Assessing Speaking Skills Through E-Assessment: A Systematic Review for Advancing Language Evaluation in Indonesian Language Education

Achmad Rizal Taufiqi[1*], Syamsul Sodiq[2], Miftachul Amri[3]
[1*] Universitas Negeri Surabaya, Surabaya, Indonesia
[2] Universitas Negeri Surabaya, Surabaya, Indonesia
[3] Universitas Negeri Surabaya, Surabaya, Indonesia

Keywords:
Artificial Intelligence,
Technology,
E-Assessment,
Indonesia Speaking Skills,
Indonesia Learning

ABSTRACT

Speaking skill assessment is a vital component of Indonesian language learning; however, its implementation often faces persistent challenges, particularly in capturing the nuances of students' spoken expressions and vocabulary use. This study proposes an innovative approach by integrating Artificial Intelligence (AI) as a supportive tool for speaking assessment. Drawing on speech recognition technology and natural language processing (NLP), AI enables the automatic recording, transcription, and analysis of learners' oral performances. Through a literature review of Scopus-indexed sources, this conceptual research examines the application of e-assessment in speaking skills, identifies the technologies and platforms involved, and evaluates both advantages and challenges. Findings suggest that AI-assisted assessment offers significant benefits, including improved objectivity, reduced teacher workload, enhanced time efficiency, and the provision of timely, in-depth feedback. Nonetheless, limitations remain, such as the need for dialect-sensitive models, infrastructure readiness, and adherence to ethical data practices. This study concludes with recommendations for future development, emphasizing contextually relevant AI systems, teacher training, and ethical frameworks to ensure fair, accurate, and culturally responsive speaking assessment in the Indonesian language context.

INTRODUCTION

The rapid advancement of technology has significantly impacted various aspects of life, including the teaching and learning of the Indonesian language. Speaking skills in Indonesian language education represent an expressive ability that requires students to articulate ideas orally with appropriate structure and meaning. However, in practice, assessing this skill is not always straightforward. Speaking is a crucial competency, as it serves as the foundation for effective communication, a fundamental need in today's world (Akhter et al., 2020). This importance is further emphasized by Teo, as cited in Hong et al. (2022), who asserts that thinking and cognition occur through verbal communication, shaped by how others' voices interact with what we say, write, and think. This highlights that speaking skills are acquired through interactive learning, by engaging with the thoughts, language, and interactive behaviors of others. In the context of teaching, performance-based assessment is a form of authentic evaluation that aims to measure specific abilities or competencies students are expected to demonstrate in the learning process (Agustina et al., 2022). Therefore, as technology evolves, there is a pressing need to enhance assessment practices, particularly in evaluating speaking performance.

In general, the teaching of speaking skills in Indonesian language classes is still dominated by traditional approaches with limited innovation (Gatra, 2018). Halidjah (2012) also argues that assessments of speaking skills have traditionally relied on fixed benchmarks and normative references, resulting in learning concepts that may be less egalitarian for students. Moreover, many students struggle to produce accurate speech sounds, leading them to remain silent and refrain from expressing their thoughts, often appearing hesitant or anxious (Rohaini, 2021). This condition, however, cannot be separated from the teacher's own conceptual understanding, which significantly influences the development of assessment literacy (Coombe et al., 2020).

Understanding valid and consistent definitions and methods of measuring learning outcomes remains a subject of ongoing discussion (Sintadewi et al., 2017). Assessment plays a crucial role in the learning process as it allows teachers to determine the extent of student development. The information gathered from such assessments serves as a basis for designing future instructional strategies (Nurhamidah, 2021). Teachers often face difficulties in capturing students' spoken utterances in detail, especially in large classroom settings or when time is limited. In speaking instruction, what matters is not only the clarity of articulation but also gestures and nonverbal cues. Moreover, some evaluation methods used to measure the quality of a program inherently involve a degree of subjectivity (Wahyono, 2017).

The advancement of Artificial Intelligence (AI) presents new opportunities in the field of education, including in the development of assessment systems. Technologies such as speech recognition and Natural Language Processing (NLP) have proven capable of identifying and processing human speech in various contexts (Jurafsky & Martin, 2025). In Indonesia, research on speech recognition has been conducted by Iqbal (2024), although it was developed in English. This presents a promising opportunity for adaptation into the Indonesian language. Within the context of Indonesian language instruction, the integration of AI holds potential to assist teachers in capturing students' spoken expressions in real time, archiving speech data, and automatically generating linguistic analysis.

Unfortunately, to date, there has been limited research specifically exploring the application of AI in speaking skills assessment at the secondary school level, particularly in Indonesian language subjects. This gap highlights an urgent area for further investigation, given the need for more accurate, fair, and in-depth assessment methods to foster students' speaking competence in the digital era.

RESEARCH METHOD

This article employs a qualitative approach using a literature study method. The research method follows a Systematic Literature Review (SLR). A systematic review aims to synthesize evidence in order to answer predefined research questions (Aytac et al., 2025). According to Chadler et al. in Gunnell et al. (2022), a systematic review uses well-established and explicit methods to provide a comprehensive synthesis of knowledge on a particular topic or field.

This study consists of seven stages: (1) planning, (2) literature search, (3) study selection, (4) data extraction, (5) data synthesis, (6) interpretation, and (7) reporting (Aytac et al., 2025). Artificial intelligence (AI) has been widely applied in education for personalized learning, automated grading, and adaptive instruction (Luckin & Holmes, 2016). In the field of language learning, tools such as Google Speech-to-Text, Amazon Transcribe, and applications like ELSA Speak have demonstrated capabilities in recognizing pronunciation, accents, and providing feedback on articulation and sentence structure (Derwing & Munro, 2021). However, the potential of AI to detect specific terms or diction spoken by students remains largely unexplored in the context of Indonesian language education.

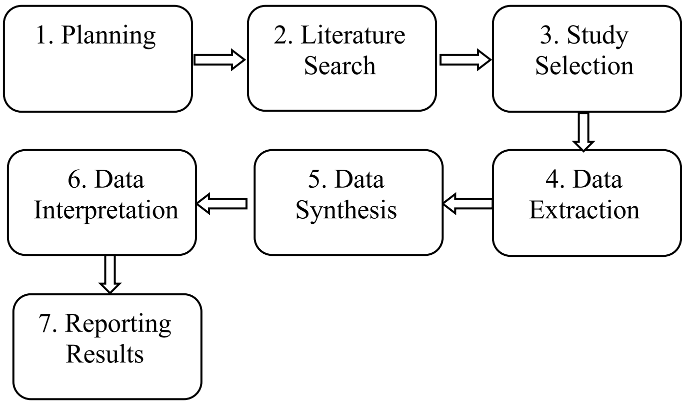


Diagram 1. Research Stages (Aytac et al., 2025)

RESULTS AND DISCUSSION

Planning

The advancement of digital technology has driven significant transformation in the field of education, including in the area of learning assessment. One emerging innovation is electronic assessment (e-assessment), which refers to technology-based evaluation methods. In the context of Indonesian language instruction, particularly speaking skills, e-assessment has become an increasingly relevant topic of inquiry. This is due to the fact that speaking skills require assessment strategies that are not only accurate, but also adaptable to technological developments and the needs of distance learning.

