Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 1–11
https://doi.org/10.54941/ahfe1008235

AI-Enhanced Live-Streamed English Language Learning: Designing and Evaluating an Intelligent Platform for Chinese Learners

Nan Jiang[1], Chengcheng Ma[2], and Peter Childs[1]

[1]Imperial College London, London, SW7 2AZ, United Kingdom
[2]Shanghai Jiao Tong University, Shanghai, 200241, China

ABSTRACT

Knowledge Sharing Live Streaming (KSLS) on Chinese commercial platforms such as Douyin (TikTok) and BiliBili is a popular choice for English language learning, offering attributes such as accessibility, low cost, and large-scale reach. However, a fundamental paradox exists: while effective language learning requires active engagement, real-time feedback, and meaningful interaction, KSLS is inherently a passive, one-to-many format that struggles to support these essential pedagogical elements. With AI technologies, particularly large language models, there is an opportunity to explore if AI can bridge this gap. A focus group study was conducted with 18 participants experienced in live-streamed English lessons. Thematic analysis identified two primary pain points: (1) uncertainty about pronunciation accuracy due to the absence of immediate corrective feedback, and (2) limited contextual speaking practice, as self-organised groups frequently collapsed without structure or topic guidance. Informed by these findings, an AI-integrated platform was designed with two key innovations: an AI-driven pronunciation feedback module providing real-time visual and auditory assessments, and AI-led conversational practice in breakout rooms where AI assumes four differentiated roles (mentor, expert, partner, servant) along two dimensions of closeness and guidance derived from social actor theory. The design was evaluated through Wizard-of-Oz experiments. For the pronunciation module, a within-subject study found that 63% of learners preferred multimodal (visual + auditory) feedback and 78% preferred sentence-level immediate feedback. For the AI roles, a between-subject study showed that the mentor role achieved the highest overall satisfaction (M = 4.3/5), while the partner role scored highest on comfort (M = 4.4/5). A subsequent deployment study using AI-powered chatbots further confirmed significant differences across conditions in engagement ($F = 10.521$, $p < 0.05$), vocabulary acquisition ($F = 81.213$, $p < 0.05$), and speaking proficiency across fluency, grammar, lexicon, and pronunciation. This study contributes to human-computer interaction and technology-enhanced language learning by providing empirical evidence of the KSLS paradox in English education, a theoretically grounded design framework, and experimental validation of AI integration in live-streamed language learning.

Keywords: Live-streamed english learning, AI-enhanced language education, Pronunciation feedback, Conversational practice, Human-computer interaction

Received July 4, 2026; Revised September 22, 2026; Accepted October 4, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

INTRODUCTION

Knowledge Sharing Live Streaming (KSLS) has emerged as a cost-effective alternative to traditional education in China, with platforms such as Douyin (TikTok) routinely reaching thousands of learners per session (Lu et al., 2018). However, the format is inherently passive: while KSLS provides broad access to expert instruction, it offers limited opportunities for active collaboration or meaningful interaction among learners. This tension between scale and engagement defines the central challenge this study addresses.

To explore the challenges of live-streamed English lessons, a focus group study was conducted with 18 participants. This study highlighted two main concerns: uncertainty about pronunciation accuracy and limited opportunities for contextual practice. Guided by the insights from the focus group study, a detailed design for this platform was developed. The proposed platform incorporates AI-driven pronunciation feedback to address learners' pronunciation uncertainties, and AI-led conversational practice in Breakout Rooms to offer contextual practice opportunities. AI in this context plays a guiding and facilitating role, ensuring a supportive and engaging learning environment.

This study explores two key research questions in the context of live-streamed English education for Chinese learners:

RQ1: How can AI-driven mechanisms be designed to effectively address pronunciation inaccuracies in real-time during live-streaming English classes?

RQ2: How could AI-led conversational practices in breakout rooms enhance contextual learning and promote a more dynamic learning environment?

