BSIJ *Business System & Innovation Journal*
Research • Innovation • Digital Transformation

YAYASAN FATHUR RAHMAN
Vol. 01, Issue 02, 2026 | e-ISSN 3163-2831

AI-Assisted Audio Learning In Educational Settings: Opportunities, Challenges, And Future Research Directions

Muhammad Fauzan Hawari[1]*, Afifah Mesha Putri[2], Akmal Andri Yantama[3], Siti Syahsudarmi[4]

1 Faculty of Social and Political Sciences, Gadjah Mada University, Yogyakarta, Indonesia
2 Information Systems Technology and Design, University of Technology and Design, Singapore
3 Faculty of Computer Science, Lancang Kuning University, Pekanbaru, Indonesia
4 Master of Management Program, Postgraduate School, Islamic University of Riau
* Correspondence: muhammadfauzanhawari@mail.ugm.ac.id

OPEN ACCESS
No article processing charge (APC)

ARTICLE INFORMATION

Page
75-85

Received
1 September 2026

Revised
10 September 2026

Accepted
11 September 2026

Published
12 September 2026

DOI
https://doi.org/10.67955/bsij.v1i2.19

Abstract. This systematic literature review examines how audio-based artificial intelligence (AI) tools are being used to support learning across educational contexts and what challenges and opportunities arise from their implementation. Using a structured search across major e-scientific databases, this study identified and analysed empirical and conceptual works on AI assisted audio learning, including audio feedback, voice tutors/assistants, adaptive audio modules, and AI generated podcasts. VOSviewer was employed to map keyword co-occurrences and thematic clusters, while a large language model (LLM)-based notebook was used to assist in summarising abstracts, methods, and key findings, and to group studies by educational level and AI technology type. The findings indicate that audio-based AI interventions generally contribute positively to cognitive outcomes, motivation, and engagement, particularly by enabling frequent, personalised feedback and flexible, on-the-go learning opportunities. At the same time, the review reveals several challenges related to instructional design, teacher orchestration, data privacy, and the risk of learner overreliance on AI feedback. The existing evidence is heavily concentrated in language learning and a few specialised domains, with a predominance of short-term studies and self-report measures, limiting the generalisability of current conclusions. This review contributes by offering an updated mapping of AI-assisted audio learning research, highlighting underexplored contexts and design configurations, and outlining clear directions for future studies, including longitudinal designs, comparative trials of audio versus text-based feedback, and investigations into how teachers and students negotiate the role of AI in classroom practice.

Keywords: artificial intelligence; learning media; technology acceptance model; perceived trust; high school students; behavioral intention

1. Introduction

Advances in artificial intelligence (AI) have driven significant changes in the world of education, particularly in how students access information, interact with learning materials, and receive more personalized learning support (Giannakos et al. 2025; Baldrich, Pérez-García, and Santamarina-Sancho 2025). The presence of AI opens opportunities for the creation of learning processes that are more flexible, faster, and adaptive to user needs (Crompton et al. 2024). In this context, this study was conducted to systematically examine the use of voice assistant-based AI along with related learning strategies. The focus of the study is directed toward the forms of technology utilization, the benefits generated, implementation challenges, and its role in supporting the learning process. Through this study, the researchers aim to gain a more comprehensive understanding of the direction of research development in this field so that it can serve as a foundation for the development of more relevant and innovative learning (Grützmacher et al. 2025; Strielkowski et al. 2025).

The urgency of this research is increasingly evident because the use of AI in education is growing rapidly, while empirical evidence regarding its effectiveness still yields mixed results (Koolivand et al. 2024; T. C. Yang, Hsu, and Wu 2025). On one hand, this technology offers easy access to information, responsive interactions, and more personalized learning support (Paradis 2023; H. Lin and Chen 2024). On the other hand, however, questions remain regarding user readiness, the quality of system responses, data security, and the alignment of the technology with learning needs (Ratten 2023). These conditions indicate that the use of AI cannot be based solely on technological trends but requires a strong academic foundation to ensure its implementation truly delivers tangible benefits. Therefore, this research holds significant value both theoretically and practically in supporting technology-based educational transformation (Praetorius et al. 2025; Darmawansah et al. 2025).

In practice, voice assistant-based AI is increasingly being used in various learning activities, such as providing instructions, answering students' questions, offering feedback, and supporting self-directed learning (Koolivand et al. 2024; Cooper, Ilich, and Sturrock 2026). Nevertheless, its implementation has not always been optimal due to various factors, such as users' digital literacy levels, device quality, network stability, and the design of the system used

(Tomchakovskyi et al. 2026). In many situations, both instructors and students are still in the process of adapting to the new interaction patterns offered by this technology. This phenomenon indicates that AI holds immense potential in education, but its implementation still requires careful refinement to ensure it operates effectively and sustainably (Li et al. 2025; Fuchs et al. 2025).

If this situation is not promptly understood and properly managed, the use of AI in education risks failing to yield significant results (F. F. Yang, Ren, and Gu 2022). The technology will merely serve as a supplement without making a tangible contribution to improving learning outcomes. In fact, if implemented without proper planning, AI can lead to confusion, excessive reliance, reduced social interaction in learning, and a decline in the quality of students' learning experiences (Liu, Ma, and Chen 2024). In the long term, such a situation could actually hinder educational innovation, which should be focused on improving the quality of learning.