Although a number of studies have addressed this area, there has yet to be a systematic review that specifically maps the approaches, tools, challenges, and advantages of e-assessment in the speaking skill domain of Indonesian language education. Therefore, this study was designed to provide a scientific synthesis of existing research findings through the Systematic Literature Review (SLR) method. The goal is to identify, analyze, and synthesize various approaches, methods, and findings related to the implementation of e-assessment for speaking skills in Indonesian language learning.

From this framework, the following research questions (RQs) are formulated:

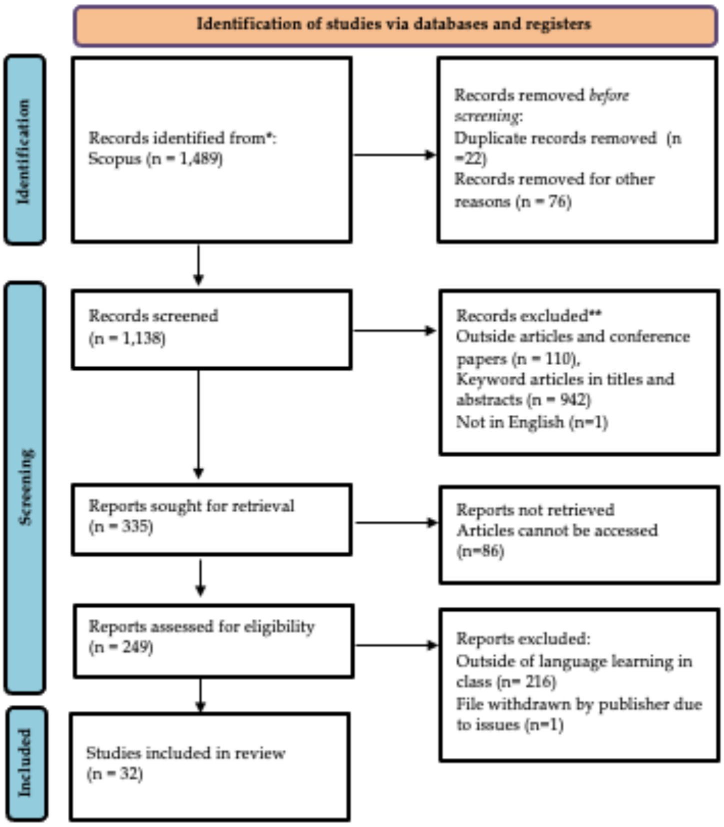
1. RQ1: How is the e-assessment approach applied to speaking skills?
2. RQ2: What technologies and platforms are used in e-assessment of speaking skills?
3. RQ3: What are the advantages and challenges of implementing e-assessment in Indonesian language speaking instruction?
4. RQ4: What are the recommendations for the future development of speaking e-assessment?

Literature Search

Data were collected by searching journals and scientific publications indexed in Scopus, using the keywords "AI in oral language assessment" and "speech recognition in education" within the title and abstract fields. The initial search yielded 1,489 documents (n = 1,489). The search was limited to peer-reviewed journal articles published between 2020 and 2025, written in English, and relevant to the application of artificial intelligence in language education.

The selection process involved screening the titles and abstracts to ensure their relevance to the research questions. Articles discussing the application of computational and AI-based technologies, such as speech recognition, automated scoring, and natural language processing (NLP), in the context of speaking skill assessment were retained. Studies unrelated to educational settings or those that did not focus on speaking skills were excluded.

The analysis was conducted using a thematic approach, enabling the identification and categorization of patterns related to the roles, potentials, and challenges of using AI in speaking assessment. The findings were synthesized to inform the discussion and draw implications for Indonesian language education.


Identification of studies via databases and registers
Identification
Records identified from*:
Scopus (n = 1,489)
Records removed *before screening*:
Duplicate records removed (n =22)
Records removed for other reasons (n = 76)
Screening
Records screened
(n = 1,138)
Records excluded**
Outside articles and conference papers (n = 110),
Keyword articles in titles and abstracts (n = 942)
Not in English (n=1)
Reports sought for retrieval
(n = 335)
Reports not retrieved
Articles cannot be accessed
(n=86)
Reports assessed for eligibility
(n = 249)
Reports excluded:
Outside of language learning in class (n= 216)
File withdrawn by publisher due to issues (n=1)
Included
Studies included in review
(n = 32)

Eliminated for other reasons (e.g., incomplete metadata or retrieval errors). As a result, 1,388 The literature search was conducted using the Scopus database, which initially generated 1,489 articles. In the identification stage, 22 duplicate records were removed, followed by an additional 76 articles remained for further screening.

In the screening phase, articles were excluded based on predefined eligibility criteria. The reasons for exclusion included: non-journal publications such as popular articles and conference proceedings (n = 110), lack of relevance based on keywords found in the title and abstract (n = 942), and one article written in a language other than English (n = 1). This screening resulted in 335 articles being retained for the next phase.

Among these 335 records, 86 articles were excluded due to inaccessibility (e.g., paywall issues or removed by publisher). The remaining 249 articles underwent full-text evaluation, during which 216 articles were excluded for not being relevant to the context of language learning in classroom settings, and 1 article was withdrawn by the journal due to failing to meet publication standards.

Finally, 32 articles were found to fully meet all inclusion criteria and were selected for in-depth analysis in this study. A total of 32 article titles and author names were documented based on the filtered results using the PRISMA flow methodology.