LITERATURE REVIEW

KSLS as an Emerging Pedagogical Medium

KSLS exemplifies how live streaming has evolved beyond entertainment, enabling instructors and learners to engage in real-time discussions and collaborative learning experiences (Wu et al., 2018). However, there are challenges unique to large-scale live streaming environments. Managing thousands of participants in a single session presents difficulties for instructors, who must balance engagement with effective content delivery (Cunningham et al., 2019). Features such as chat filters and polls offer some support, but ensuring meaningful participation requires well-designed interaction mechanisms (Lu et al., 2018). Without these mechanisms, learners risk becoming passive participants, reducing the educational value of the session (Ma & Cao, 2017).

The Evolution of AI in Education

Addressing this passivity challenge requires innovative technological intervention. The introduction of generative AI tools such as ChatGPT represents a shift towards direct AI-user interaction. Unlike platform-based systems, generative AI tools engage directly with students, offering immediate

feedback and conversational support. This mode of interaction encourages self-directed learning by allowing students to seek answers independently and explore topics interactively (Ouyang & Jiao, 2021; Tzirides et al., 2023). However, misinformation and fabricated responses from generative models have raised concerns about the reliability of these tools in educational contexts. Students may encounter biased or inaccurate information, which can undermine trust in the system. Additionally, transparency becomes a critical issue; students need to understand when they are interacting with AI and be aware of its limitations (Grassini, 2023; Akgun & Greenhow, 2022).

PRE-STUDY: USER PAIN POINTS

Participants, Materials, and Procedures

Semi-structured focus group interviews were conducted with 18 participants experienced in English language learning via live-streaming platforms such as Douyin, BiliBili, and other specialized websites. The cohort was recruited through professional networks and social media advertisements, with diversity in age, gender, educational background and language proficiency. To ensure substantive experience, eligibility was restricted to individuals who had engaged with English live streams for at least three hours per week over three months. Notably, participants were selected based on their commitment to learning for personal or professional growth rather than exam-oriented purposes.

Preliminary observations identified two primary pedagogical categories in these sessions: phonetic focus, which utilizes visual aids and demonstrations to teach pronunciation and intonation, and expressions and contextual usage, which employs contemporary resources including podcasts and literature to enhance idiomatic fluency. For the study, the 18 participants were divided into three groups of six to facilitate diverse discourse. Each one-hour video-conferenced session followed a semi-structured protocol exploring learning preferences and perspectives on AI integration. All sessions were recorded with informed consent and analyzed using thematic analysis. To ensure data integrity and validity, anonymized transcripts were returned to participants for member checking.

Challenges

A major concern among participants was the uncertainty about pronunciation accuracy during live-streamed lessons. As participant P4 mentioned, practicing pronunciation without knowing if it is correct can be discouraging despite the effort invested. This issue is significantly compounded by the lack of immediate feedback; for instance, P9 noted that while they can submit homework to WeChat groups for evaluation, they prefer instant feedback during lessons to correct errors and boost confidence. Such uncertainty often leads to frustration and reduced motivation, as evidenced by P7's experience.

In addition, participants expressed concern regarding limited practice opportunities to use English in real-life contexts, desiring more chances to apply newly learned expressions. Although self-organized practice groups were

suggested as a solution, they often faced functional difficulties. As P12 explained, these groups can become "awkward" due to a lack of structured discussion topics, making it hard to maintain conversation flow and collective interest.

Design Goals & Hypotheses

These challenges directly informed the following design goals and hypotheses.

G1: Real-Time Pronunciation Feedback. As shown in Figure 1 (Goal G1), the AI acts as a mediator between teachers and students, sending aggregated data to the teacher while delivering pronunciation feedback directly to the student. Incorporating HCI principles (Rogers et al., 2023), our first hypothesis postulates that integrating AI-driven real-time pronunciation feedback can elevate learning efficiency. By offering immediate, contextual corrections, we expect the system to not only bolster pronunciation accuracy but also intensify learners' engagement and expedite their language learning progress.

