TESOL Journal

WILEY

MEDIA REVIEW

D-ID Studio: Empowering Language Teaching With AI Avatars

Chenghao Wang | Bin Zou

Department of Applied Linguistics, Xi'an Jiaotong-Liverpool University, Suzhou, China

Correspondence: Bin Zou (bin.zou@xjtlu.edu.cn)

Received: 14 September 2024 | **Revised:** 17 February 2025 | **Accepted:** 6 March 2025

Funding: This study was funded by TDF (2324-R28-240) and the University Research Centre for Culture, Communication, and Society (CCCS) at Xi'an Jiaotong-Liverpool University.

Keywords: AI avatar | D-ID | GenAI agent | language teaching

ABSTRACT

Avatars play a significant role in Artificial Intelligence (AI)-powered education, supported by various human-computer interactions and second language acquisition theories. AI avatars have become increasingly anthropomorphic and realistic with advancements in speech synthesis, speech-driven lip-syncing, and speech-to-facial animation. D-ID Studio (D-ID), a versatile platform for creating AI avatar videos and Generative AI (GenAI) agents, leverages large language models (LLMs) to provide enriched educational resources and speaking practice materials for language learning. Through its multilingual, multi-accent, and customizable avatar settings, it promotes personalized language learning experiences. This media review aims to explore and discuss D-ID's key features and potential applications in language education.

1 | Introduction

AI avatars (digital humans) have been extensively utilized in language education. Equipped with speech recognition, speech synthesis, and LLMs, avatars significantly enhance interactive language learning by enabling multimodal input, providing feedback, and simulating practice (Wang et al. 2024; Zou et al. 2024). Vivid avatars further increase language learners' enjoyment, social, and cognitive presence, leading to improved learning outcomes (Wang et al. 2022). Echoing the Cognitive Theory of Multimedia Learning and Interaction Hypothesis, the presence of avatars simulates real-life conversational scenarios. Their anthropomorphic features, body language, and synchronized lip movements with speech provide learners with multisensory input, facilitating immersive learning and enhancing the likelihood of successful language acquisition (Mayer 2024; Wang et al. 2024). However, designing personalized AI avatars and chatbots tailored to specific teaching contexts is often technically complex and time-consuming for language teachers. D-ID, a user-friendly platform for creating AI avatar videos and GenAI agents (LLMs-powered embodied chatbots), skillfully navigates this challenge, making GenAI-powered language teaching and learning more diverse and accessible.

2 | Overview

D-ID (https://studio.d-id.com) is a GenAI-powered platform designed for creating animated and realistic AI avatar videos and GenAI agents. With a streamlined creation interface (see Figure 1), it simplifies the process of generating personalized avatar-based videos. The platform offers a wide range of templates and options, allowing for various integration of appearance, facial expressions, body movements, languages, speed of speech and accents, making it an ideal tool for developing multimodal instructional materials for language learners. Additionally, under the "Create an Agent" section (see Figure 2), teachers can design GenAI agents tailored to their classrooms and student needs. These agents can then be accessed through a web link or other software platforms through its Application Programming Interface (API).