No	Authors	Title
1	Aizat K.; Mohamed O.; Orken M.; Ainur A.; Zhumazhanov B. (2020)	Identification And Authentication Of User Voice Using Dnn Features And I-Vector
2	Che Dalim C.S.; Sunar M.S.; Dey A.; Billinghurst M. (2020)	Using Augmented Reality With Speech Input For Non-Native Children's Language Learning
3	Chen H.; Mei K. (2024)	Exploring A Mobile Technology-Driven Model For Intercultural Communication Education
4	Detey S.; Fontan L.; Le Coz M.; Jmel S. (2020)	Computer-Assisted Assessment Of Phonetic Fluency In A Second Language: A Longitudinal Study Of Japanese Learners Of French
5	Geng L. (2021)	Evaluation Model Of College English Multimedia Teaching Effect Based On Deep Convolutional Neural Networks
6	Henkel O.; Horne-Robinson H.; Hills L.; Roberts B.; McGrane J. (2025)	Supporting Literacy Assessment In West Africa: Using State-Of-The-Art Speech Models To Assess Oral Reading Fluency
7	Hong Z.-W.; Tsai M.-H.M.; Ku C.S.; Cheng W.K.; Chen J.-T.; Lin J.-M. (2024)	Utilizing Robot-Tutoring Approach In Oral Reading To Improve Taiwanese Efl Students' English Pronunciation
8	Ji Y. (2024)	A Deep Learning Assessment Model For Oral Learning In English Online Education
9	Lee Y.; Cho J. (2020)	Design Of A Foreign Language Conversation Learning System Using Machine Learning
10	Li W.; Mohamad M. (2023)	An Efficient Probabilistic Deep Learning Model For The Oral Proficiency Assessment Of Student Speech Recognition And Classification
11	Liu Y.-F.; Luthfi M.I.; Hwang W.-Y. (2024)	Developing An Ai-Enhanced Video Drama-Making Learning System To Support Efl Learners In Authentic Contexts
12	Turniyazova S.; Ne'matova Y.; Turdikulov S.; Makhkamova N.; Sapaev I.; Mukhitdinova B.; Abdullaev R.; Odilova G. (2025)	Wireless Mobile Network With Transfer Learning Algorithm For Multilingual Education And Historical Research

13	Orosoo M.; Raash N.; Treve M.; M. Lahza H.F.; Alshammry N.; Ramesh J.V.N.; Rengarajan M. (2025)	Transforming English Language Learning: Advanced Speech Recognition With Mlp-Lstm For Personalized Education
14	Penning de Vries B.W.F.; Cucchiarini C.; Strik H.; van Hout R. (2020)	Spoken Grammar Practice In Call: The Effect Of Corrective Feedback And Education Level In Adult L2 Learning
15	Qian Y.; Ubale R.; Lange P.; Evanini K.; Ramanarayanan V.; Soong F.K. (2020)	Spoken Language Understanding Of Human-Machine Conversations For Language Learning Applications
16	Sun H. (2023)	Advancing English Language Learning With Biotechnology: Utilizing Domain Adaptive Convolutional Neural Networks
17	Tolba R.M.; Elarif T.; Taha Z.; Hammady R. (2024)	Interactive Augmented Reality System For Learning Phonetics Using Artificial Intelligence
18	Uppalapati P.J.; Dabbiru M.; Kasukurthi V.R. (2025)	Ai-Driven Mock Interview Assessment: Leveraging Generative Language Models For Automated Evaluation
19	Wang H.; Sharma A.; Shabaz M. (2022)	Research On Digital Media Animation Control Technology Based On Recurrent Neural Network Using Speech Technology
20	Wang L. (2022)	A Machine Learning Assessment System For Spoken English Based On Linear Predictive Coding
21	Xie Y. (2023)	Application Of Speech Recognition Technology Based On Machine Learning For Network Oral English Teaching System
22	Revenor R Y. (2024)	On The Adequacy Of Elsa Speak In Formal Education: A Survey Of Teacher-Users
23	Zhu M. (2022)	The Application Of Intelligent Speech Analysis Technology In The Spoken English Language Learning Model
24	Zou B.; Lyu Q.; Han Y.; Li Z.; Zhang W. (2023)	Exploring Students' Acceptance Of An Artificial Intelligence Speech Evaluation Program For Efl Speaking Practice: An Application Of The Integrated Model Of Technology Acceptance
25	Dang R. (2024)	English-Speaking Learning Strategies In University Based On Artificial Intelligence
26	Filighera A.; Bongard L.; Steuer T.; Tregel T. (2022)	Towards A Vocalization Feedback Pipeline For Language Learners
27	Gonzales K.; Maranan J.; Macale N.; Renovalles E.J.; Palafox N.A.; Santelices F.P.; Mendoza J.M. (2024)	Bk3at: Bangsamoro K-3 Children's Speech Corpus For Developing Assessment Tools In The Bangsamoro Languages
28	Guney C.; Akinci O.; Camoglu K. (2021)	Artificial Learning-Based Proctoring Solution For Remote Online Assessments: "Vproctor"
29	Hou R.; Fütterer T.; Bühler B.; Bozkir E.; Gerjets P.; Trautwein U.; Kasneci E. (2024)	Automated Assessment Of Encouragement And Warmth In Classrooms Leveraging Multimodal Emotional Features And Chatgpt
30	Nasriddinov F.; Kocielnik R.; Gupta A.; Yang C.; Wong E.; Anandkumar A.; Hung A.J. (2025)	Automating Feedback Analysis In Surgical Training: Detection, Categorization, And Assessment
31	Pandey D.; Subedi A.; Mishra D. (2022)	Improving Language Skills And Encouraging Reading Habits In Primary Education: A Pilot Study Using Nao Robot
32	Sloan J.; Maguire D.; Carson-Berndsen J. (2021)	Emotional Response Language Education For Mobile Devices

Table 1. Names and Titles of Research Studies

The line graph presented below illustrates the dynamics of publication output over the past six years. Based on the visual data, there is a clear fluctuation that reflects the variability in scientific productivity from year to year. In 2020, a total of 7 publications were recorded, followed by a significant decline to 3 publications in 2021. This decline may be attributed to internal or external factors such as a shift in research priorities, operational challenges, or transitions in academic policy. The year 2022 showed signs of recovery, with the number of publications increasing to 6, before slightly decreasing again to 5 in 2023. The peak of productivity was reached in 2024 with 10 publications, which could reflect intensified research activity or the successful implementation of collaborative academic strategies. However, this figure dropped again to 4 publications in 2025. Overall, the graph demonstrates that despite occasional surges in productivity, the general trend reveals an unstable pattern likely influenced by changing institutional policies, shifting research agendas, and the integration of AI technologies in the assessment of speaking skills.