H1: *Real-Time Pronunciation Feedback Enhances Learning Efficiency.*

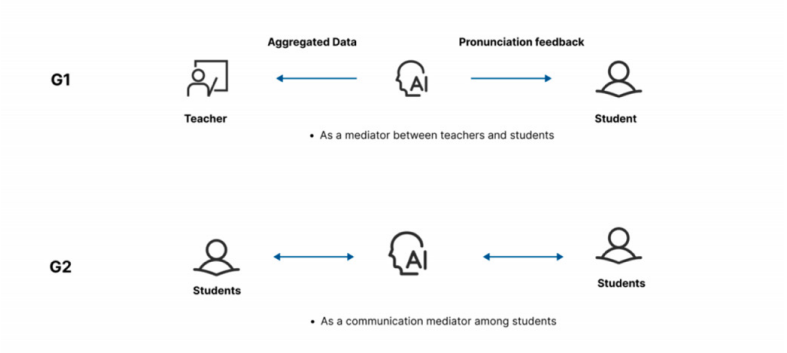


Figure 1: Design goals.

G2: AI-led Conversational Practice in Breakout Rooms. As illustrated in Figure 1 (G2), the implementation of AI serves as a communication mediator among students, enabling bidirectional interaction within the breakout room setting. Echoing HCI's emphasis on meaningful user interactions (Rogers et al., 2023), our second hypothesis suggests that AI-led and facilitated conversational practice will propel learners into more realistic English usage scenarios. We anticipate this enriched interaction to culminate in enhanced vocabulary, fluency, and overall language proficiency.

H2: *AI-led Conversational Practice Boosts Language Proficiency.*

METHODOLOGY

A mixed methodology approach was adopted utilizing observational scales, interview guidelines, and questionnaires as the primary research tools. All studies complied with the ethical requirements of Imperial College London (Science Engineering Technology Research Ethics Committee study approval

number 6704842), and informed consent was obtained from all participants prior to data collection.

Three sequential studies were conducted: a pre-study to identify user pain points, Study 1 to evaluate the pronunciation feedback module (RQ1), and Study 2 to examine AI roles in conversational practice (RQ2), followed by a deployment study using functional AI chatbots.

STUDY 1: PRONUNCIATION FEEDBACK MODULE

Participants, Materials and Procedure

Participants were recruited through social media and online English-learning communities based on three eligibility criteria: basic English proficiency, prior experience with live-streamed lessons, and a stated interest in AI-enhanced language learning. A total of 50 participants with diverse demographic and professional backgrounds were enrolled, including 18 returning members from a previous focus group study who had no prior exposure to the system prototype. To evaluate the pronunciation feedback module, a within-subject design was used where each participant experienced four counterbalanced conditions during a 20–25 minute live-streamed session: No Feedback (C0), Visual Feedback (C1), Auditory Feedback (C2), and Visual + Auditory Feedback (C3). The system interface used in the study, including the feedback console and learner-facing display, is illustrated in Figure 2.

The experimental tasks comprised instructor-led exercises at the phoneme, word, and sentence levels. To ensure a controlled yet realistic environment, a Wizard-of-Oz (WoZ) setup was implemented, where a trained operator used a dedicated console to provide real-time, templated feedback on correctness, error types, and actionable improvements. Data collection involved quantitative Likert-scale measures (1–5) assessing perceived usefulness, clarity, disruption, and confidence after each block, followed by post-session questionnaires and semi-structured interviews to explore qualitative preferences regarding feedback modality and frequency.

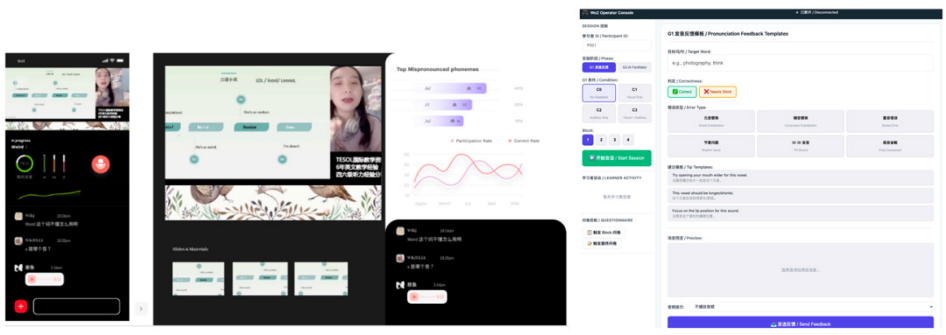


Figure 2: Design interface.

