, the prompt requires Alex to summarize which aspects of the student’s reasoning go towards the right direction and gives further hints (e.g., “*If the answer is wrong, do not give the solution – give a hint that supports the learner’s thinking*”).

The prompt of Alex paid attention that learners sequentially developed understanding about the concepts defined by the content anchors without jumping between the stations and that Alex didn’t disclose a solution until learners had articulated a plausible reasoning chain. Overall, the prompt design had to balance flexibility (e.g., for following the learners’ reasoning) and required constraints (meeting the learning goals).

2.2.3 Reflection

It is essential to realize that there is some kind of goal conflict in such kind of interactive teacher scenario: the interactivity of knowledge construction and adaptivity would benefit if the avatar follows the learner’s reasoning towards an individual learning path. This path, however, may not necessarily reflect the flow of the water and the progress in the learning environment to meet the learning goals. In contrast, each station has some specific learning goals that should be met during the learning process. Thus, the prompt requires Alex to be responsive to the learner’s reasoning without losing track of the learning goals.

What worked well. Observations and questionnaire responses indicate that the scenario elicited interactions in the intended style for most of the learners. In this line, several learners reported that station-wise sequencing and the requirement to verbalize reasoning supported sustained engagement and helped them build explanations. Some learners reported spending longer at individual stations, which may reflect high interest and motivation, although also a novelty effect may be plausible.

What was challenging. A general challenge was the perceived trade-off between depth of elaboration and efficiency, as already mentioned in the context of the goal conflict. Prompting the avatar to ask for the learner’s articulation and to respond with contingent hints can slow down progression in the learning environment. Such kind of guidance may be experienced as coercive, especially if a learner has for some reason no clue about the intentions of the avatar. Giving more freedom to the discourse, however, resulted in some discussions losing track of the learning goals. Furthermore, we were not yet far enough to establish a sound coherence between the 360° sphere and the learning discourse. It would have been desirable that the avatar would have guided the learner’s visual exploration of the 360° sphere, which the current version was not yet able to fulfill. This, however, is not trivial, as such an endeavor would require that the learner and avatar—that is, the AI—share the same view in the sphere, which also would mean providing a visual layer, in addition to the verbal layer, for the AI interaction.

Detailed descriptions of the respective setting and avatar design principles of the interactive teacher scenario are provided in complementary publications (see Ertl & Fink, 2025; Fink et al., 2026).

3 EFFECTS OF AVATAR APPEARANCE

While constructing the learning scenarios and the prompting, we also got the impression that we should check learners' perceptions of the avatars' appearance. This impression was raised by peer feedback that some of the available avatars looked less sympathetic, yet also by research on digital avatars in human–computer interaction that reports users applying familiar social rules to virtual characters, even when they know that no real person is present. Like in our scenario, avatars are treated as social actors, and visible gender cues are among the first categories people use when forming impressions of them (Nag & Yalçın, 2020; Nowak & Rauh, 2008). In virtual environments, visible appearance cues such as gendered features and clothing shape social perception, with variations in avatar gender expression and attire influencing stereotype-consistent attributions related to social roles, professionalism, and anthropomorphism (DeWester et al., 2009; Nowak & Rauh, 2005; Nowak et al., 2009).

Beyond social categorization processes, perceptual and design-related features such as facial structure, level of anthropomorphism, and lighting shape how realistic and credible virtual humans are perceived, particularly at high levels of human likeness (Nowak & Rauh, 2008; Shevchenko et al., 2024; Schwind et al., 2018; Schwind et al., 2015). Experimental work on facial design shows that even subtle variations in facial features influence perceptions of attractiveness, likeability, and eeriness, and that designs deviating more strongly from human norms are more likely to evoke uncanny impressions (Schwind et al., 2015). Findings from video-mediated communication further indicate that camera framing and viewing distance affect social presence and perceived proximity (Fauville et al., 2022), mechanisms that are also relevant for avatar-based interactions. In applied contexts, visual design cues that signal professional appropriateness influence credibility judgments of virtual agents, particularly in educational settings (Ashrafi et al., 2025). Correspondingly, more credible and natural-looking teacher avatars are associated with higher perceived teaching effectiveness and greater willingness to engage (Garcia, 2025).

Against this background, we created four humanlike avatars, two male and two female, and wanted to know how students interpret their personality, credibility, and occupational fit after very brief exposures. First, we just compared a female and a male avatar (original female/male), but when we realized differences between the perception of both, we systematically refined their avatars' appearance (refined female/male). In a second run, we wanted to find out to what extent subtle design changes can stabilize social impressions and reduce unintended bias in AI-based learning scenarios.