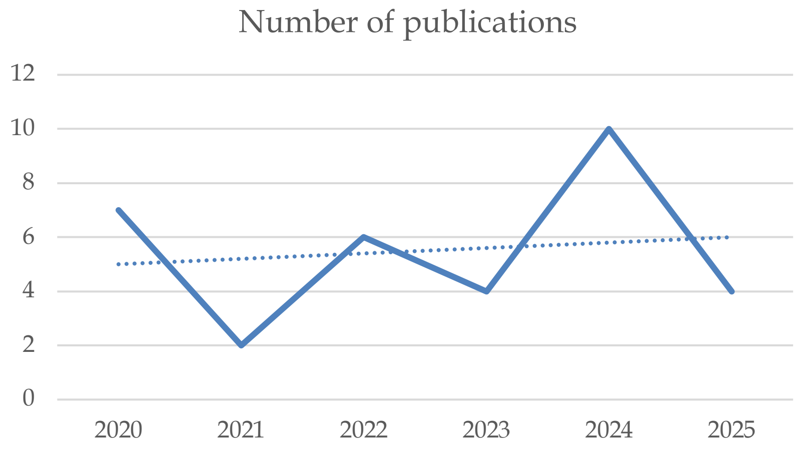


Diagram 2. Researcher's processing chart

Overall, the trendline indicates a positive trajectory, suggesting an increase in the number of publications over time, despite an irregular or non-linear annual pattern. The observed year-to-year fluctuations reflect the influence of both internal and external factors on scientific productivity within this research domain. Nonetheless, the upward trend as a whole suggests growing interest and development in AI-based speaking assessment, particularly within the context of language education. Therefore, while annual inconsistencies persist, the data offers an encouraging outlook for continued scholarly exploration and future research opportunities.

No	Authors	Cited by
1	Che Dalim C.S.; Sunar M.S.; Dey A.; Billinghurst M.	96
2	Geng L.	31
3	Wang H.; Sharma A.; Shabaz M.	19

4	Aizat K.; Mohamed O.; Orken M.; Ainur A.; Zhumazhanov B.	16
5	Zou B.; Lyu Q.; Han Y.; Li Z.; Zhang W.	16
6	Penning de Vries B.W.F.; Cucchiarini C.; Strik H.; van Hout R.	14
7	Detey S.; Fontan L.; Le Coz M.; Jmel S.	11
8	Qian Y.; Ubale R.; Lange P.; Evanini K.; Ramanarayanan V.; Soong F.K.	9
9	Pandey D.; Subedi A.; Mishra D.	8
10	Guney C.; Akinci O.; Camoglu K.	4
11	Tolba R.M.; Elarif T.; Taha Z.; Hammady R.	3
12	Hou R.; Fütterer T.; Bühler B.; Bozkir E.; Gerjets P.; Trautwein U.; Kasneci E.	3
13	Hong Z.-W.; Tsai M.-H.M.; Ku C.S.; Cheng W.K.; Chen J.-T.; Lin J.-M.	2
14	Lee Y.; Cho J.	2
15	Li W.; Mohamad M.	2
16	Orosoo M.; Raash N.; Treve M.; M. Lahza H.F.; Alshammry N.; Ramesh J.V.N.; Rengarajan M.	2
17	Wang L.	2
18	Zhu M.	2
19	Uppalapati P.J.; Dabbiru M.; Kasukurthi V.R.	1
20	Sloan J.; Maguire D.; Carson-Berndsen J.	1
21	Chen H.; Mei K.	0
22	Henkel O.; Horne-Robinson H.; Hills L.; Roberts B.; McGrane J.	0
23	Ji Y.	0
24	Liu Y.-F.; Luthfi M.I.; Hwang W.-Y.	0
25	Mukhitdinova B.; Abdullaev R.; Odilova G.; Turniyazova S.; Ne'matova Y.; Turdikulov S.; Makhkamova N.; Sapaev I.	0
26	Sun H.	0
27	Xie Y.	0
28	Yamamoto Ravenor R.	0
29	Dang R.	0
30	Filighera A.; Bongard L.; Steuer T.; Tregel T.	0
31	Gonzales K.; Maranan J.; Macale N.; Renovalles E.J.; Palafox N.A.; Santelices F.P.; Mendoza J.M.	0
32	Nasriddinov F.; Kocielnik R.; Gupta A.; Yang C.; Wong E.; Anandkumar A.; Hung A.J.	0

Table 2. Number of research citations

The citation data presented above reflects the academic impact and reach of several researchers involved in studies related to artificial intelligence (AI)-based learning technologies, speech recognition, and speaking assessment. Citation count serves as one of the indicators of the scholarly influence of a publication or researcher.

The authors with the highest citation count are Che Dalim C.S., Sunar M.S., Dey A., and Billinghurst M., with a total of 96 citations, indicating that their work holds significant influence and is frequently referenced by other scholars. This is followed by Geng L. with

31 citations, as well as Wang H., Sharma A., and Shabaz M. with 19 citations, demonstrating a relatively high level of visibility within the academic community.

Other researchers, such as Aizat K. and colleagues, along with Zou B. and collaborators, received 16 citations each, marking a notable level of scholarly relevance. Meanwhile, authors like Detey S. and team (11 citations), Qian Y. and team (9 citations), and Pandey D. and colleagues (8 citations) show a presence in the field, though on a smaller scale.

On the other hand, some authors, such as Chen H. and Mei K., Henkel O. and colleagues, among others, have yet to receive citations. This may be due to various factors, including the recency of publication, limited accessibility, or the relatively underexplored nature of the topic within the broader academic network.

Overall, this citation distribution pattern illustrates the dominance of a few highly influential researchers, while the majority exhibit moderate to low levels of academic impact. It also highlights the dynamics of knowledge dissemination in the field and underscores the importance of visibility and thematic relevance in determining the extent to which scholarly work is cited within the global academic community.

RQ1: How Is the E-Assessment Approach Applied to Speaking Skills?

The e-assessment approach in the context of speaking skills refers to the use of digital technologies to evaluate individuals' oral communication abilities. This method utilizes various computer-based tools and applications, such as speech recognition software, audio-video recordings, online learning platforms, and automated scoring systems, to assess critical aspects of oral performance, including fluency, pronunciation, sentence structure, and clarity of speech.

E-assessment enables a more objective, efficient, and flexible evaluation process by leveraging digital platforms and tools. In several studies, advanced technologies such as ELSA-Speak [22], GPT-based language models [11], [18], [30], and deep learning frameworks like LDiDL (Language Development using Deep Learning) [10] have demonstrated effectiveness in enhancing speaking competencies.

Experimental research [5] has shown that e-assessment algorithms can assist learners in identifying inconsistencies between their pronunciation and standard pronunciation models, offering corrective feedback that leads to improved oral English performance. However, several technical challenges must be addressed, including complex phonetic articulation, large volumes of data in speech signals, high-dimensional voice function parameters, and substantial computational demands for accurate speech recognition and evaluation [6].

In practice, speaking skill e-assessment involves various task formats, such as monologues submitted via voice recordings, simulation-based interactive dialogues, and the use of chatbots or AI-driven applications to train and assess fluency and accuracy. Research conducted in digital drama contexts [11] has shown promising results, though further refinement is needed. Additionally, an empirical study [28] affirms that the application of AI can significantly enhance students' spoken expression through personalized and semantically-oriented learning strategies.