Study 1 Results and Discussion

Data analysis utilized means and standard deviations for Likert ratings and descriptive distributions for preference outcomes. Thematic analysis of interview transcripts identified the benefits and design implications of integrating real-time pronunciation feedback into live-streamed environments. Overall, participants highly valued the feedback (Mean: 4.2/5), demonstrating clear preference for multimodal delivery. Specifically, 63% of learners favored the combined visual + auditory condition (C3), noting that visual guidance provided interpretable cues while auditory references offered concrete exemplars for imitation. This preference for multimodal feedback aligns with Zhou et al. (2022), who found that combined visual and auditory cues from pedagogical agents produced stronger learning outcomes and higher motivation than single-modality feedback. In contrast, the no-feedback baseline (C0) was deemed insufficient for identifying prosodic errors such as stress or intonation. This is consistent with Neri et al. (2008), whose seminal study on computer-based speech corrective feedback demonstrated that the absence of automated feedback significantly impedes learners' ability to detect and self-correct phonemic errors. Participants favoured immediate post-utterance feedback, with 78% preferring reinforcement after every sentence. While 68% identified a 10–30 second window as optimal for error processing, 30% cautioned that excessive frequency could disrupt lesson momentum. Finally, 80% of learners desired agency over feedback frequency and duration, underscoring the necessity of personalized delivery to maintain the balance between immediacy and pedagogical continuity.

Furthermore, the scoring component served as a significant motivator rather than a source of pressure; for instance, learners (e.g., P11) reported that scores fostered a "game-like" sense of progress, thereby enhancing practice engagement and learner confidence. To maximize pedagogical impact, future systems must balance the clarity and immediacy of multimodal feedback with the need to minimize cognitive load and preserve lesson immersion. Overly granular or frequent interventions risk fragmenting the speaking flow; therefore, exploring adaptive feedback strategies that align with learners' unique error patterns and proficiency levels remains a vital direction for system optimization.

STUDY 2: AI ROLES

Participants, Materials and Procedure

This study involved the same 50 English learners from Study 1, maintaining consistent demographic and proficiency profiles. Grounded in Social Actor Theory (Reeves & Nass, 1996), we conceptualized the AI's pedagogical roles across two dimensions: Closeness (interpersonal intimacy) and Guidance (instructional supervision). The intersection of these dimensions yielded four distinct AI personas for Breakout Room interactions: Mentor (High Closeness/High Guidance), Expert (Low Closeness/High Guidance), Partner (High Closeness/Low Guidance), and Servant (Low Closeness/Low Guidance), as illustrated in Figure 3.

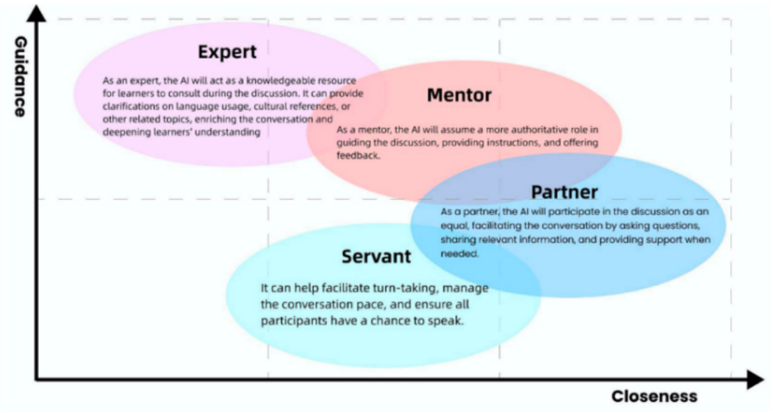


Figure 3: AI pedagogical role framework.