The approach used in this study is a systematic literature review (SLR). This method was chosen because it allows researchers to search for, select, evaluate, and analyze articles in a structured manner based on predetermined criteria (Hamman-Ortiz et al. 2025; Shi and Aryadoust 2024; Grützmacher et al. 2025). With an SLR, researchers can filter out the most relevant findings, minimize bias, and avoid overly general discussions. This approach also yields stronger conclusions because they are based on a selected and verified body of scientific evidence. Thus, the research can be organized systematically and is academically accountable (Zerkouk, Mihoubi, and Chikhaoui 2025; Zhang et al. 2025).

2. Related Works

2.1. Audio Feedback In Technology-Enhanced Learning

Research on AI-based voice assistants in education indicates that this technology is used to provide instant assistance, brief explanations, reminders, and navigation through course materials within digital learning environments e.g., LMS or mobile apps (Wang et al. 2024). Several studies report that voice assistants improve perceptions of ease of use, accessibility, and student engagement, particularly when used to support self-directed learning and project-based assignments (Krsti et al. 2022; Hong and Tai 2025; Tomchakovskyi et al. 2026). However, the direct impact on learning outcomes tends to vary and is highly dependent on task design, course context, and intensity of use (Kasneci et al. 2023). These findings position voice assistants as a component supporting the learning process, rather than a full replacement for the instructor's role (Kerr 2017; Li et al. 2025; Sungkur and Shibdeen 2025).

2.2. Audio Feedback In Technology-Enhanced Learning

Studies on audio feedback indicate that voice-based feedback is used to evaluate complex tasks and provide more detailed guidance for improvement compared to written comments (Cooper, Ilich, and Sturrock 2026; Fuchs et al. 2025). Students generally perceive audio feedback as more personal, easier to understand, and helpful in clarifying assessment expectations. Recent research on AI-assisted vocal training indicates that a combination of audio comparison tools, dialogic feedback with generative AI, and reflective journals significantly enhances metacognitive abilities, but does not result in a significant difference in singing performance improvement between the experimental and control groups (Li et al. 2025). These findings suggest that in the context of vocal training, the integration of AI is more effectively utilized to develop awareness and regulation of the learning process through structured reflection mechanisms, rather than being expected to directly produce short-term improvements in vocal performance.

2.3. Audio Feedback In Technology-Enhanced Learning

In recent years, generative AI has begun to be viewed not merely as an automation tool, but as a strategic component that directly shapes how learners receive material, practice, and reflect on their learning process. In EFL academic writing studies, multimodal generative AI is used to support drafting and revision, thereby enhancing academic vocabulary, syntactic complexity, and text coherence through visual feedback and iterative prompting (Tomchakovskyi et al. 2026). In medical education, generative AI is used to create cinematic clinical narratives that significantly enhance situational interest, engagement, and retention of pharmacology material compared to traditional case studies (Bland 2025). In vocal training, generative AI is combined with audio comparison tools and reflective journals to strengthen metacognition, resulting in improved metacognitive abilities without significant differences in singing performance between groups (Li et al. 2025). Additionally, generative AI-based pronunciation trainers provide real-time pronunciation feedback with high levels of satisfaction and task

success, although interpreting the details of phonetic feedback remains a challenge for users (Huang 2024).

3. Methods And Review Process

This study employed a systematic literature review (SLR) methodology in accordance with the guidelines of Kitchenham et al. (2010) and the search protocol strategy of Higgins and Green (2008). The review process comprised several key stages: formulation of the research question, design of the search schema, study selection, data extraction, and synthesis of findings. These stages are visualized in a PRISMA flow diagram (see Figure 1) to ensure that the article selection process is conducted objectively and is traceable by readers.

3.1. Search Technique In E-Scientific Database

The literature search was conducted exclusively in the Scopus e-scientific database due to its extensive coverage of reputable journals in the fields of education and educational technology, as well as its support for filtering by journal quartile. The search used the following keywords: (TITLE-ABS-KEY("voice assistant") OR TITLE-ABS-KEY("audio feedback")) AND TITLE-ABS-KEY("generative AI") AND (TITLE-ABS-KEY("education") OR TITLE-ABS-KEY("teaching") OR TITLE-ABS-KEY("learning")). The search was limited to English-language journal articles and yielded 31 documents, which were then further screened for eligibility according to the PRISMA flowchart Figure 1 and Table 1.

3.2. Study Selection Based On Title And Keywords

In the initial screening phase, all 31 documents retrieved from Scopus were filtered based on titles and keywords. This screening aimed to identify articles that explicitly examined the use of voice assistants, audio feedback, generative AI, or multimodal AI tools in the context of learning, teaching, or assessment, while excluding studies that only discussed the technical aspects of AI without a pedagogical focus. Articles that clearly fell outside the domain of education, were purely conceptual, or did not contain terms related to voice/audio-based AI and generative AI in the title or keywords were eliminated at this stage. This screening process reduced the number of documents from 31 to 16, as shown in Figure 1.