The primary advantage of this approach lies in its ability to accommodate diverse learning settings, including remote education, and support both formative and summative assessment through real-time or delayed feedback, as well as the documentation of learners' spoken outputs for reflection and continuous improvement. At the same time, successful implementation requires careful attention to assessment validity, technological readiness, and the digital competence of both teachers and students.

RQ2: What Technologies and Platforms Are Used in E-Assessment of Speaking Skills?

This research draws upon the analysis of data collected from a set of relevant peer-reviewed scientific articles. The findings provide an overview of the diverse technologies employed in the development of speaking skill instruction, particularly within the context of artificial intelligence (AI) integration and digitally mediated systems. Each technology identified demonstrates a distinct role and contribution to the effectiveness of speaking skill development, covering areas such as pronunciation training, automated assessment, speech recognition, as well as enhancement of learner motivation and interaction.

The following is a summary of the key technologies utilized in the reviewed literature:

Deep Neural Network (DNN), as identified in [1], is utilized for advanced speech processing and recognition, employing brain-inspired architectures to improve pronunciation accuracy in language learning.

TeachAR, in [2], demonstrates the application of Augmented Reality to enhance interactive visualization in oral language instruction.

Mean Opinion Score (MOS) and Mel Spectrum Distortion (MSD), as used in [3], serve as metrics for evaluating audio quality in speech synthesis and automatic speech recognition systems.

Computer-Assisted Pronunciation Training (CAPT) and Automatic Speech Recognition (ASR) in [4] provide real-time feedback on learners' pronunciation.

A combination of Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) in [5] represents classical yet still relevant techniques for evaluating pronunciation alignment with standard models.

ASR technologies consistently emerge as dominant tools in [6], [7], [12], [26], and [30], reinforcing their central role in streamlining the evaluation of speaking skills.

Bidirectional Encoder Representations from Transformers (BERT) and BiLSTM models in [8] illustrate the integration of transformer-based contextual understanding with bidirectional long-term memory networks for deeper speech comprehension.

Word2Vec-based learning systems in [9] support semantic processing of student utterances.

The integration of Support Vector Machines (SVM) and Recurrent Neural Networks (RNN) in [10] aids in pronunciation classification and enhancement of automated scoring systems.

Text-to-Speech (TTS) combined with the Technology Acceptance Model (TAM) in [11] evaluates user acceptance of technologies while generating spoken feedback.

The MLP-LSTM model in [14] represents a neural architecture for sequential speech analysis.

The fusion of ASR and LTM-RRN in [15] seeks to balance long- and short-term memory components in speech processing.

Convolutional Neural Networks (CNN) in [16] are employed for acoustic feature extraction from student speech data.

Augmented Reality (AR) and MRI Recording technologies in [17] offer both medical and visual approaches for observing vocal production.

BLSTM-RNN models in [19] incorporate bidirectionality and memory retention for robust speech processing.

Integration of ILPC and CNN in [20] combines linear predictive coding with image-based pattern recognition in speech signals.

Mel-Frequency Cepstral Coefficients (MFCC) in [21] remain a foundational method for acoustic feature extraction.

Additive White Gaussian Noise (AWGN) in [23] is applied to assess system robustness in noisy environments.

The convergence of IMTA and CALL in [24] reflects interactive language instruction supported by computer-assisted learning technologies.

Database-Driven CALL (DB-CALL) in [25] facilitates large-scale storage and analysis of student speech responses.

Wav2Vec 2.0, as shown in [27], introduces a self-supervised learning model for efficient speech recognition.

vProctor in [28] demonstrates voice-based automated proctoring for oral examinations.

NAO, in [31], an interactive robot, is utilized to provide oral responses and feedback in speaking skill training.

The Emotional Response Language Education (ERLE) model in [32] integrates emotional feedback with human-machine interaction for spoken language learning.

If classified, the technologies in assessment can be organized as follows:

1. **Artificial Intelligence (AI) & Machine Learning-Based Technologies**

This category focuses on machine learning models, neural networks, and Natural Language Processing (NLP) techniques. Technologies in this group include deep learning architectures such as Deep Neural Networks (DNN), Bidirectional Long Short-Term Memory (BiLSTM), and Convolutional Neural Networks (CNN), as well as semantic representations like Word2Vec and BERT. These models are applied to a variety of NLP tasks, including text classification, phonetic feature extraction, and speech synthesis. The presence of models such as Wav2Vec 2.0 illustrates the evolution of technology toward more efficient and accurate speech signal processing through pretrained representation learning.

Technology	**Article**
Deep Neural Network (DNN)	[1]
BERT & BiLSTM	[8]
Word2Vec	[9]
SVM & RNN	[10]
MLP-LSTM	[14]
BLSTM-RNN	[19]
CNN	[16]
ILPC + CNN	[20]
Wav2Vec 2.0	[27]

2. **Automatic Speech Recognition (ASR) and Computer-Assisted Pronunciation Training (CAPT) Technologies**
   This category refers to automatic speech recognition technologies that enable computer systems to comprehend and process human speech input. ASR functions through a combination of acoustic feature extraction, such as Mel-Frequency Cepstral Coefficients (MFCC), and statistical models like Hidden Markov Models (HMM) or Dynamic Time Warping (DTW). In the context of language learning, ASR is integrated with CAPT to provide automatic feedback on pronunciation, allowing for adaptive, efficient, and large-scale pronunciation training. These technologies support the automation of speech recognition and pronunciation practice in educational settings.

Technology	**Article**
ASR	[4], [6], [7], [12], [15], [26], [30]
CAPT	[4]
DTW & HMM	[5]
MFCC	[21]
TTS	[11]

3. **Augmented Reality (AR) and Interactive Multimedia Technologies**
   This category refers to the use of Augmented Reality (AR), platforms such as TeachAR, and other interactive visual systems to enhance the language learning experience. AR provides a contextual and multimodal learning environment, enabling learners to engage with language materials through visual and kinesthetic interaction. The integration of AR with tools such as MRI recording, for instance, opens up possibilities for analyzing articulatory movements in pronunciation. Meanwhile, systems like vProctor illustrate the direction of AR integration with automated evaluation systems for remote monitoring and assessment.

Technology	**Article**
TeachAR	[2]
AR & MRI Recording	[17]
vProctor	[28]

4. **Evaluation and Measurement of Speech System Quality**
Technologies in this category are employed to assess the acoustic and perceptual quality of speech processing systems, such as Text-to-Speech (TTS) and Automatic Speech Recognition (ASR). The Mean Opinion Score (MOS) evaluates perceptual quality based on human judgments, while Mel Spectrum Distortion (MSD) and Additive White Gaussian Noise (AWGN) serve as quantitative parameters to assess signal distortion and noise levels. Such evaluations are essential for determining the performance of TTS and ASR systems to ensure their alignment with human perception and usability standards.