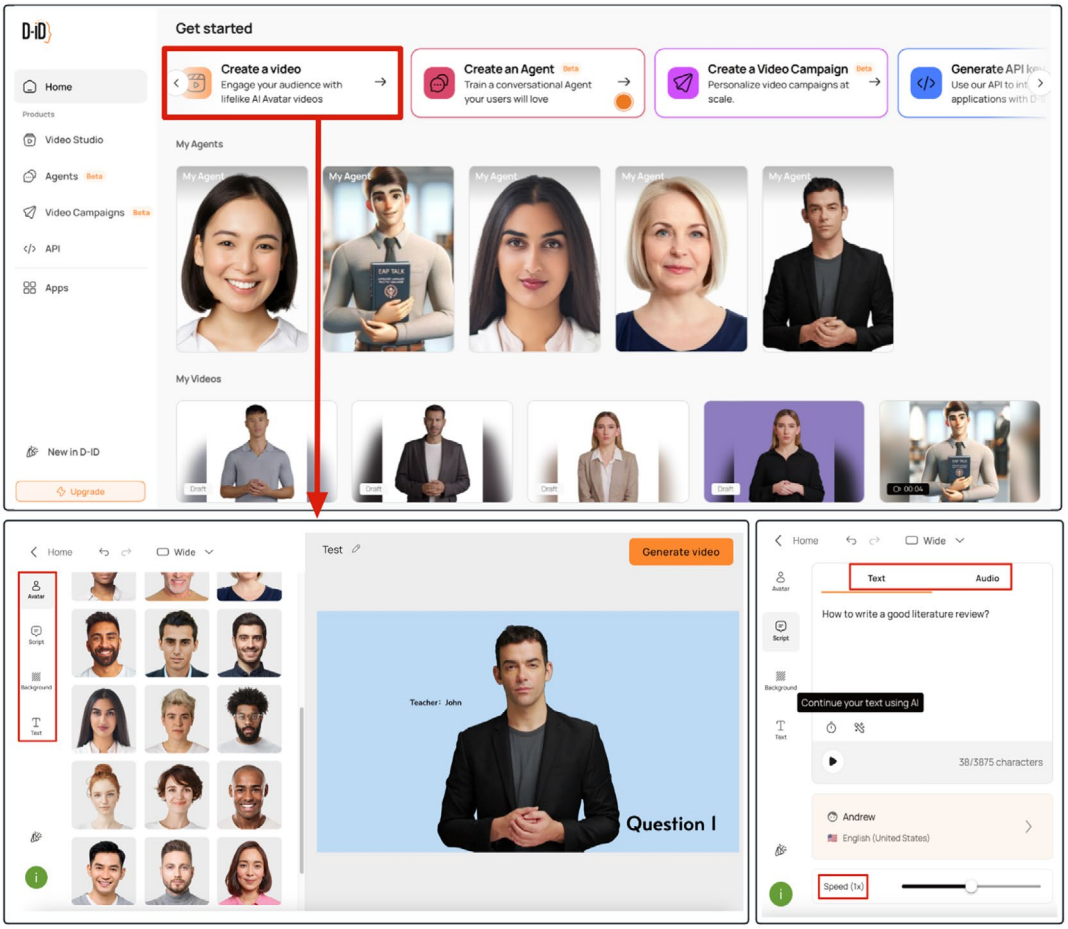

FIGURE 1 | Avatar video creation.

3 | Fostering Language Development Through a Diversity of Avatars, Languages, Accents, and Dialects

D-ID presently provides 18 free-access avatars, alongside nearly 70 additional avatars available through a premium subscription. Teachers also have the option to upload their photographs to generate personalized avatars. Moreover, the platform supports over 120 languages (see Figure 3), which provide resources for teachers of various languages, especially those teaching less commonly spoken languages. Teachers can create comparison videos in different languages, enabling learners to compare and analyze various linguistic structures, thereby enhancing comprehension and cognitive flexibility.

Notably, English encompasses 14 distinct accents, such as British, American, and Indian, reflecting a wide range of linguistic and regional variations. For English learners, exposure to such accent diversity closely mirrors authentic communicative contexts, thereby fostering improvements in listening proficiency and pronunciation (Bent and Atagi 2017).

D-ID also offers a comprehensive range of resources tailored to the dialects of countries and regions, such as Welsh in the United Kingdom, Cantonese, and Wu in China. Teachers can create avatar videos in which the same content is spoken in different dialects, offering insights into language use's cultural, historical, and social contexts. Familiarity with dialectal variations aids language learners in engaging more authentically and effectively with native speakers in specific regions. Such diverse linguistic resources foster the development of cross-cultural communication skills, which are crucial in an increasingly globalized society (Schoonmaker-Gates 2017).

Regarding vocal characteristics, the "Use case" feature in Figure 3 offers different vocal qualities. "Animation" is ideal for generating listening materials in kindergarten and primary school classrooms, while "Narration" and "News" suit higher grade levels. This versatility, with adjustable speech speed, makes it valuable for teachers to accommodate different proficiency levels.