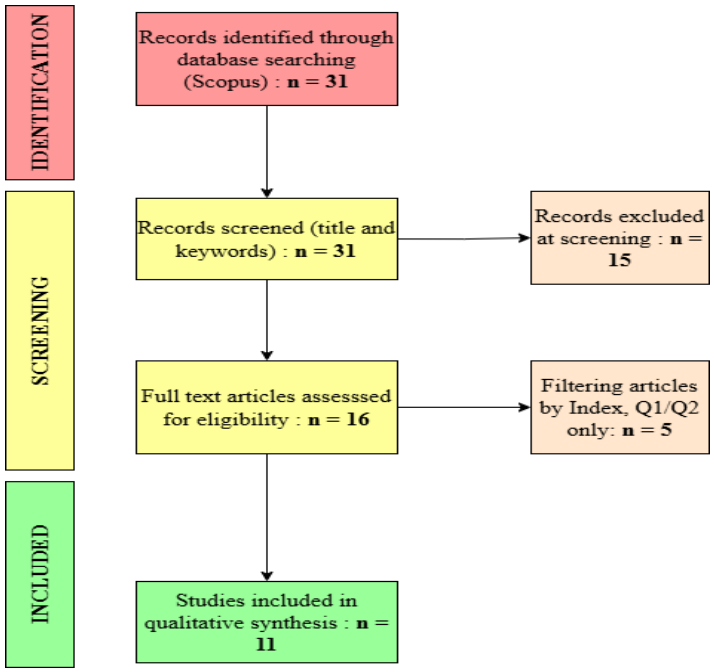

Figure 1 Study selection process based on title, keywords, and index

Table 1 Search scheme

Key categories	Exploration / search terms
Technology / Modality	"voice assistant", "virtual assistant", "audio feedback", "speech-based feedback", "pronunciation trainer"
AI approach	"generative AI", "large language model", "LLM", "multimodal AI", "AI chatbot"

Educational context	"education", "higher education", "university", "teaching", "learning", "training", "EFL", "medical education"
Learning focus	"speaking", "pronunciation", "writing", "metacognition", "engagement", "assessment", "feedback"
System role	"digital assistant", "AI tutor", "AI-based feedback", "AI-driven assessment", "virtual docent"

3.3. Full-Text Screening for Inclusion and Exclusion Criteria

At this stage, all articles that had been screened based on title, keywords, and database indices were read in full text to ensure compliance with the established inclusion and exclusion criteria. The primary focus of the selection process was direct relevance to the use of AI-based audio in educational contexts, the clarity of the reporting of methods and findings, and the empirical or conceptual contribution to the topic under study. Articles that do not provide sufficient information about intervention design, educational context, or learning outcomes are eliminated at this stage so that only substantial studies are included in further analysis.

3.4. Research Mapping and Synthesis

The studies that passed the full-text screening were then systematically analyzed by combining the results of the initial mapping using VOSviewer and an LLM notebook with an in-depth reading of each article. For each study, researchers extracted information regarding the objectives, educational context, research design, type of AI-based audio intervention, as well as learning outcome indicators and key findings, then mapped these findings into pre-defined thematic categories. This process enabled the identification of established research clusters, under-explored areas, and recurring patterns of challenges and opportunities, which then formed the basis for the results and discussion sections, as well as the formulation of research gaps for future studies

4. Results And Integrated Scenarios

This section presents the main findings of the systematic review, a mapping of learning scenarios involving AI-based audio, and a discussion of their implications for practice and future research. The presentation begins with a description of the characteristics and categorization of the 11 included studies, followed by an analysis of challenges and opportunities, a mapping of keywords and integrated scenarios, and a discussion of the findings in relation to the relevant literature.

4.1. Research Mapping and Synthesis

This section synthesizes the findings from 11 empirical studies that implemented various forms of AI-based audio tools in educational contexts. These studies span primary, secondary, and higher education levels, as well as adult learners, with research designs including experimental, quasi-experimental, survey, and mixed-methods approaches. To provide a structured overview, all studies were categorized based on educational level, type of AI audio technology used, context or subject area, research design, and the type of learning outcomes measured, as summarized in Table 2. In general, the reported learning outcomes are grouped into four main categories: the cognitive domain, the metacognitive domain, the affective domain, and indicators of learning engagement (e.g., time-on-task, participation, and persistence).

Table 2 Categorising of reviewed

Article	C1	C2	C3	C4	C5	C6	C7	C8
(Ariya et al. 2025)	√	-	√	√	√	-	√	-
(Tomchakovskyi et al. 2026)	√	√	-	√	√	√	-	-
(Michou et al. 2026)	√	-	-	-	-	-	-	-
(Huang 2024)	√	-	√	√	√	√	√	-
(Li et al. 2025)	√	√	-	√	√	√	√	√
(Bland 2025)	√	-	√	√	√	-	-	√

(Sungkur and Shibdeen 2025)	√	-	√	√	√	√	√	√
(Cooper, Ilich, and Sturrock 2026)	-	-	√	√	-	√	-	-
(Rienties et al. 2025)	-	-	√	√	√	-	√	-
(Fuchs et al. 2025)	-	-	-	-	-	√	-	-
(Uppalapati, Dabbiru, and Kasukurthi 2025)	√	-	√	√	√	√	-	-
(Tomchakovskyi et al. 2026)	√	√	-	√	√	√	-	-

Notes:
C1: Cognitive outcomes
C2: Metacognitive outcomes
C3: Affective outcomes
C4: Engagement / behavior
C5: Distance / blended learning context
C6: AI audio feedback
C7: AI voice tutor / assistant
C8: AI audio modules / quizzes / microlearning

4.2. Mapping of Keywords for Research on AI Audio in Education

Bibliometric mapping was conducted using VOSviewer software to understand the research landscape related to AI-based audio in education. Data extracted from scientific databases (article titles, abstracts, and keywords) were imported into VOSviewer, followed by a keyword co-occurrence analysis that set a minimum occurrence threshold to ensure only frequently appearing terms were included in the map. Each keyword is represented as a node, while the lines connecting the nodes indicate how often the two keywords appear together in the same article. The thicker the line and the closer the distance between nodes, the stronger the conceptual relationship within the literature.