Technology	**Article**
Mean Opinion Score (MOS) & Mel Spectrum Distortion (MSD)	[3]
Additive White Gaussian Noise (AWGN)	[23]

5. **Computer-Assisted Language Learning (CALL) and Its Variants**
This category encompasses various forms of computer-assisted language learning technologies. CALL and its derivatives, such as the Integrated Multimedia Teaching Approach (IMTA) and Database-Driven CALL (DB-CALL), facilitate language learning through data-driven, multimedia-based, and interactive approaches. These systems support flexible, autonomous, and personalized learning while also enabling the storage and analysis of learner behavior through integrated databases and performance tracking mechanisms.

Technology	**Article**
IMTA dan CALL	[24]
DB-CALL	[25]

6. **Psychological Models and Human–Machine Interaction**
Technologies such as the Technology Acceptance Model (TAM) and Emotional Response Language Education (ERLE) explore the affective and perceptual dimensions of technology use in language learning. TAM is employed to assess the degree to which a technology is accepted by users, based on perceived usefulness and ease of use. Meanwhile, ERLE emphasizes the role of emotional responses in influencing learner motivation and the effectiveness of language acquisition, highlighting the importance of psychological factors in the design of educational technologies.

Technology	**Article**
Technology Acceptance Model (TAM)	[11]
Emotional Response Language Education (ERLE)	[32]

7. **Robotics and Social Interaction in Language Learning**
The use of social robots such as NAO in language education represents an innovative approach that integrates human–machine interaction with principles of interactive pedagogy. These robots are programmed to provide appropriate linguistic and

gestural responses within learning scenarios, enhancing learners' cognitive and emotional engagement. This implementation reflects a new direction in language education, one that is more immersive and socially adaptive.

Technology	Article
NAO	[31]

RQ3: What Are the Advantages and Challenges of Implementing E-Assessment in Indonesian Language Speaking Instruction?

In today's digital era, the integration of information technology into education has become an inevitable necessity. One notable innovation that has emerged is electronic assessment (e-assessment). In the context of Indonesian language instruction, particularly in speaking skills, e-assessment offers several significant advantages. First, it enables a more flexible and efficient assessment process in terms of both time and location. Learners can record and submit their speaking tasks online, allowing educators to evaluate them more thoroughly and objectively. Second, technologies such as speech recognition and artificial intelligence (AI) have the potential to assist in capturing, recognizing, and analyzing students' speech automatically, providing accurate quantitative data related to pronunciation, intonation, and vocabulary precision.

Nevertheless, the implementation of speaking e-assessment also faces a number of challenges. The primary issues lie in curriculum readiness, infrastructure, and digital literacy, both among educators and learners. Limited access to devices and stable internet connectivity remains a significant barrier in various regions, particularly in Indonesia's 3T areas frontier, outermost, and underdeveloped regions (*terdepan, terluar, dan tertinggal*). Data from the Ministry of Communication and Information Technology indicates a persistent digital divide between urban and rural areas. This directly affects the feasibility of conducting online assessments, which require reliable connectivity and appropriate technological equipment.

In-depth aspects, such as observation of nonverbal expressions, emotional intonation, and the fluency of spontaneous communication, dimensions that are not always fully captured by digital platforms, remain a limitation. The restricted features of online assessment applications may result in assessments that do not entirely reflect students' actual speaking abilities.

Another phenomenon observed in Indonesia is the uneven distribution of training for teachers in utilizing digital assessment technologies. A survey conducted by the Center for Educational Research and Policy (Puslitjak) of the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) revealed that many teachers still struggle to integrate technology into language skill assessments, particularly in speaking. This underscores the need to enhance the pedagogical and technological competencies of Bahasa Indonesia teachers to ensure that the implementation of e-assessment is effective and carried out with accountability.

In addition, concerns over validity and academic honesty in the implementation of online assessments remain significant. In some cases, it has been found that speaking tasks submitted by students were not their original work, but recordings assisted or produced by others. This potential for data manipulation highlights the urgent need to strengthen digital ethics and implement identity verification systems within assessment platforms.

Therefore, although e-assessment in speaking holds great potential for supporting innovative Indonesian language instruction, its implementation must be accompanied by adequate infrastructure readiness, enhanced teacher competencies, strengthened digital literacy, and the development of assessment systems that ensure authenticity and fairness for all students across diverse regions in Indonesia.

CONCLUSION

The implementation of technology integrated with artificial intelligence (AI) in the assessment of speaking skills in Indonesian language instruction presents promising opportunities for educators to conduct more accurate, efficient, and objective evaluations. AI-based systems can capture, process, and analyze lexical patterns, pronunciation, intonation, and speech fluency, thereby addressing long-standing challenges in oral assessments, particularly those related to documentation, reliability, and subjectivity in scoring.

Despite its potential, the application of AI in this area still requires significant refinement. A primary challenge lies in the system's ability to recognize local linguistic nuances, such as regional dialects, sociolects, and culturally embedded communicative norms, that are essential for valid evaluation of students' speaking performance. Without proper adaptation, AI systems may misinterpret student utterances, leading to biased or inaccurate assessments. Moreover, the ethical dimensions of data collection and usage, particularly concerning the recording of students' voices, must be taken seriously to ensure data protection, informed consent, and pedagogical integrity.

This study offers a significant early contribution by illustrating how AI can serve not only as an assistive tool in language assessment, but also as a foundation for reimagining assessment frameworks within the broader digital learning ecosystem. It underscores the need for interdisciplinary collaboration among linguists, educators, computer scientists, and policymakers to develop robust, context-sensitive technologies for language education.

Looking ahead, we recommend the systematic integration of AI-driven assessment tools into national language curricula, supported by teacher training programs that strengthen both technological and pedagogical competencies. Furthermore, future research should focus on the development of localized speech corpora, the refinement of speech recognition algorithms for Indonesian and its regional varieties, and the construction of valid assessment rubrics that blend machine precision with human judgment. In doing so, Indonesia can take a leading role in advancing equitable, culturally responsive, and technologically enhanced language education for the digital age.