3.1.1 Implementation

The avatars consisted of two male and two female characters created with Character Creator 4 v.4.62 (Reallusion Inc., 2025), shown in Figure 4, each designed to resemble a plausible professional in an educational context. Students viewed a short introduction video for each avatar, approximately ten seconds in length and presented three times to support stable first impressions. The presentation order alternated across groups to minimize sequence effects.

After each viewing, participants completed a brief questionnaire. They indicated the profession they believed the avatar represented, provided three spontaneous character descriptors, rated credibility on a ten-point scale, and reported their willingness to interact socially with the avatar. They also evaluated each avatar on a single item for each of the Big Five personality dimensions, allowing the capture of rapid, intuitive judgements.



FIGURE 4: Avatars created with Character Creator 4 (Reallusion Inc., 2025)

Following the original comparison ($n = 43$), the avatars were refined to improve visual coherence and reduce unintended variation. Adjustments included harmonizing eye color and clothing, modifying selected facial features such as cheek fullness, and revising the male avatar's hairstyle for a more natural appearance. The camera framing was standardized, as the initial recordings had positioned the female avatar farther from the viewer than the male avatar. Lighting was also warmed to reduce the artificial appearance typical of unrefined digital renders. Two further student groups ($n = 32$) compared the refined avatars using the same procedure, allowing direct comparison of impressions before and after the visual modifications.

3.1.2 Results

In the comparison of the original avatars, several statistically reliable gender-related differences emerged. The female avatar received higher credibility ratings and was also rated as more agreeable and more open. The male avatar was rated as more extraverted. Conscientiousness and neuroticism showed no meaningful differences. Social willingness produced a medium-sized difference favoring the male avatar.

Participants provided three free descriptors for each avatar. The female avatar was most often described with terms such as “friendly”, “supportive”, “patient” and “trustworthy”. Descriptions of the male avatar were broader and more polarized, frequently including “confident”, “assertive” and “outgoing”, alongside occasional terms such as “distant” or “stern”. Occupational attributions differed significantly. The female avatar was predominantly assigned to teaching, counseling, or health-related professions, whereas the male avatar was more often linked to technical, administrative, or policing roles.

The two refined avatars received largely comparable ratings across all measured traits. Credibility was high for both, and no statistically significant differences were found. Ratings of agreeableness, openness, extraversion, conscientiousness, and neuroticism varied only slightly, and none of the differences reached significance. Social willingness showed a small, non-significant tendency favoring the female avatar.

Free character descriptions were more homogeneous. Both avatars were frequently described as “professional”, “competent”, “approachable” and “calm”. Strongly gendered adjectives appeared less often, and descriptors indicating dominance or marked warmth were rare. Job category assignments were also more evenly distributed. Both avatars were most often placed in educational or academic professions, and no significant association between avatar gender and occupational category was observed.

3.1.3 Reflection

The results show that the two avatar versions prompted different patterns of perception, indicating that design features meaningfully shaped how consistently the avatars were estimated. In the initial version, participants differentiated more strongly between avatars, and the written character descriptions reflected a broad set of impressions, often drawing on familiar social categories such as warmth, confidence, or assertiveness. In the revised version, ratings were more aligned, and the descriptors were more uniform and professionally neutral, suggesting that specific visual adjustments reduced cues that had previously encouraged more divergent judgements.

Although the differences between avatars decreased substantially in the refined version, small numerical variations remained across several dimensions. These residual differences may indicate that some aspects of impression formation are less responsive to design refinement or that not all features exert comparable effects. The shift in job category assignments follows a similar pattern. Whereas the first version produced clearly different occupational associations, the revised version yielded more balanced distributions, though the data does not clarify which design elements contributed most to this change.

A further limitation concerns the absence of demographic information about the participants. The study did not record the gender of respondents, even though prior work suggests that perceiver characteristics can influence responses to virtual avatars. Collecting this information in future studies would enable a more precise assessment of whether evaluation patterns reflect consistent group-level tendencies or interactions between avatar design and user characteristics.

4 CONCLUSION

In this practice paper we illustrated how AI-based avatars can support different learning scenarios. We gave examples of two learning scenarios, the respective roles of the avatars, the respective prompting, and our experiences. We also discussed that the appearance of the avatar is a communicator of credibility and other perceptions, and thereby the avatar’s appearance should at least be coherent with the learning goals.