4 | Alleviating Teachers' Workload Through the Use of AI Avatar Videos

As shown in Figures 1 and 4, teachers can upload their lecture slides in the Background and input the textual content into the Script. If the lecture content involves more than one language, they can select the multilingual option in the voice section. D-ID will then generate a natural bilingual or multilingual instructional video. For instance, Figure 4 demonstrates an AI avatar delivering content in both English and Spanish. Teachers can assign these videos for learners to watch in advance as a preview, thus reducing their workload. Alternatively, teachers can incorporate this feature into Task-Based Language Teaching by assigning group tasks where students use the platform to

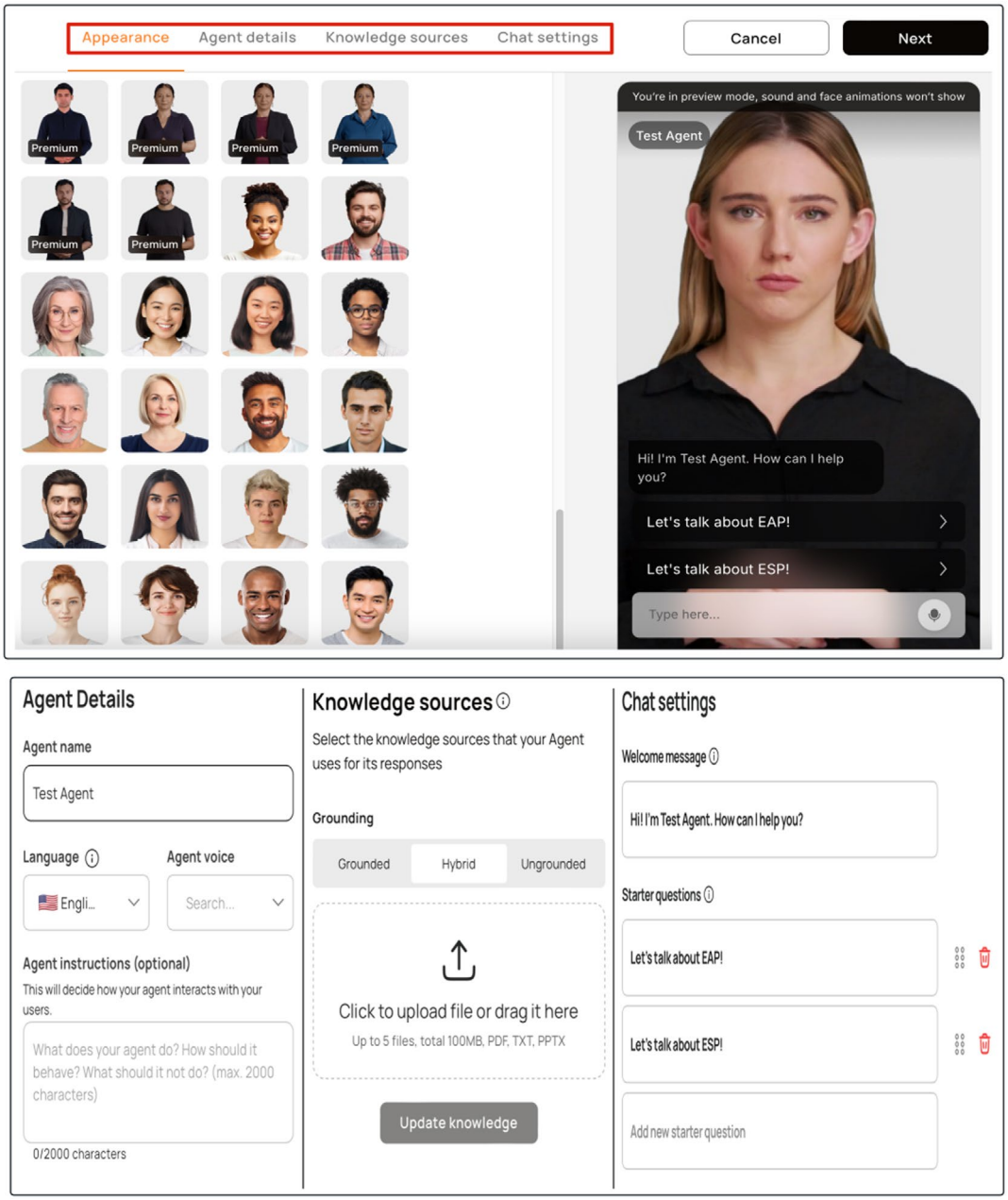

FIGURE 2 | GenAI agent creation.

convert an article into a presentation video. This facilitates students' exposure to and interaction with embodied AI-intergrated learning oppotunities, enhancing their digital literacy, learning motivation and engagement.