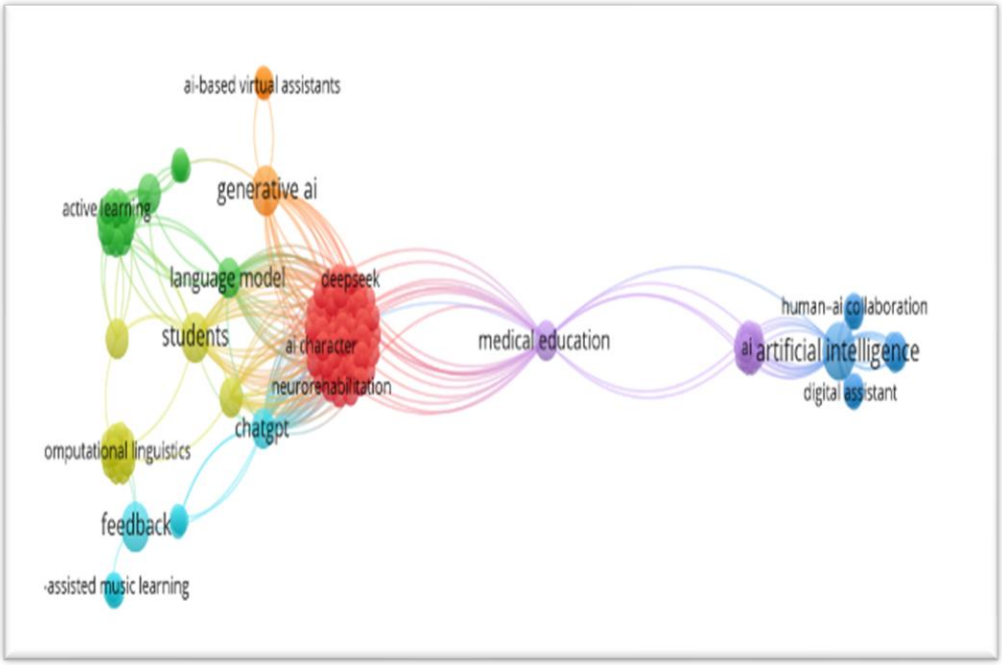

Figure 2 VOSviewer keyword map of audio AI in learning

Figure 2 shows several colored clusters representing broad thematic groups. The green–yellow cluster on the left indicates connections between keywords such as generative AI,

language model, students, feedback, and chatbot, illustrating the use of generative models and AI-based assistants to support student learning. In the center, a large red cluster focuses on medical education and is connected to terms like deep learning and neural networks; this indicates a concentration of research on the use of audio and AI in medical education and clinical training. On the right side, the blue cluster contains keywords such as artificial intelligence, digital assistant, and human–AI collaboration, emphasizing the technical and design dimensions of intelligent assistants in the context of learning. Several other keywords related to speech recognition, text-to-speech, and natural language form a sub-cluster that highlights the foundational speech and language processing technologies underpinning these various applications.

The patterns of keyword density and inter-cluster connectivity on this map indicate that the research not only focuses on improving language skills but also extends to the contexts of medical education, vocal training, and the development of core technologies for speech recognition and synthesis. At the same time, the connecting pathways between clusters remain relatively limited, particularly between technical terms (e.g., speech recognition, language model) and explicit pedagogical terms, indicating that research that truly integrates technical innovations with detailed voice-based learning design remains an important opportunity to be explored in the future.

4.3. Challenges in Implementing AI Audio for Education

Of the eleven studies reviewed, three main categories of challenges emerged in the implementation of AI-based audio tools in an educational context: technical, pedagogical, and socio-ethical.

4.3.1. Technical

Technically, several studies have reported issues with sound quality, response latency, and the accuracy of speech recognition and transcription, particularly when accents, ambient noise, or recording quality vary (Ariya et al. 2025; Huang 2024). Generative systems also risk producing inaccurate or excessive information, so AI outputs still require manual verification by instructors (Fuchs et al. 2025).

4.3.2. Pedagogical

At the pedagogical level, these studies highlight the challenges of integrating AI-generated audio into meaningful instructional design, rather than simply adding a new tool. Educators report a high initial workload in designing prompts, aligning feedback with assessment rubrics, and ensuring students do not become overly reliant on AI so that self-regulation and learning initiative are maintained (Tomchakovskyi et al. 2026; Sungkur and Shibdeen 2025).

4.3.3. Social Ethics

From an ethical and social perspective, prominent issues include voice data privacy and security, potential bias in models, disparities in infrastructure access, as well as concerns regarding academic integrity and shifts in professional identity when certain feedback roles are automated (Rienties et al. 2025). These findings underscore the need for robust institutional policies and human oversight to ensure that the use of audio AI truly supports rather than replaces the role of educators (Cooper, Ilich, and Sturrock 2026).

4.4. Challenges in Implementing AI Audio for Education

AI-based audio tools offer significant opportunities to enhance cognitive, metacognitive, and affective learning outcomes, as well as student engagement. Several studies in this corpus report improvements in cognitive achievements, such as academic writing proficiency, pronunciation accuracy, singing performance, and comprehension of complex material when audio feedback, voice tutors, or audio modules are systematically used in structured tasks and exercises. In the domains of metacognition and self-regulation, the use of audio feedback and dialogic interactions with large language models fosters processes of reflection, self-awareness, and gradual revision in both writing and vocal training contexts, although the risk of over-reliance on AI remains a concern. On the affective side, findings from eleven studies indicate that the use of voice tutors and personalized audio feedback tends to reduce speaking anxiety, increase motivation, and make the learning experience more enjoyable for students. Adaptive voice interactions and instant feedback also appear to foster higher engagement and more consistent learning habits, particularly when implemented in online and blended learning environments.