REFERENCES

Agustina, M., Pujiati, P., & Perdana, R. (2022). Pengembangan Instrumen Penilaian Kinerja Berbasis Model Project Based Learning untuk Meningkatkan Keterampilan Berbicara Peserta Didik di Sekolah Dasar. *Jurnal Basicedu*, *6*(4), 6900–6910. https://doi.org/10.31004/basicedu.v6i4.3281

Aizat, K., Mohamed, O., Orken, M., Ainur, A., & Zhumazhanov, B. (2020). Identification and authentication of user voice using DNN features and i-vector. *Cogent Engineering*, *7*(1), 17515. https://doi.org/10.1080/23311916.2020.1751557

Akhter, S., Haidov, R., Rana, A. M., & Qureshi, A. H. (2020). Exploring the significance of speaking skill for efl learners. *PalArch's Journal of Archaeology of Egypt / Egyptology*, *17*(9), 6019–6030. Retrieved from https://archives.palarch.nl/index.php/jae/article/view/5149

Aytac, S., Scheinfeld, L., & Tran, C. Y. (2025). How to do a systematic review. *The Reference Librarian*, 1–12. https://doi.org/10.1080/02763877.2025.2461780

Che Dalim, C. S., Sunar, M. S., Dey, A., & Billinghurst, M. (2020). Using augmented reality with speech input for non-native children's language learning. *International Journal of Human-Computer Studies*, *134*, 44–64. https://doi.org/10.1016/j.ijhcs.2019.10.002

Chen, H., & Mei, K. (2024). Exploring a Mobile Technology-Driven Model for Intercultural Communication Education. *International Journal of Interactive Mobile Technologies (iJIM)*, *18*(18), 62–75. https://doi.org/10.3991/ijim.v18i18.51491

Coombe, C., Vafadar, H., & Mohebbi, H. (2020). Language assessment literacy: What do we need to learn, unlearn, and relearn? *Language Testing in Asia*, *10*(1), 3. https://doi.org/10.1186/s40468-020-00101-6

Dang, R. (2024). English-speaking learning strategies in university based on artificial intelligence. *2024 IEEE 7th Eurasian Conference on Educational Innovation (ECEI)* (pp. 176–179). IEEE. https://doi.org/10.1109/ECEI60433.2024.10510791

Detey, S., Fontan, L., Le Coz, M., & Jmel, S. (2020). Computer-assisted assessment of phonetic fluency in a second language: A longitudinal study of Japanese learners of French. *Speech Communication*, *125*, 69–79. https://doi.org/10.1016/j.specom.2020.10.001

Filighera, A., Bongard, L., Steuer, T., & Tregel, T. (2022). Towards a vocalization feedback pipeline for language learners. *2022 International Conference on Advanced Learning Technologies (ICALT)*, 248-252. https://doi.org/10.1109/ICALT55010.2022.00081

Gatra, I. M. (2018). Meningkatkan keterampilan berbicara siswa SMA Dwijendra Gianyar melalui model pembelajaran contextual teaching and learning. *Journal of Education Action Research*, *2*(4), 322–330. https://doi.org/10.23887/jear.v2i4.16323

Geng, L. (2021). Evaluation model of college English multimedia teaching effect based on deep convolutional neural networks. *Mobile Information Systems*, *2021*(1), 1874584. https://doi.org/10.1155/2021/1874584

Gonzales, K. D., Maranan, J. R., Santelices, F. P. D., Renovalles, E. J. M., Macale, N. D., Palafox, N. A. A., & Mendoza, J. M. A. (2024). *BK3AT: Bangsamoro K-3 Children's Speech Corpus for Developing Assessment tools in the Bangsamoro Languages*. 59–65. https://aclanthology.org/2024.sigul-1.8/

Guney, C., Akinci, O., & Çamoğlu, K. (2021). Artificial Learning-Based Proctoring Solution for Remote Online Assessments: "Vproctor." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVI-4/W5-2021*, 235–238. https://doi.org/10.5194/isprs-archives-XLVI-4-W5-2021-235-2021

Gunnell, K. E., Belcourt, Veronica J., Tomasone, Jennifer R., & and Weeks, L. C. (2022). Systematic review methods. *International Review of Sport and Exercise Psychology*, *15*(1), 5–29. https://doi.org/10.1080/1750984X.2021.1966823

Halidjah, S. (2012). Evaluasi keterampilan berbicara dalam pembelajaran Bahasa Indonesia. *Jurnal Visi Ilmu Pendidikan*, *2*(1). https://doi.org/10.26418/jvip.v2i1.367

Henkel, O., Horne-Robinson, H., Hills, L., Roberts, B., & McGrane, J. (2025). Supporting Literacy Assessment in West Africa: Using State-of-the-Art Speech Models to Assess Oral Reading Fluency. *International Journal of Artificial Intelligence in Education*, *35*(1), 282–303. https://doi.org/10.1007/s40593-024-00435-9

Hong Y, Chen L-G, Huang J-H, Tsai Y-Y and Chang T-Y (2022) The Impact of Cooperative Learning Method on the Oral Proficiency of Learners of the Training Program for English Tourist Guides. *Front. Psychol.* 13:866863. https://doi.org/10.3389/fpsyg.2022.866863

Hong, Z.-W., Tsai ,Ming-Hsiu Michelle, Ku ,Chin Soon, Cheng ,Wai Khuen, Chen ,Jian-Tan, & and Lin, J.-M. (2024). Utilizing robot-tutoring approach in oral reading to improve Taiwanese EFL students' English pronunciation. *Cogent Education*, *11*(1), 2342660. https://doi.org/10.1080/2331186X.2024.2342660

Hou, R., Fütterer, T., Bühler, B., Bozkir, E., Gerjets, P., Trautwein, U., & Kasneci, E. (2024).. Automated Assessment of Encouragement and Warmth in Classrooms Leveraging Multimodal Emotional Features and ChatGPT. Dalam A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, & I. I Bittencourt (Ed.), *Artificial Intelligence in Education* (hlm. 60–74). Springer Nature Switzerland. https://doi.org/10.48550/arXiv.2404.15310

Iqbal, M. (2024). Automatic speech recognition (ASR) based on progressive web apps to develop pronunciation learning. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, *11*(1), 175–182. https://doi.org/10.33330/jurteksi.v11i1.3635

Ji, Y. (2024). *A deep learning assessment model for oral learning in english online education*. https://doi.org/10.1155/2022/6931796

Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* (3rd ed.). https://web.stanford.edu/~jurafsky/slp3/

Lee, Y., & Cho, J. (2020). Design of a Foreign Language Conversation Learning System Using Machine Learning. *Int'l Journal of Innovation, Creativity and Change*, *13*(4), 235–247. https://www.ijicc.net/images/vol_13/Iss_4/13426_Lee_2020_E_R.pdf

Li, W., & Mohamad, M. (2023). An efficient probabilistic deep learning model for the oral proficiency assessment of student speech recognition and classification. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(6), 411–424. https://doi.org/10.17762/ijritcc.v11i6.7734