5 | Empowering Interactive Activities With Customized GenAI Agents

According to Figure 2, teachers can configure their own teaching aim-oriented GenAI agents by customizing the avatar's appearance, language, and voice. Moreover, the "Instructions" can be tailored to define how it interacts with students, including its role, objectives, knowledge base, and any specific guidelines to follow. As illustrated in Figure 5, the agent is designed to support learners in mastering and understanding English for Academic Purposes, serving as either a learning companion or a virtual teacher. Specifically, as discussed in Wang et al. (2024), students engage in conversational interactions with the designed agent to learn new vocabulary and sentence patterns. This avatar-guided human-computer interaction not only reduces teaching load but also enhances students' willingness to communicate and emotional learning experience.

Compared to off-the-shelf GenAI chatbots (e.g., ChatGPT 4.0), such GenAI agents offer seamless integration of LLMs, voice recognition, synthesis APIs, and avatars, while also allowing enhanced control to restrict interactive AI-generated content. Selecting Grounded or Hybrid options (see Figure 6) can help prevent the agent from delivering excessive information, making it a safer and easier-to-use tool for controlling the cognitive load brought by learning materials.

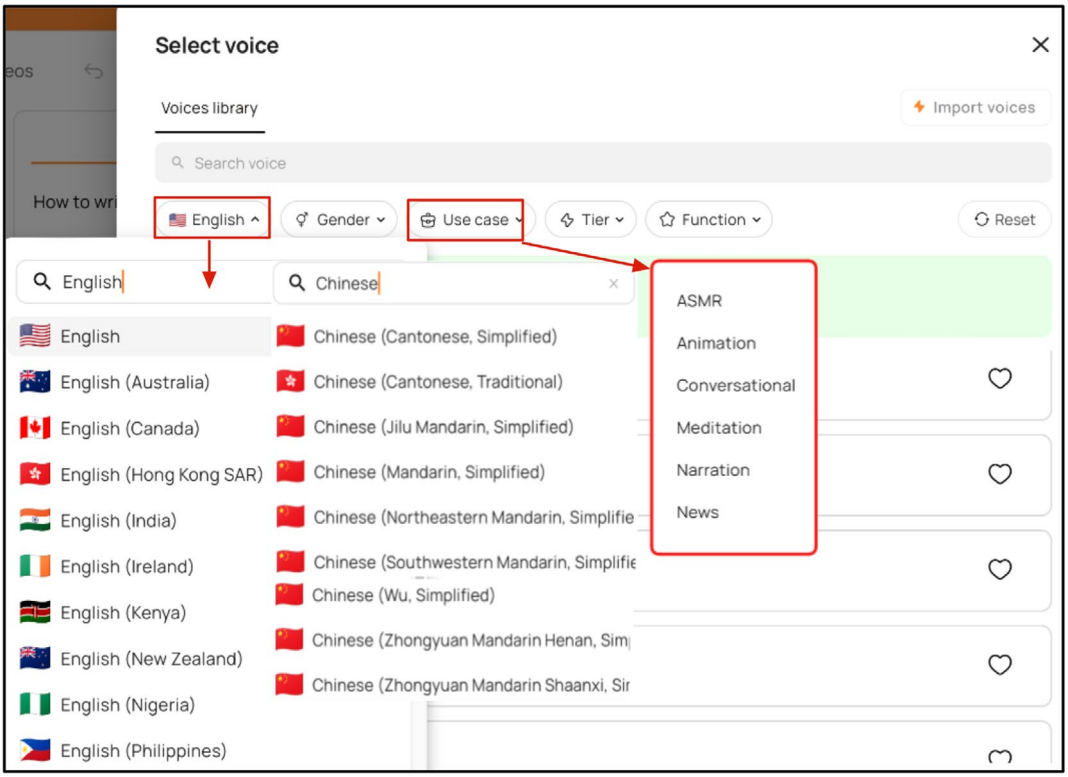

FIGURE 3 | Avatar voice setting.