4.5. Challenges in Implementing AI Audio for Education

A synthesis of eleven studies indicates that the implementation of AI-based audio in education tends to cluster into several recurring usage patterns. To facilitate visualization and instructional design, these patterns are presented in a single integrated scenario map that illustrates how teachers design courses, how students interact with various AI audio tools, and the types of learning outcomes targeted. Figure 3 presents four main pathways that can be implemented independently or combined within a single course according to the objectives, context, and needs of the students.

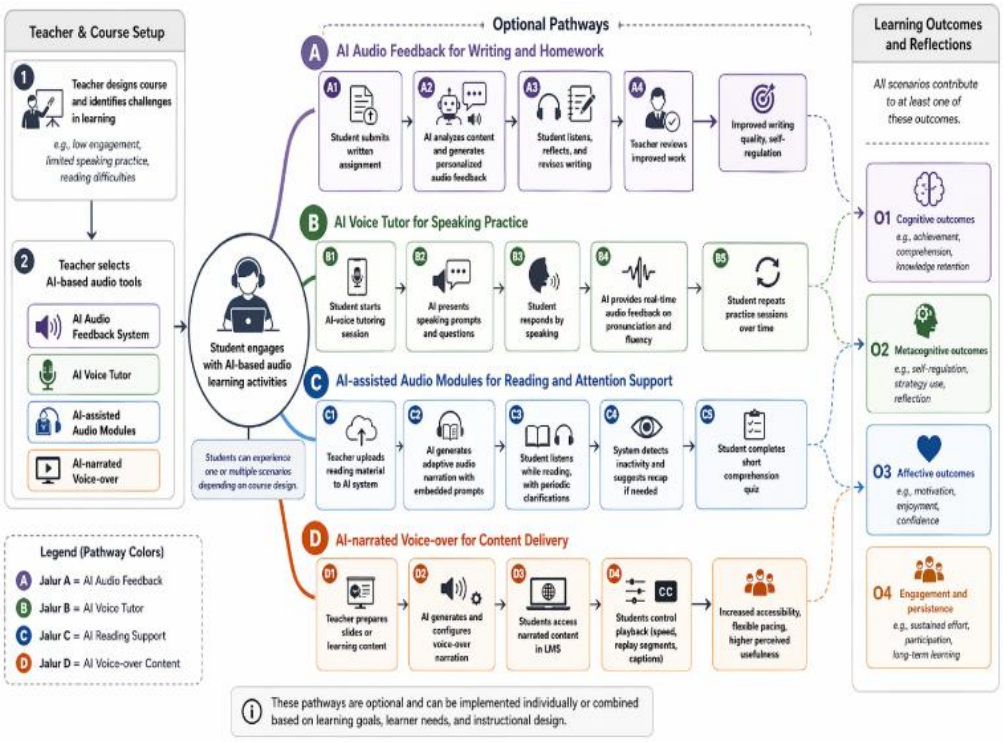

Figure 3 Integrated learning scenarios with AI-based audio

4.5.1. Scenario A – AI Audio Feedback for Writing and Performance

In this scenario, AI powered audio serves as a mediator between students' work and the iterative revision process. Verbal feedback helps highlight aspects of language, structure, or performance quality in a more natural and accessible way, including for students who struggle to read lengthy written feedback. Several studies within the corpus indicate that this practice is associated with improvements in writing quality, vocal performance, and support for reflection and self-regulation, although it still requires teacher supervision to ensure the accuracy and tone of the feedback.

4.5.2. Scenario B – AI Voice Tutor for Speaking Practice

This scenario positions AI as a conversation partner that is always available for speaking practice or subject-matter consultation, particularly in language learning and remote learning support. Two-way voice interaction allows for more frequent and flexible practice compared to relying entirely on in-person schedules with instructors. Based on the reviewed studies, this scenario is associated with improved speaking fluency, self-confidence, and students' positive perceptions of the availability of instant learning assistance; however, task design and usage restrictions are still necessary to prevent students from passively relying on AI responses.

4.5.3. Scenario C – Ai-Assisted Audio Modules For Reading And Practice

In this scenario, AI-based audio modules are used to facilitate reading comprehension, maintain attention, or practice specific skills such as pronunciation or singing. Several studies indicate that this approach helps students with concentration difficulties or auditory learning preferences, and can improve comprehension and time-on-task when narration and audio interventions are designed progressively. The main challenge lies in developing the content and adjusting the level of adaptability so that the module is not merely an "audiobook," but truly supports active learning strategies.

4.5.4. Scenario D – AI-narrated Voice-over for Content Delivery

This scenario leverages AI-generated audio to enhance content delivery, for example through cinematic clinical cases or lecture materials narrated in a more engaging style. According to the study's findings, this approach is associated with increased engagement, knowledge retention, and perceived ease of access, particularly for students learning online or who have difficulty reading long texts. However, to avoid being merely passive, voice-overs often need to be combined with follow-up activities (discussions, quizzes, or application-based assignments) so that students are encouraged to actively process the material, rather than just listen.

