THE ROLE OF ARTIFICIAL INTELLIGENCE IN ASSESSING SPEAKING

Musulmonova Sug'diyona[1]

Abdullayeva Sevara[2]

ABSTRACT

The emergence of Artificial Intelligence (AI) has transformed language learning and assessment, particularly in evaluating speaking skills, one of the most complex and dynamic components of language proficiency. Traditional human-based assessments often lack consistency and scalability, whereas AI-powered systems offer automated, data-driven evaluations that are increasingly sophisticated. This paper explores how AI technologies especially Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Machine Learning (ML)—are employed to assess speaking. It evaluates the benefits, challenges, and ethical implications of AI use in this context and underscores the importance of combining AI with human expertise to ensure fairness, cultural sensitivity, and pedagogical relevance.

INTRODUCTION

Language assessment plays a vital role in education, employment, and migration contexts. Among the four language skills—reading, writing, listening, and speaking— speaking is arguably the most challenging to assess due to its real-time and performative nature. Traditionally evaluated through face-to-face interviews or recorded responses assessed by human raters, speaking assessments often suffer from issues such as subjectivity, inconsistency, and logistical complexity (O'Sullivan, 2012). Artificial Intelligence (AI) offers an alternative that promises objectivity, scalability, and instant feedback.

As AI systems continue to develop, their use in language testing—especially for speaking— has grown, with organizations like Educational Testing Service (ETS), Pearson, and Duolingo integrating AI-based technologies into their assessments. This paper investigates the

---

English Language Teachers

mechanisms behind AI-powered speaking assessments, highlights their strengths and limitations, and discusses implications for educators, learners, and policymakers.

Core Technologies Behind AI Speaking Assessment

1.   Automatic Speech Recognition (ASR):

ASR technology converts spoken language into written text. It is the foundation of many AI-speaking assessment tools. Advances in deep learning and large-scale acoustic modeling have significantly improved ASR accuracy even across diverse accents and speech rates (Hinton et al., 2012). For example, Google's speech-to-text system uses neural networks trained on thousands of hours of speech data.

2.   Natural Language Processing (NLP):

NLP enables computers to understand and generate human language. In the context of speaking assessment, NLP helps analyze lexical diversity, syntactic complexity, and

discourse coherence (Burstein et al., 2013). Tools like Coh-Metrix and SpaCy are examples of NLP applications in education.

3.   Machine Learning (ML):

ML algorithms are trained on large datasets to detect patterns in speech and assign scores based on features such as fluency, pronunciation, grammar, and vocabulary (Zechner et al., 2014). The systems can be continuously improved using supervised learning models based on human-annotated training data.

Benefits of AI in Assessing Speaking Objectivity and Reliability

Human raters may vary in their interpretations due to fatigue, mood, or implicit bias. AI offers consistent, repeatable scoring (Williamson et al., 2012), reducing the variability found in traditional assessments.

Scalability and Efficiency

AI systems enable large-scale administration of tests without the need for multiple trained assessors, thus lowering costs and increasing access, especially in remote or underserved areas (Chung, 2021).

Immediate and Actionable Feedback

AI tools provide real-time formative feedback to learners on pronunciation, fluency, and vocabulary usage.

Inclusivity

Online, AI-based assessments are accessible to learners regardless of location, helping bridge geographic and economic divides (Hardy & Thompson, 2020).

Limitations and Ethical Considerations Recognition Errors and Bias

ASR systems may misinterpret speakers with strong regional accents or speech disorders, potentially leading to unfair scores (Koenecke et al., 2020). Most systems perform best on speech that resembles the training data, which is often skewed toward specific varieties of English.

Lack of Pragmatic and Cultural Sensitivity

AI lacks the nuanced understanding of cultural norms, intonation, and discourse

249

pragmatics—critical elements in evaluating communicative competence (Taylor, 2006).

Data Privacy and Security

The collection and processing of voice data raise ethical concerns regarding consent, anonymity, and data storage. Regulations like the General Data Protection Regulation (GDPR) in the EU provide legal frameworks, but challenges remain (Popenici & Kerr, 2017).

Overreliance on Technology

AI should support, not replace, human judgment. Teachers and assessors must remain involved to ensure assessments align with educational goals and values (Loukina et al., 2019).

Case Studies

• Duolingo English Test: Utilizes AI to assess speaking, writing, and listening in an adaptive, online format. AI scores are reviewed by human raters for validation (Duolingo, 2022).

• ETS SpeechRater: This system evaluates spontaneous speech using over 100 features, including fluency, pronunciation, and grammar. Research has shown strong correlations between SpeechRater scores and human ratings (Zechner et al., 2009).

• Pearson Versant: Uses ASR and NLP to provide automated scoring of oral language proficiency. Widely used in corporate and educational settings, it offers rapid results with robust reliability (Townshend et al., 2013).

Conclusion

AI is reshaping the landscape of speaking assessment in language education. It offers consistency, scalability, and immediate feedback, helping educators and learners alike. However, AI should be seen as a complement—not a replacement—for human evaluators. While technological advances continue to improve scoring algorithms, the ethical, pedagogical, and cultural dimensions of language assessment require human insight. A hybrid model that combines the strengths of AI and human expertise represents the most promising path forward.

REFERENCES:

1. Burstein, J., Chodorow, M., & Leacock, C. (2013). Automated Essay Evaluation: The
2. Criterion Online Writing Service. AI Magazine, 34(1), 27–36.
3. Chung, J. (2021). Artificial Intelligence in Language Education: Benefits and Ethical Challenges. Language Learning Journal, 49(3), 314–326.
4. Duolingo. (2022). Duolingo English Test Research. Retrieved from
5. https://englishtest.duolingo.com/research
6. Hardy, Q., & Thompson, M. (2020). AI and Global Education Equity. International Review of Education, 66(2-3), 193–210.