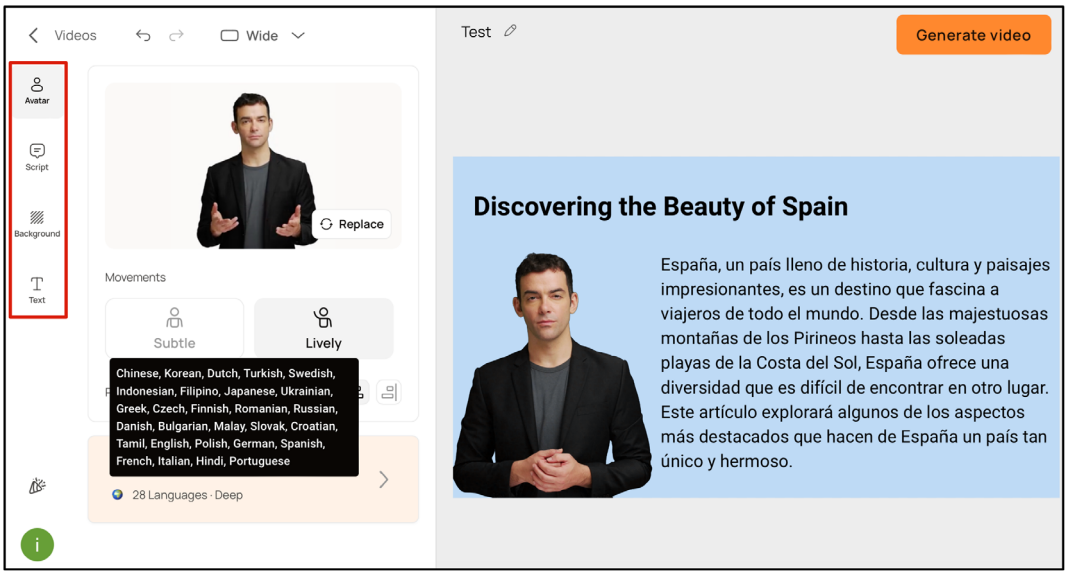

FIGURE 4 | Avatar lecture video creation.

Finally, in the Chat settings, teachers can design a "Welcome message" that helps set the tone and context for the interaction. Meanwhile, starter questions (see Figure 7) can be set up to guide the conversation, serving as scaffolding to provide a more structured and directed interaction, thereby helping users engage more effectively with the agent. Subsequently, the GenAI agent can be shared with students via a link, enabling them to interact with it anytime and anywhere. This not only enhances usability but also facilitates communicative language learning. Notably, the customization features of these agents are not available on other AI avatar creation platforms, such as Heygen and KreadoAI, thus offering greater potential for personalized GenAI-assisted oral language instruction. However, this introduces new challenges for teachers, particularly in classroom management and AI literacy. It is essential for teachers to be able to design GenAI tools thoughtfully and guide students in using the technology rationally, scientifically, and critically. By providing timely technology support, teachers can ensure the effective and safe integration of GenAI into language learning, while also reflecting on their limitations.