4.5.5. Combining Scenarios in a Single Course

In educational practice, the four scenarios in Figure 3 tend to emerge as a series of interconnected activities rather than as standalone components. For example, a narrative of the material with an AI-narrated voiceover can be used at the beginning to introduce concepts, followed by exercises using ASR/TTS-based audio modules to deepen understanding, and concluded with a voice tutor session or personalized audio feedback to help students produce language or demonstrate their best performance. This workflow provides a clear path from content presentation, continuing through structured practice, and culminating in feedback that encourages revision and self-reflection.

At the classroom level, the combination of audio feedback and a voice tutor can also be structured as a learning cycle in which students practice independently with a voice assistant, then submit formal assignments and receive more in-depth audio feedback from the system or an instructor. This approach helps provide opportunities for repeated practice without compromising the quality of guidance, even when there are a large number of participants or when learning takes place online. Conceptually, paths A–D in Figure 3 can be viewed as pathways that can be linked and adapted according to learning objectives, available resources, and learner characteristics, thereby forming a more adaptive AI audio learning ecosystem.

5. Discussion

These SLR findings indicate that AI-based audio is effective when used as a learning mediation tool, rather than as a substitute for teachers or instructors. Various forms of intervention, such as audio feedback, voice tutors/assistants, adaptive audio modules, and AI voice-overs, support conceptual understanding, repeated self-practice, and more frequent and personalized feedback, provided they are integrated into a learning workflow with explicit objectives, tasks, rubrics, and a cycle of practice-feedback-revision (Y. Lin et al. 2023).

In practical terms, educators need to clearly define when AI audio should be used for presenting concepts (narration or voice-over), for guided practice (voice tutor or speech shadowing), and for immediate, actionable feedback. Usage limits also need to be established, for example by restricting the types of assistance that may be requested from AI and specifying which parts of the task must be completed independently, so that AI support does not undermine students' self-regulation and learning initiative (Uppalapati, Dabbiru, and Kasukurthi 2025). Within this framework, teachers and instructors serve as designers of the learning experience as well as the final interpreters of AI outputs, particularly for summative assessments and material requiring high-level analysis.

At the system and policy level, a learning platform is needed that enables the integration of various AI audio scenarios within a single ecosystem, along with mechanisms for educators to edit feedback, monitor logs of student interactions when using AI, and adjust configurations according to class characteristics. Institutions need to establish clear guidelines regarding voice data privacy and security, the use of recordings and transcripts for model training, and updates to academic integrity regulations that explicitly govern the use of voice assistants and automated feedback in assignments and assessments. Meanwhile, the analyzed body of studies remains limited to language and communication contexts or specific fields such as medical education, with relatively short intervention durations and metacognitive and affective measurements dominated by perceptual data rather than directly observed behavioral indicators.

These limitations open the door to longitudinal research examining the impact of AI audio on cognitive, metacognitive, and affective development over a longer period of time, as well as comparative experiments testing the effectiveness of various configurations, such as audio feedback versus text feedback or voice tutors versus passive audio modules. Research in more diverse contexts including elementary schools beyond language learning, technical vocational education, and teacher training programs is essential to enhance the generalizability of findings. In-depth qualitative studies on how teachers and students negotiate the role of AI in the classroom are also needed to understand shifts in professional identity, perceptions of the source of feedback,

and the dynamics of trust in AI-based systems. Thus, audio AI can be positioned as a strategic component within the learning ecosystem, provided its implementation is grounded in a well-defined pedagogical, technical, and ethical framework.

6. Conclusion

This systematic literature review indicates that AI-based audio makes a positive contribution to learning outcomes, engagement, and metacognitive development when integrated into structured instructional designs, rather than used on an ad hoc basis. Various scenarios ranging from automated audio feedback, voice tutors/assistants, adaptive audio modules, to AI-generated podcasts consistently support conceptual understanding, accelerate the practice-feedback-revision cycle, and expand opportunities for self-directed learning, particularly through increased motivation and engagement. Recent studies also indicate that combining AI-based audio feedback with teacher guidance can foster deeper self-reflection, the development of self-monitoring habits, and the design of more effective learning strategies.

Nevertheless, the available evidence remains concentrated in the contexts of language and communication as well as a few specific fields, with relatively short intervention durations and a predominance of perception-based measurements. These gaps limit the generalizability of findings and underscore the need for longitudinal studies, cross-scenario comparative experiments, and research in more diverse educational contexts. Furthermore, issues regarding voice data privacy, bias, and academic integrity demand a more robust policy and governance framework at the institutional level. Overall, the results of this review confirm that audio AI holds strategic potential to enhance the quality of learning, but these benefits can only be maximized if the technology is positioned as a pedagogical partner within an ecosystem designed with clear pedagogical principles, robust data protection, and a teacher role that remains central.

Acknowledgements

The authors express their sincere gratitude to Universitas Lancang Kuning and Universitas Islam Negeri Sultan Syarif Kasim Riau for their institutional support and resources that facilitated this research.

Funding

This research received no external funding.

Author Contributions (CRediT)

Muhammad Fauzan Hawari contributed to the conceptualization, methodology, investigation, and writing of the original draft. Afifah Mesha Putri contributed to supervision, validation, and review and editing of the manuscript. Siti Syahsudarmi contributed to data curation, formal analysis, and visualization. Akmal Andri Yantama contributed to supervision, resources, and review and editing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest

Data Availability

This article utilizes previously published articles—listed in the references—as the primary material for the literature review.