Employing a between-subjects design, participants were stratified by baseline proficiency and randomly assigned to a single AI condition to mitigate carryover effects. The experimental procedure comprised three phases: **Pre-test (Baseline, 8–10 min).** Participants were required to complete (a) a brief speaking task on a travel-related prompt (submitted as an audio message) and (b) a short self-report measure of speaking anxiety and confidence. **Breakout Room Practice (30 min).** Participants joined a 4–6 person Breakout Room facilitated by a Wizard-of-Oz operator enacting the assigned AI role. The session design included (i) warm-up prompts (3 min), (ii) structured role-play tasks (20 min), and (iii) wrap-up feedback (7 min). The operator protocol was to follow a role-specific guideline to ensure consistent behavior within each condition. The post-test (10–15 min) protocol included the requirement for participants to complete (a) the same speaking task format with a parallel prompt, (b) a role evaluation questionnaire, and (c) a short semi-structured interview about preferences and perceived helpfulness.

Study 2 Results and Discussion

The four conditions on post-session engagement, comfort/anxiety change (pre–post), and pre–post-performance gains have been compared. In addition, quantitative results were triangulated with thematic analysis of interview data to interpret why certain roles may feel more motivating, more structured, or more pressure-inducing in live group practice. Overall satisfaction ratings (1–5 scale) were highest for the Mentor ($M = 4.3$), followed by Partner ($M = 4.1$), Expert ($M = 3.8$), and Servant ($M = 3.6$). Specifically, the Mentor was lauded for its motivational tone and actionable guidance, while the Partner role achieved the highest comfort rating ($M = 4.4$) by fostering a low-pressure environment, followed by Mentor ($M = 4.2$), Servant ($M = 4.0$), and Expert ($M = 3.7$). This aligns with Holmes et al. (2019), who argue that adaptive AI systems offering personalised guidance are more likely to sustain learner engagement and satisfaction than directive, accuracy-focused approaches. These results suggest that expert-like precision may come at the cost of learner comfort. Perceived contribution to engagement mirrored this

pattern: Mentor (M = 4.5), Partner (M = 4.2), Expert (M = 3.9), and Servant (M = 3.7). Consequently, participants expressed a preference for expanded Mentor- and Partner-style facilitation to boost motivation, while also recognizing the Expert's utility for focused improvement and the Servant's consistency in establishing practice routines.

In-depth interviews revealed varying learner experiences with different AI roles. The Mentor was praised for providing a supportive environment, boosting confidence and motivation, though some learners desired more autonomy. The Expert was valued for its precise feedback on complex language rules, but some learners felt overwhelmed by the focus on accuracy, preferring more freedom. The Partner role fostered a stress-free environment, encouraging speaking practice, though some wanted more structured learning opportunities. The Servant was appreciated for its reliability and consistency in establishing a practice routine, but some found it less engaging, wishing for more interactive learning experiences. Taken together, these divergent experiences across roles echo Kumar and Saini (2024), who observed that the effectiveness of human-AI collaboration in learning hinges on aligning the AI's interactional stance with learners' emotional and motivational states, and that overly task-focused AI behaviour can reduce perceived relatedness and engagement.

EXPERIMENTAL EXPLORATION: AI PERSONAS AND GROUP CHAT IMPLEMENTATION

While the WoZ studies provided valuable initial insights, their ecological validity was limited by the simulated nature of AI behavior. To further validate the findings from the WoZ studies in a naturalistic environment, we implemented a functional AI system using the ChatGPT API. The study aims to pinpoint which personas offer enhanced interaction experiences and whether these experiences surpass traditional human-led interactions. A ChatGPT-powered Telegram bot was developed using Cloudflare Workers, enabling real-time group chat interactions as summarised in Figure 4.

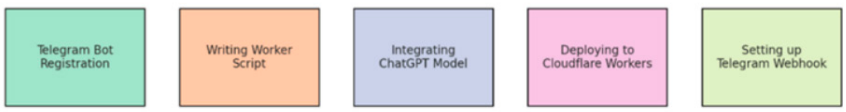


Figure 4: Overview of the end-to-end workflow.