Liu, Y.-F., Luthfi, M. I., & Hwang, W.-Y. (2024). Developing an AI-enhanced Video Drama-Making Learning System to Support. *Global Chinese Conference on Computers in Education Main Conference Proceedings (English Paper)*, 34-37. https://scholars.ncu.edu.tw/en/publications/developing-an-ai-enhanced-video-drama-making-learning-system-to-s

Luckin, R., & Holmes, W. (2016). *Intelligence unleashed: An argument for AI in education*. UCL Knowledge Lab. https://static.googleusercontent.com/media/edu.google.com/en//pdfs/Intelligence-Unleashed-Publication.pdf

Nasriddinov, F., Kocielnik, R., Gupta, A., Yang, C., Wong, E., Anandkumar, A., & Hung, A. (2025). Automating Feedback Analysis in Surgical Training: Detection, Categorization, and Assessment. Dalam S. Hegselmann, H. Zhou, E. Healey, T. Chang, C. Ellington, V. Mhasawade, S. Tonekaboni, P. Argaw, & H. Zhang (Ed.), *Proceedings of the 4th Machine Learning for Health Symposium* (Vol. 259, hlm. 787–804). PMLR. https://proceedings.mlr.press/v259/nasriddinov25a.html

Nurhamidah, D. (2021). Pengembangan instrumen penilaian berbasis media nearpod dalam mata kuliah bahasa Indonesia. *Pena Literasi*, *4*(2), 80–91. https://doi.org/10.24853/pl.4.2.80-91

Orosoo, M., Raash, N., Treve, M., M. Lahza, H. F., Alshammry, N., Ramesh, J. V. N., & Rengarajan, M. (2025). Transforming English language learning: Advanced speech recognition with MLP-LSTM for personalized education. *Alexandria Engineering Journal*, *111*, 21–32. https://doi.org/10.1016/j.aej.2024.10.065

Pandey, D., Subedi, A., & Mishra, D. (2022). Improving language skills and encouraging reading habits in primary education: A pilot study using NAO robot. *2022 IEEE/SICE International Symposium on System Integration (SII)*, 827-832. https://doi.org/10.1109/SII52469.2022.9708843

Penning de Vries, B. W., Cucchiarini, C., Strik, H., & van Hout, R. (2020). Spoken grammar practice in CALL: The effect of corrective feedback and education level in adult L2 learning. *Language Teaching Research*, *24*(5), 714–735. https://doi.org/10.1177/1362168818819027

Qian, Y., Ubale, R., Lange, P., Evanini, K., Ramanarayanan, V., & Soong, F. K. (2020). Spoken Language Understanding of Human-Machine Conversations for Language Learning Applications. *Journal of Signal Processing Systems*, *92*(8), 805–817. https://doi.org/10.1007/s11265-019-01484-3

Ravenor, R. Y. (2024). On the Adequacy of Elsa Speak in Formal Education: A Survey of Teacher-Users. *Advances in Artificial Intelligence and Machine Learning*. 04. 2387-2394. https://doi.org/10.54364/AAIML.2024.42138

Rohaini, B. (2021). Meningkatkan keterampilan berbicara siswa matapelajaran bahasa indonesia dengan menggunakan model time token di kelas x sma negeri 5 medan. *LANGUAGE : Jurnal Inovasi Pendidikan Bahasa dan Sastra*, *1*(2), 198–209. https://doi.org/10.51878/language.v1i2.759

Sintadewi, N. G. A., Sriasih, S. A. P., & Sudiana, I. N. (2017). Teknik penilaian keterampilan berbicara dalam pembelajaran bahasa Indonesia di SMA Negeri 4 Denpasar. *E-Journal Pendidikan Bahasa Dan Sastra Indonesia*, *7*(2), 1–12. https://doi.org/10.23887/jjpbs.v7i2.12001

Sloan, J., Maguire, D., & Carson-Berndsen, J. (2021). Emotional Response Language Education for Mobile Devices. *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. https://doi.org/10.1145/3406324.3417603

Sun, H. (2023). Advancing English Language Learning with Biotechnology: Utilizing Domain Adaptive Convolutional Neural Networks. *Journal of Commercial Biotechnology*, *28*(5), 283–294. https://doi.org/DOI:10.5912/jcb1198

Tolba, R. M., Elarif, T., Taha, Z., & Hammady, R. (2024). Interactive augmented reality system for learning phonetics using artificial intelligence. *IEEE Access*, 12, 78219-78231. https://doi.org/10.1109/ACCESS.2024.3406494

Turniyazova, S., Ne'matova, Y., Turdikulov, S., Makhkamova, N., & Sapaev, I. (2025). Wireless Mobile Network with Transfer Learning Algorithm for Multilingual Education and Historical Research. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*. https://doi.org/10.58346/JOWUA.2025.I1.035

Uppalapati, P. J., Dabbiru, M., & Kasukurthi, V. R. (2025). AI-driven mock interview assessment: Leveraging generative language models for automated evaluation. *International Journal of Machine Learning and Cybernetics*. https://doi.org/10.1007/s13042-025-02529-9

Wahyono, H. (2017). Penilaian Kemampuan Berbicara di Perguruan Tinggi Berbasis Teknologi Informasi Wujud Aktualisasi Prinsip-Prinsip Penilaian. *Transformatika: Jurnal Bahasa, Sastra, dan Pengajarannya*, *1*(1), 19–34. https://core.ac.uk/download/pdf/228479906.pdf

Wang, H., Sharma, A., & Shabaz, M. (2022). Research on digital media animation control technology based on recurrent neural network using speech technology. *International Journal of System Assurance Engineering and Management*, *13*(1), 564–575. https://doi.org/10.1007/s13198-021-01540-x

Wang, L. (2022). A Machine Learning Assessment System for Spoken English Based on Linear Predictive Coding. *Mobile Information Systems*, *2022*(1), 6131572. https://doi.org/10.1155/2022/6131572

Xie, Y. (2023). Application of speech recognition technology based on machine learning for network oral English teaching system. *International Journal of System Assurance Engineering and Management*. https://doi.org/10.1007/s13198-023-02143-4

Zhu, M. (2022). The Application of Intelligent Speech Analysis Technology in the Spoken English Language Learning Model. *Mobile Information Systems*, *2022*(1), 3192892. https://doi.org/10.1155/2022/3192892

Zou, B., Lyu ,Qinglang, Han ,Yining, Li ,Zijing, & and Zhang, W. (2023). Exploring students' acceptance of an artificial intelligence speech evaluation program for EFL speaking practice: An application of the Integrated Model of Technology Acceptance. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2023.227860