References

Ariya, P, S Khanchai, K Intawong, and K Puritat. 2025. “Enhancing Textile Heritage Engagement through Generative AI-Based Virtual Assistants in Virtual Reality Museums.” *Computers and Education: X Reality* 7. https://doi.org/10.1016/j.cexr.2025.100112.

Baldrich, Kevin, Carmen Pérez-García, and María Santamarina-Sancho. 2025. “Artificial Intelligence in Academic Literacy: Empirical Evidence on Reading and Writing Practices in Higher Education.” *Frontiers in Education* 10 (November): 1–10. https://doi.org/10.3389/feduc.2025.1701238.

Bland, T. 2025. “Enhancing Medical Student Engagement Through Cinematic Clinical Narratives: Multimodal Generative AI–Based Mixed Methods Study.” *JMIR Medical Education* 11. https://doi.org/10.2196/63865.

Cooper, G, K Ilich, and K Sturrock. 2026. "Using Generative Artificial Intelligence to Reimagine Feedback in Higher Education: A Collaborative Autoethnography." *Assessment and Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2026.2653889.

Crompton, H, A Edmett, N Ichaporia, and D Burke. 2024. "AI and English Language Teaching: Affordances and Challenges." *BRITISH JOURNAL OF EDUCATIONAL TECHNOLOGY* 55 (6): 2503–29. https://doi.org/10.1111/bjet.13460.

Darmawansah, D, G J Hwang, C J Lin, and F Febiyani. 2025. "An Artificial Intelligence-Supported GFCA Learning Model to Enhance L2 Students' Role-Play Performance, English Speaking and Interaction Mindset." *ETR&D-EDUCATIONAL TECHNOLOGY RESEARCH AND DEVELOPMENT* 73 (3): 1451–79. https://doi.org/10.1007/s11423-025-10453-6.

Fuchs, S, A Werth, C Méndez, and J Butcher. 2025. "Leveraging AI-Generated Synthetic Data to Train Natural Language Processing Models for Qualitative Feedback Analysis." *Journal of Engineering Education* 114 (4). https://doi.org/10.1002/jee.70033.

Giannakos, M, R Azevedo, P Brusilovsky, M Cukurova, Y Dimitriadis, D Hernandez-Leo, S Jaervelae, M Mavrikis, and B Rienties. 2025. "The Promise and Challenges of Generative AI in Education." *BEHAVIOUR & INFORMATION TECHNOLOGY* 44 (11): 2518–44. https://doi.org/10.1080/0144929X.2024.2394886.

Grützmacher, L, B Herbert, D Holzberger, A Naumann, M Steffensky, and S Vieluf. 2025. "Critical Examination of the Measurement Quality of Student Outcomes: A Systematic Review." *SCHOOL EFFECTIVENESS AND SCHOOL IMPROVEMENT* 36 (2): 211–41. https://doi.org/10.1080/09243453.2025.2482579.

Hamman-Ortiz, L, C Doughtery, Z F Tian, D Palmer, and L Poza. 2025. "Translanguaging at School: A Systematic Review of US PK-12 Translanguaging Research." *SYSTEM* 129. https://doi.org/10.1016/j.system.2025.103594.

Higgins, J, and S Green, eds. 2008. "Cochrane Handbook for Systematic Reviews of Interventions." West Sussex: John Wiley & Sons.

Hong, Jon-Chao, and Tzu-Yu Tai. 2025. "Exploring the Role of Internet Self-Efficacy, Perceived Enjoyment, and Anxiety on Learning Outcome in Intelligent Personal Assistant-Based EFL Learning." *Innovation in Language Learning and Teaching* 19 (2): 136–54. https://doi.org/10.1080/17501229.2024.2347884.

Huang, J. 2024. "Enhancing EFL Speaking Feedback with ChatGPT's Voice Prompts." *International Journal of TESOL Studies* 6 (3): 4–13. https://doi.org/10.58304/ijts.20240302.

Kasneci, Enkelejda, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, et al. 2023. "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education." *Learning and Individual Differences* 103:102274. https://doi.org/https://doi.org/10.1016/j.lindif.2023.102274.

Kerr, Sharon. 2017. "Exploring the Impact of Artificial Intelligence on Teaching and Learning in Higher Education." https://doi.org/10.1186/s41039-017-0062-8.

Kitchenham, Barbara, Rialette Pretorius, David Budgen, O Pearl Brereton, Mark Turner, Mahmood Niazi, and Stephen Linkman. 2010. "Systematic Literature Reviews in Software Engineering – A Tertiary Study." *Information and Software Technology* 52 (8): 792–805. https://doi.org/https://doi.org/10.1016/j.infsof.2010.03.006.

Koolivand, H, M M Shooreshi, R Safari-Faramani, M Borji, M S Mansoory, H Moradpoor, M Bahrami, and S M Azizi. 2024. "Comparison of the Effectiveness of Virtual Reality-Based Education and Conventional Teaching Methods in Dental Education: A Systematic Review." *BMC MEDICAL EDUCATION* 24 (1). https://doi.org/10.1186/s12909-023-04954-2.

Krsti, Lazar, Veljko Aleksi, Marija Krsti, Vocational Studies, Southern Serbia, and Technical Sciences. 2022. "Artificial Intelligence in Education : A Review," no. September. https://doi.org/10.46793/TIE22.223K.

Li, W, X Cui, P Manoharan, L Dai, K Liu, and L Huang. 2025. "AI-Assisted Feedback and Reflection in Vocal Music Training: Effects on Metacognition and Singing Performance." *Frontiers in Psychology* 16. https://doi.org/10.3389/fpsyg.2025.1598867.