Participants, Materials, and Procedures

Participants were recruited through language learning forums, social media, and university bulletin boards, with candidates undergoing initial screening via an online questionnaire to assess their language proficiency, learning preferences, and prior experience with AI-based tools. The study employed a

2×6 factorial design to examine the effects of session type (AI-led vs. Human-led) and facilitator role (Mentor, Expert, Partner, Servant, No facilitation, and Live teaching assistant). Following recruitment, participants completed a baseline language assessment via Telegram to establish their initial proficiency through brief instructor-led conversations. The grading focused on fluency, coherence, lexical resources, grammatical range, accuracy, and pronunciation, using a rubric adapted from the IELTS. Scores were averaged for a final proficiency score on a scale of 0 to 9, with inter-rater reliability assessed using Krippendorff's alpha.

Based on these results, participants were stratified and assigned to one of six groups: four AI-facilitated groups (A–D), where participants interacted with distinct AI personas via the Telegram Bot API; a control group (E) without assistance; and a control group (F) featuring live teaching assistants. Participants in the AI groups first completed a customization survey to calibrate the chatbot's communication style and empathy levels. Over a two-day intervention period, participants engaged in daily 15–20 minute conversation sessions on travel-related topics within their assigned Telegram groups. Post-experiment, proficiency was re-evaluated using the initial assessment protocol, and engagement was measured through the User Engagement Scale Short Form (UES-SF) and the Second Language Speaking Anxiety Scale (SLSAS). To ensure academic rigor and participant safety, all chat logs were anonymized for linguistic analysis, and the study adhered to strict data protection regulations, supported by a dedicated channel for technical assistance and participant well-being.

Results and Discussion

ANOVA was applied to evaluate the impact of various instructional approaches on learner outcomes, yielding significant variances across all measured dimensions ($P < 0.05$). Regarding engagement, the User Engagement Scale Short Form revealed an F value of 10.521, with Control Group F (human TAs) scoring highest (4.40), followed closely by the AI Partner (4.35) and AI Mentor (4.25) groups; conversely, the AI Expert (4.10), AI Servant (4.15), and the unsupported Control Group E (3.85) demonstrated lower resonance. Notably, all AI-led groups, particularly the AI Partner group, achieved a significant reduction in language anxiety compared to both control settings, suggesting that interactive AI personas foster a more supportive learning environment.

In terms of vocabulary acquisition ($F = 81.213$), the AI Mentor and AI Expert groups outperformed others with averages of 8.5 and 8.0 new words respectively, surpassing the AI Partner (7.5), AI Servant (7.0), and both Control Group F (6.8) and E (6.5). The F-values obtained from the ANOVA—9.484 for Fluency, 10.645 for Grammar, 10.235 for Lexicon, and 12.863 for Pronunciation—each correspond to P-values that decisively reject the null hypothesis for the respective tests ($P < 0.05$). Such outcomes provide compelling evidence of significant disparities in learning achievements as influenced by the employed teaching methods.

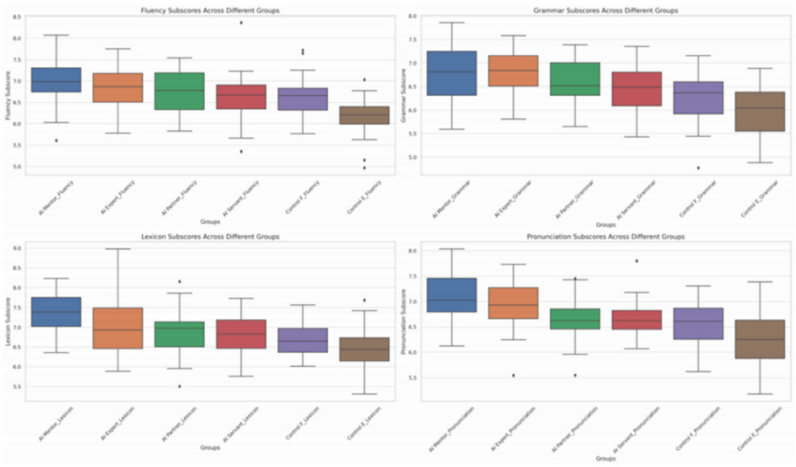


Figure 5: Fluency, Grammar, Lexicon, and Pronunciation subscores across different groups.