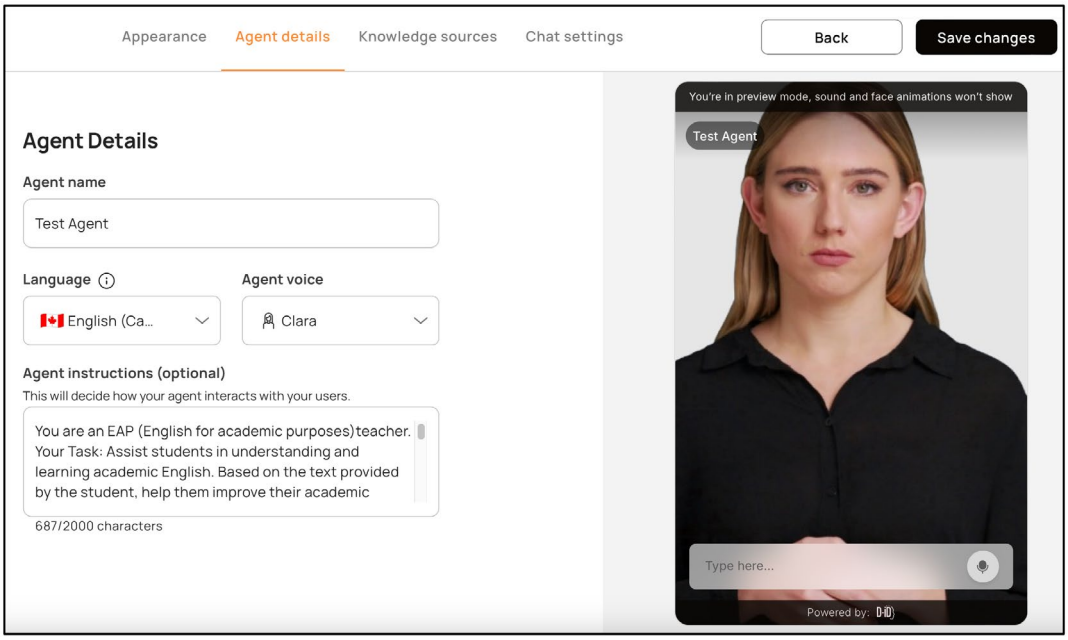

FIGURE 5 | The GenAI agent settings.

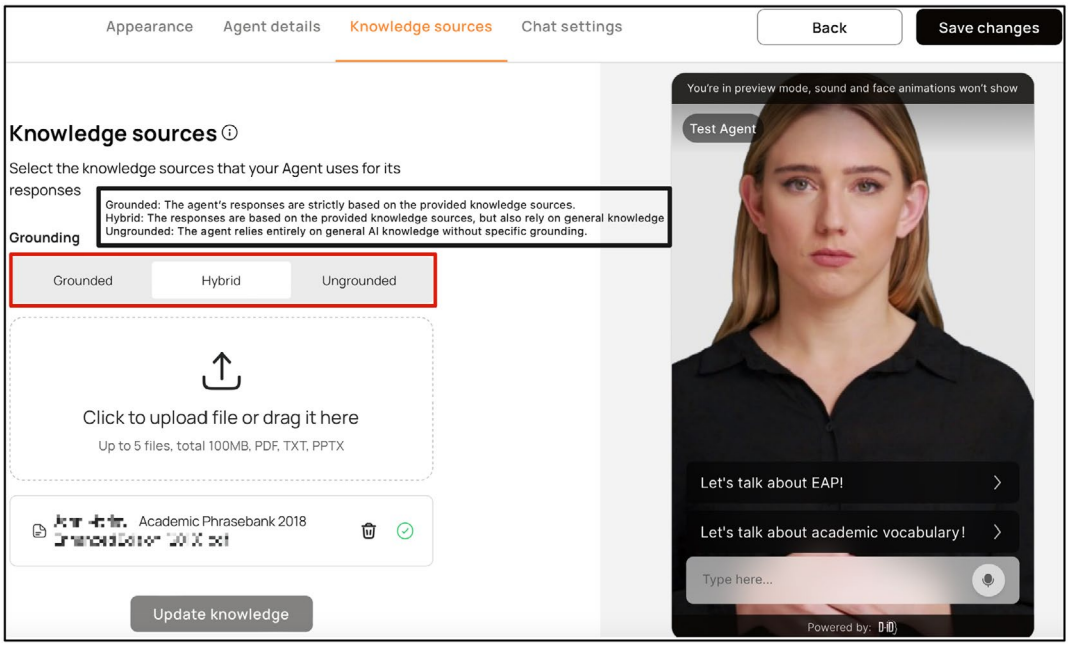

FIGURE 6 | Knowledge sources settings interface.

6 | Conclusion

D-ID is a powerful platform for designing multilingual teaching resources, significantly enhancing language education by creating personalized AI avatar videos and GenAI agents. The platform offers a wide range of customizable avatars with various languages, accents, dialects, and background settings that support both in-class and extracurricular language learning activities. These resources are applicable for lesson lead-ins, speaking and listening materials, vocabulary expansion, and cross-cultural communication. However, the platform has some limitations. For instance, the Knowledge resources for GenAI agents are restricted to a maximum upload of 100 MB, with each file not exceeding 20 MB. The free version of D-ID also limits the number of video creations and the variety of avatar appearances available to users. Upgrading to the lite, pro, or advanced versions can remove these restrictions. As LLMs advance, GenAI agents are expected to deliver more accurate and comprehensive output and respond more quickly. The innovative use of D-ID enriches the expansion of future language education resources while also presenting new challenges for teachers' digital literacy.