Lin, Haozhuo, and Qiu Chen. 2024. "Artificial Intelligence (AI) -Integrated Educational Applications and College Students' Creativity and Academic Emotions." https://doi.org/10.1186/s40359-024-01979-0.

Lin, Y, X Wang, H Ma, L Wang, F Hao, and Z Cai. 2023. "An Efficient Approach to Sharing Edge Knowledge in 5G-Enabled Industrial Internet of Things." *IEEE Transactions on Industrial Informatics* 19 (1): 930–39. https://doi.org/10.1109/TII.2022.3170470.

Liu, Y, S Ma, and Y Chen. 2024. "The Impacts of Learning Motivation, Emotional Engagement

and Psychological Capital on Academic Performance in a Blended Learning University Course." *FRONTIERS IN PSYCHOLOGY* 15. https://doi.org/10.3389/fpsyg.2024.1357936.

Michou, I, A Fouras, D Chrysanthakopoulou, M Theodoritsi, S Mariettou, S Stellatou, and C Koutsojannis. 2026. "AI-Powered Physiotherapy: Evaluating LLMs Against Students in Clinical Rehabilitation Scenarios." *Applied Sciences (Switzerland)* 16 (3). https://doi.org/10.3390/app16031165.

Paradis, J. 2023. "Sources of Individual Differences in the Dual Language Development of Heritage Bilinguals." *JOURNAL OF CHILD LANGUAGE* 50 (4): 793–817. https://doi.org/10.1017/S0305000922000708.

Praetorius, A K, A Jentsch, J Luoto, S D Keller, and B Fauth. 2025. "Context in Research on Teaching Quality: Bringing a Complex Idea into the Spotlight." *SCHOOL EFFECTIVENESS AND SCHOOL IMPROVEMENT* 36 (2): 263–96. https://doi.org/10.1080/09243453.2025.2482577.

Ratten, V. 2023. "The Post COVID-19 Pandemic Era: Changes in Teaching and Learning Methods for Management Educators." *INTERNATIONAL JOURNAL OF MANAGEMENT EDUCATION* 21 (2). https://doi.org/10.1016/j.ijme.2023.100777.

Rienties, B, J Domingue, S Duttaroy, C Herodotou, F Tessarolo, and D Whitelock. 2025. "What Distance Learning Students Want from an AI Digital Assistant." *Distance Education* 46 (2): 173–89. https://doi.org/10.1080/01587919.2024.2338717.

Shi, H W, and V Aryadoust. 2024. "A Systematic Review of AI-Based Automated Written Feedback Research." *RECALL* 36 (2): 187–209. https://doi.org/10.1017/S0958344023000265.

Strielkowski, Wadim, Veronika Grebennikova, Alexander Lisovskiy, Guzalbegim Rakhimova, and Tatiana Vasileva. 2025. "AI-Driven Adaptive Learning for Sustainable Educational Transformation." *Sustainable Development* 33 (2): 1921–47. https://doi.org/10.1002/sd.3221.

Sungkur, R K, and N Shibdeen. 2025. "Pronunciation Trainer for Second Language Learning Using Generative AI." *International Journal of Educational Technology in Higher Education* 22 (1). https://doi.org/10.1186/s41239-025-00561-x.

Tomchakovskyi, O, Y Tomchakovska, L Strochenko, L Shvelidze, and S Yeryomenko. 2026. "Multimodal AI Tools as Mediators of Linguistic Development in EFL Academic Writing." *Arab World English Journal* 2026-January (Special Issue 3): 390–403. https://doi.org/10.24093/awej/A3.25.

Uppalapati, P J, M Dabbiru, and V R Kasukurthi. 2025. "AI-Driven Mock Interview Assessment: Leveraging Generative Language Models for Automated Evaluation." *International Journal of Machine Learning and Cybernetics* 16 (12): 10057–79. https://doi.org/10.1007/s13042-025-02529-9.

Wang, Shan, Fang Wang, Zhen Zhu, Jingxuan Wang, Tam Tran, and Zhao Du. 2024. "Artificial Intelligence in Education : A Systematic Literature Review." *Expert Systems With Applications* 252 (PA): 124167. https://doi.org/10.1016/j.eswa.2024.124167.

Yang, F F, L F Ren, and C Gu. 2022. "A Study of College Students' Intention to Use Metaverse Technology for Basketball Learning Based on UTAUT2." *HELIYON* 8 (9). https://doi.org/10.1016/j.heliyon.2022.e10562.

Yang, T C, Y C Hsu, and J Y Wu. 2025. "The Effectiveness of ChatGPT in Assisting High School Students in Programming Learning: Evidence from a Quasi-Experimental Research." *INTERACTIVE LEARNING ENVIRONMENTS* 33 (6): 3726–43. https://doi.org/10.1080/10494820.2025.2450659.

Zerkouk, Meriem, Miloud Mihoubi, and Belkacem Chikhaoui. 2025. "A Comprehensive Review of AI-Based Intelligent Tutoring Systems: Applications and Challenges." http://arxiv.org/abs/2507.18882.

Zhang, W, W Zhang, T Daım, and H Yalcin. 2025. "AI Challenges Conventional Knowledge Management: Light the Way for Reframing SECI Model and Ba Theory." *Journal of Knowledge Management* 29 (5): 1618–54. https://doi.org/10.1108/JKM-02-2024-0203.