As shown in Figure 5, while the AI Mentor group showed comprehensive gains in fluency and pronunciation, the AI Partner group excelled in lexicon and grammar. Although Control Group F showed moderate improvements through live interaction, these gains were less pronounced than those in the AI-facilitated groups, while Control Group E exhibited only marginal progress, underscoring the limitations of traditional, non-interactive methods.

CONCLUSIONS AND LIMITATIONS

This study demonstrates that AI can meaningfully augment live-streamed English learning by providing real-time pronunciation feedback, scalable conversational practice, and adaptable social roles. While prior research has emphasized content delivery and automated feedback, our findings illustrate how diverse AI personas—specifically the Mentor, Partner, Expert, and Servant—can modulate learners' affective experiences, cognitive load, and social engagement in real time. Despite these promising implications, the study has limitations: the efficacy of the system remains heavily dependent on the AI's core capabilities, and our reliance on the "Wizard of Oz" methodology may not fully represent the performance or constraints of a fully autonomous, production-scale system. Consequently, future research should focus on developing and testing fully functional AI systems in real-world settings to provide more reliable, longitudinal data. Furthermore, exploring multimodal models for prosody-level assessment and expanding AI applications to address broader challenges in live-streamed education will be essential for advancing adaptive, human-AI collaborative learning.

ACKNOWLEDGMENT

The principal author would like to express sincere gratitude to Professor Peter Childs for his valuable advice and guidance. The author also appreciates the AHFE organizing committee for providing the opportunity to present this research.

REFERENCES

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, *2*(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7

Cunningham, S., Craig, D., & Lv, J. (2019). China's livestreaming industry: Platforms, politics, and precarity. *International Journal of Cultural Studies*, *22*(6), 719–736. https://doi.org/10.1177/1367877919834942

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, *13*(7), 692. https://doi.org/10.3390/educsci13070692

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Kumar, A. K., & Saini, A. (2024). Human and AI collaboration in the higher education environment: Opportunities and concerns. *Cognitive Research: Principles and Implications*, *9(1)*, 39.

Lu, Z., Xia, H., Heo, S., & Wigdor, D. (2018). You watch, you give, and you engage: A study of live streaming practices in China. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Paper 466, 13 pages). ACM. https://doi.org/10.1145/3173574.3174040

Ma, X., & Cao, N. (2017). Video-based evanescent, anonymous, asynchronous social interaction: Motivation and adaptation to medium. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW'17) (pp. 770–782). https://doi.org/10.1145/2998181.2998256

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *Language Learning & Technology*, *12*(3), 373–401.

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, *2*, 100020. https://doi.org/10.1016/j.caeai.2021.100020

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Rogers, Y., Sharp, H., & Preece, J. (2023). *Interaction design: Beyond human-computer interaction* (6th ed.). Wiley.

Tzirides, A.-O., Saini, A., Zapata, G., Searsmith, D., Cope, B., Kalantzis, M., Castro, V., Kourkoulou, T., Jones, J., da Silva, R. A., Whiting, J., & Kastania, N. P. (2023). *Generative AI: Implications and applications for education*. arXiv. https://doi.org/10.48550/arXiv.2305.07605

Wu, Q., Sang, Y., Zhang, S., & Huang, Y. (2018). Danmaku vs. forum comments: Understanding user participation and knowledge sharing in online videos. In Proceedings of the 2018 ACM Conference on Supporting Groupwork (GROUP '18) (pp. 209–218). https://doi.org/10.1145/3148330.3148344

Zhou, Y., Wang, X., & Li, H. (2022). The impact of emotional feedback and elaborated feedback provided by a pedagogical agent on learners' emotions, intrinsic motivation, agent perception, cognitive load, and transfer performance in multimedia learning. *Frontiers in Psychology*, *13*, 810194. https://doi.org/10.3389/fpsyg.2022.810194