Our experiences indicate that AI-based avatars can support individualized and adaptive learning scenarios, but not by default. Rather, individualization emerges only under specific technical, didactic, and design-related conditions. For example, it is currently not possible to allow participants to personalize the visual representation of the avatar, as such adaptations would comprise standardized data collection and introduce additional sources of variance. However, creating a personalized teacher avatar might also compromise its credibility; as previously discussed in the section about avatar appearance, diverging from a standardized design can evoke unintended social stereotypes that interfere with a neutral and professional educational appearance. Future research should examine whether allowing learners

to personalize the visual representation of the avatar influences perceived expertise, social presence, and learning outcomes, while carefully balancing personalization with the need for standardized assessment conditions. This finding directly relates to the problem outlined in the introduction, namely the tension between the pedagogical demand for individualized learning and the limited human resources in educational practice.

The interactive setting with the avatar has, however, some limitations regarding *technology and speech* and *network and latency*. The *technology and speech* limitation requires that learners speak comprehensibly/convertibly. While human interaction partners usually comprise a set of grounding strategies (see Clark & Brennan, 1991) that allow them to prevent and clarify misunderstanding, such strategies are currently missing in *speech-to-text systems*, and therefore naturally speaking AI avatars are dependent on learners' clear pronunciation as well as on a technical infrastructure/microphone quality that allows clear voice recording. These systems may have to be even better than traditional videoconferencing infrastructure in which human users are able to apply their grounding strategies. Thus, it is essential for successful implementations that the respective hardware as well as user interaction allow a clear and unambiguous speech-to-text translation. *Network and latency* relate to the network infrastructure. Many of these tools are provided as web applets that must be started (and downloaded) before each session to allow flexible use from almost any system. This, however, implies a notable size of the web applet that may comprise 200 to 500 MBs. Furthermore, while learners communicate in real time, one must consider that the communication setting requires several sub-processes like (a) the speech to text conversion, (b) the GPT inquiry and response, and (c) the text to speech conversion with (d) a server orchestrating this all. This may require up to 4 internet transfers of audio and up to 4 internet transfers of text data, requiring sound internet connections as well as the times and latencies for the respective (up to 6) conversions. Recent GPT models are more powerful but may also require more response time that may, depending on the questions and the reasoning level, take more than 10 seconds (d4b, 2025). These times sum up and may reduce real-time interactivity by large latencies. Thus, it may be promising to choose a less powerful but quicker GPT model.

The GPT models and the respective prompting, however, are a key challenge and a key vulnerability of these learning scenarios. Like different GPT models may show different response latencies, they also may interpret the prompts in the learning environment quite differently and thereby result in quite different avatar behavior. Therefore, it is essential to ensure coherence between the prompt and the GPT model.

A further aspect that we mentioned rather implicitly is the modeling of the didactic processes in the prompt. This modeling can be quite explicit, like in the role-play example, and include decision trees and if-then clauses to resolve different content-specific responses and interactions – or rather vague by defining abstract, “core” interaction styles and core contents by fixed rules like in the interactive teacher. Both have advantages and challenges. The explicit modeling needs a clear content-specific framework about which kind of answers or hints are given best in a specific situation. This procedure implies that designers need a clear vision of which content to teach how – but it will reach some kind of standardization at the cost of putting much effort into the design of the content framework and the prompt. The vague modeling means to define communication processes and learning goals but without modeling how to reach the learning goals on a content-specific layer. Such modeling appears to be quite easy to implement, and it allows the avatar to be quite adaptive to the learner – but at the cost of giving the AI much flexibility about ways and examples to choose and how to reach the learning goals. While this worked quite well in most cases, we also experienced occurrences of inappropriate or silly examples or circular discussions. Thus, it seems promising to model concept-specific misconceptions and how to deal with them. A final remark relates to prompting the AI for the spoken discourse, which means suppressing all kinds of output that can hardly be transformed by text-to-speech systems, like emojis, text

formatting, bullet point lists, sentences in brackets, etc., as well as meta-commentaries about prompts or software that would make the avatar leave its role.

In summary, AI-based avatars offer substantial potential for implementing individualized and adaptive learning scenarios at scale, but only when technical reliability, prompt coherence, and didactic modeling are systematically aligned. The central contribution of this practice paper thus lies not in demonstrating what AI-based avatars can do, but in specifying the conditions under which they can meaningfully support instructional goals. Designing effective avatar-based learning environments, therefore, requires considerable pedagogical and conceptual effort, positioning AI-based avatars not as autonomous solutions but as didactically programmable interaction partners.

FUNDING

This paper is partially funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr [project RISK.twin]. dtec.bw is funded by the European Union – NextGenerationEU.