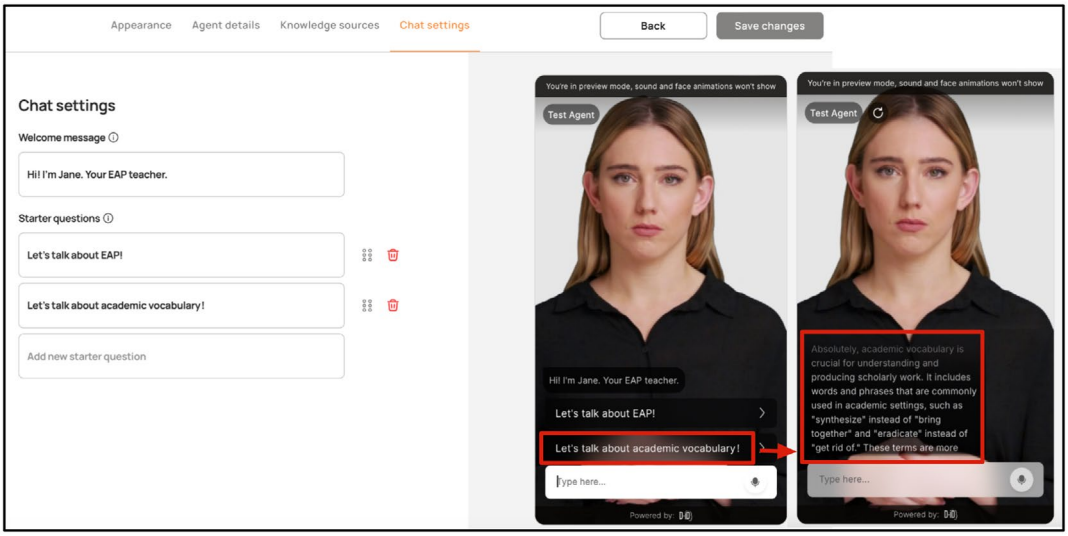

FIGURE 7 | Chat settings.

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The authors have nothing to report.

References

Bent, T., and E. Atagi. 2017. "Perception of Nonnative-Accented Sentences by 5- to 8-Year-Olds and Adults: The Role of Phonological Processing Skills." *Language and Speech* 60: 110–122. https://doi.org/10.1177/0023830916645374.

Mayer, R. E. 2024. "The Past, Present, and Future of the Cognitive Theory of Multimedia Learning." *Educational Psychology Review* 36, no. 8: 1–25. https://doi.org/10.1007/s10648-023-09842-1.

Schoonmaker-Gates, E. 2017. "Regional Variation in the Language Classroom and Beyond: Mapping Learners' Developing Dialectal Competence." *Foreign Language Annals* 50: 177–194. https://doi.org/10.1111/FLAN.12243.

Wang, C., B. Zou, Y. Du, and Z. Wang. 2024. "The Impact of Different Conversational Generative AI Chatbots on EFL learners: An Analysis of Willingness to Communicate, Foreign Language Speaking Anxiety, and Self-Perceived Communicative Competence." *System* 127: 103533. https://doi.org/10.1016/j.system.2024.103533.

Wang, X., H. Pang, M. P. Wallace, Q. Wang, and W. Chen. 2022. "Learners' Perceived AI Presences in AI-Supported Language Learning: A Study of AI as a Humanized Agent From Community of Inquiry." *Computer Assisted Language Learning* 37, no. 4: 814–840. https://doi.org/10.1080/09588221.2022.2056203.

Zou, B., C. Wang, Y. Yan, X. Du, and Y. Ji. 2024. "Exploring English as a Foreign Language Learners' Adoption and Utilisation of ChatGPT for Speaking Practice Through an Extended Technology Acceptance Model." *International Journal of Applied Linguistics*. https://doi.org/10.1111/ijal.12658.