REFERENCES

- Ashrafi, N., Graf, P., Marquardt, M., Vona, F., Schorlemmer, J., & Voigt-Antons, J.-N. (2025). Avatar appearance beyond pixels: User ratings and avatar preferences within health applications. In V. G. Duffy (Ed.), *Digital human modeling and applications in health, safety, ergonomics and risk management* (Lecture Notes in Computer Science, Vol. 15793). Springer. https://doi.org/10.1007/978-3-031-93508-4_11
- Chernikova, O., Sommerhoff, D., Stadler, M., Holzberger, D., Nickl, M., Seidel, T., Kasneci, E., Küchemann, S., Kuhn, J., Fischer, F., & Heitzmann, N. (2025). Personalization through adaptivity or adaptability? A meta-analysis on simulation-based learning in higher education [Review]. *Educational Research Review*, 46, Article 100662. <https://doi.org/10.1016/j.edurev.2024.100662>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219-243. <https://doi.org/10.1080/00461520.2014.965823>
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick (Ed.), *Perspectives on socially shared cognition* (pp. 127-149). American Psychological Association. <https://doi.org/10.1037/10096-006>
- d4b. (2025, September 28). *Average response time metrics for OpenAI GPT-5 models (by reasoning level)*. <https://www.d4b.dev/blog/2025-09-28-gpt5-response-time-metrics>
- DeWester, D., Nah, F. F.-H., Gervais, S., & Siau, K. (2009). An examination of perceptions of male and female avatars. In *MG 2009 Proceedings* (Paper 5). <https://aisel.aisnet.org/mg2009/5>
- Ertl, B., & Fink, M. C. (2025). Alter Wein in neuen Schläuchen? KI-basierte Lernsysteme und Lernavatare. In B. Schäffer & F. R. Lieder (Eds.), *Maschinen wie wir? Wie Künstliche Intelligenz Bildung, Wirtschaft und Gesellschaft herausfordert* (pp. 143-166). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-48522-1_7
- Fauville, G., Queiroz, A. C. M., Luo, M., Hancock, J., & Bailenson, J. N. (2022). Impression formation from video conference screenshots: The role of gaze, camera distance, and angle. *Technology, Mind, and Behavior*, 3(1), 133–143. <https://doi.org/10.1037/tmb0000055>
- Fink, M. C., Galler, C., Watzka, B., & Ertl, B. (2026). Effects of annotations and quizzes in a classroom-based immersive virtual field trip. *Computers & Education: X Reality*, 8, 100140. <https://doi.org/https://doi.org/10.1016/j.cexr.2026.100140>
- Fink, M. C., Robinson, S. A., & Ertl, B. (2024). AI-based avatars are changing the way we learn and teach: benefits and challenges [Perspective]. *Frontiers in Education*, 9. <https://doi.org/10.3389/educ.2024.1416307>
- Fink, M. C., Walter, L., Eska, B., & Ertl, B. (2025). A Web Application and Authoring Tool for Supporting the Generation of Ai-Based Avatars for Communication Scenarios. In *Two Decades of TEL. From Lessons Learnt to Challenges Ahead*. Cham.

- Fortuna, A., Prasetya, F., Samala, A. D., Rawas, S., Criollo-C, S., Kaya, D., Raihan, M., Andriani, W., Safitri, D., & Nabawi, R. A. (2025). Artificial intelligence in personalized learning: A global systematic review of current advancements and shaping future opportunities [Review]. *Social Sciences and Humanities Open*, 12, Article 102114. <https://doi.org/10.1016/j.ssaho.2025.102114>
- Fuchs, L. S., & Fuchs, D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research & Practice*, 16(3), 174-181. <https://doi.org/10.1111/0938-8982.00018>
- Garcia, M. B. (2025). Teachers in the metaverse: The influence of avatar appearance and behavioral realism on perceptions of instructor credibility and teaching effectiveness. *Interactive Learning Environments*, 33(7), 4334-4350. <https://doi.org/10.1080/10494820.2025.2462144>
- Gebhardt, M. (2024). *Pädagogische Diagnostik. Leistung, Kompetenz und Entwicklung messen, bewerten und für individuelle Förderung interpretieren* (Version 0.3). Ludwig-Maximilians-Universität München. https://epub.uni-muenchen.de/110013/1/Diagnostik_in_der_Schule_Version_0.3.pdf
- Glaser, R. (Ed.). (1965). *Teaching machines and programmed learning II: data and directions*. Association for Educational Communications and Technology.
- Graesser, A. C., VanLehn, K., Rose, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent Tutoring Systems with Conversational Dialogue. *AI Magazine*, 22(4), 39. <https://doi.org/10.1609/aimag.v22i4.1591>
- Leutner, D. (2002). Adaptivität und Adaptierbarkeit multimedialer Lehr- und Informationssysteme. In L. J. Issing & P. Klimsa (Eds.), *Information und Lernen mit Multimedia und Internet: Lehrbuch für Studium und Praxis* (3., vollst. überarb. Aufl., pp. 115-126). Beltz PVU.
- Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*, Atlanta, GA, USA, 359-366. <https://doi.org/10.1145/258549.258797>
- Mousavinasab, E., Zarifasanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods [Review]. *Interactive Learning Environments*, 29(1), 142-163. <https://doi.org/10.1080/10494820.2018.1558257>
- Nag, P., & Yalçın, Ö. N. (2020). Gender stereotypes in virtual agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents* (Article 41, pp. 1-8). Association for Computing Machinery. <https://doi.org/10.1145/3383652.3423876>
- Nowak, K. L., Fox, J., & Ranjit, Y. S. (2015). Inferences about avatars: Sexism, appropriateness, anthropomorphism, and the objectification of female virtual representations. *Journal of Computer-Mediated Communication*, 20(5), 554-569. <https://doi.org/10.1111/jcc4.12130>
- Nowak, K. L., & Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, 11(1), 153-178. <https://doi.org/10.1111/j.1083-6101.2006.tb00308.x>
- OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., & et al. (2023). GPT-4 technical report. *arXiv [Preprint]*. <https://doi.org/10.48550/arXiv.2303.08774>
- Reallusion Inc. (2025). *Character Creator 4* (Version 4.62.4603.1) [Computer software]. Reallusion. <https://www.reallusion.com/character-creator/>
- Robinson, S. A. (2023). *GPTAvatar* [Computer software]. GitHub. <https://github.com/SethRobinson/GPTAvatar>
- Schwind, V., Wolf, K., Henze, N., & Korn, O. (2015). Determining the characteristics of preferred virtual faces using an avatar generator. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (pp. 221-230). Association for Computing Machinery. <https://doi.org/10.1145/2793107.2793116>
- Schwind, V., Wolf, K., & Henze, N. (2018). Avoiding the uncanny valley in virtual character design. *Interactions*, 25(5), 45-49. <https://doi.org/10.1145/3236673>
- Shevchenko, Y., Heisenstein, J., & Reips, U.-D. (2024). How do avatars make a positive impression: The effect of facial cues on avatar evaluation. *Journal of Perceptual Imaging*, 7, 1-16. <https://doi.org/10.2352/J.Percept.Imaging.2024.7.000406>
- Wilcox, G., Fernandez Conde, C., & Kowbel, A. (2021). Using evidence-based practice and data-based decision making in inclusive education. *Education Sciences*, 11(3), 129. <https://doi.org/10.3390/educsci11030129>

- Younas, M., Ismayil, I., El-Dakhs, D. A. S., & Anwar, B. (2025). Exploring the Impact of Artificial Intelligence in Advancing Smart Learning in Education: A Meta-Analysis with Statistical Evidence [Article]. *Open Praxis*, 17(3), 594-610. <https://doi.org/10.55982/openpraxis.17.3.842>
- Zellner, J. (2026, February 27). Project B - AI-Avatars in Teacher Education. [OSF project]. <https://doi.org/10.17605/OSF.IO/YMZT8>
- Zellner, J., Koch, J., Fink, M., Ebenbeck, N., & Gebhardt, M. (2025a). Fostering Diagnostic Decision-Making in Special Education Students through Simulation-Based Games and AI Pedagogical Agents: A Case-Based Learning Comparison in Higher Education. *International Journal of Special Education*, 40(2), 92–103. <https://doi.org/10.52291/ijse.2025.40.24>
- Zellner, J., Koch, J., Hagmann-von Arx, P., Tönnissen, L., Link, C.-P. & Gebhardt, M. (2025b, July). *Simulationen zur Förderung der diagnostischen Kompetenzen in der heil- und sonderpädagogischen Hochschulbildung: Strukturiertes Klickspiel oder Einsatz GPTAvatar – ein Vergleich mit Noviz:innen und Expert:innen*. [Poster presentation]. Jahreskongress 2025 der Schweizerischen Gesellschaft für Bildungsforschung (SGBF) und der Schweizerischen Gesellschaft für Lehrerinnen- und Lehrerbildung (SGL). <https://doi.org/10.13140/RG.2.2.31321.22886>