



AI-Based Language Models (LLMs) in Assessing ESL Learners' Speaking Skills: Opportunities and Limitations

Leyla Hajiyeva^a

^a Azerbaijan University of Languages, Baku, Azerbaijan

E-mail address: leyla.hajiyeva.78@gmail.com

DOI : <https://doi.org/10.55248/gengpi.07.1026.3240>

ABSTRACT

Speaking is widely regarded as the most difficult skill to teach, practise, and assess in English as a Second Language (ESL) classrooms, largely because valid assessment of oral proficiency has traditionally depended on scarce resources: trained human raters, controlled testing conditions, and considerable time. The recent emergence of large language models (LLMs) capable of real-time speech-to-text conversion, conversational interaction, and rubric-referenced feedback generation has renewed interest in automating parts of this process. This article critically reviews the opportunities and limitations of LLM-based tools for assessing and supporting the development of ESL learners' speaking skills. It traces the evolution from earlier automatic speech-scoring systems grounded in automatic speech recognition (ASR) and statistical feature extraction to the current generation of generative, instruction-following language models capable of producing holistic and analytic feedback in natural language. Five categories of opportunity are identified — scalable formative feedback, on-demand conversational practice, rubric-referenced scoring support, fine-grained pronunciation and fluency diagnostics, and reduced speaking anxiety through low-stakes practice — alongside five categories of limitation: construct underrepresentation, accent- and dialect-related bias, hallucination and inconsistent reliability, washback effects and learner overreliance, and unresolved data-privacy and ethical concerns. Building on this synthesis, the article proposes a five-criterion framework — validity, reliability, fairness, transparency, and pedagogical alignment — intended to guide the responsible integration of LLM-based tools into ESL speaking instruction and assessment, and discusses implications for English-language teaching contexts such as Azerbaijan, where access to trained oral examiners is often constrained. The article concludes that LLMs are best positioned as a complement to, rather than a replacement for, human assessment of spoken English, and it outlines a research agenda for validating LLM-based speaking assessment against established testing standards.

Keywords: Large language models, ESL speaking assessment, Automated speech scoring, Artificial intelligence in language education, Formative feedback, Oral proficiency, Second language testing

1. Introduction

Of the four traditionally recognised language skills, speaking is consistently identified by both teachers and learners as the most anxiety-inducing and the hardest to assess reliably [1]. Unlike reading or grammar, which can be measured through discrete-item tests scored automatically with high reliability, oral proficiency is inherently interactive, context-dependent, and multidimensional, encompassing pronunciation, fluency, grammatical accuracy, lexical range, coherence, and pragmatic appropriateness simultaneously [2]. Reliable assessment of these dimensions has traditionally required trained human examiners working from calibrated rating scales, a resource-intensive process that is difficult to scale to large classes, frequent formative assessment, or contexts with limited numbers of qualified oral examiners — a constraint familiar to many English-language departments in non-English-dominant countries, including Azerbaijan.

Automatic speech scoring is not new: systems combining automatic speech recognition (ASR) with statistical models of fluency, pronunciation, and prosody have been used operationally in large-scale tests such as the TOEFL iBT SpeechRater since the late 2000s [3]. What has changed in the past several years is the arrival of generative large language models — systems such as those underlying ChatGPT and comparable conversational AI tools — that can not only transcribe and score speech but also generate natural-language explanations, simulate conversational partners, and adapt feedback dynamically to a learner's proficiency level [4]. This shift moves automated assessment from a narrowly psychometric exercise toward something closer to an interactive tutoring system, with corresponding new opportunities and new risks, a tension already visible in the everyday use of AI tools in ESL classrooms [5].

This article addresses three questions. First, how do LLM-based tools differ from earlier ASR-based automatic speech-scoring systems in their capacity to assess ESL speaking? Second, what specific opportunities do these tools offer for formative and summative speaking assessment? Third, what limitations — psychometric, ethical, and pedagogical — must be addressed before LLM-based speaking assessment can be responsibly integrated into ESL instruction? The article synthesises research from second language assessment, computer-assisted language learning (CALL), and the emerging

literature on generative AI in education to answer these questions, and it proposes a five-criterion evaluative framework intended to support teachers and institutions in appraising LLM-based speaking-assessment tools before adopting them.

2. Background: The Challenge of Assessing L2 Speaking

2.1. The Multidimensional Construct of Speaking Proficiency

Speaking proficiency is not a unitary skill but a composite construct. Widely used analytic rating scales, such as those underlying IELTS Speaking and the Common European Framework of Reference (CEFR) descriptors, decompose oral performance into components including fluency and coherence, lexical resource (including command of idiomatic and colloquial expressions [6], [7]), grammatical range and accuracy, and pronunciation [8]. Because these components interact — a learner may be grammatically accurate but disfluent, or fluent but lexically limited — valid assessment requires raters (human or automated) to weigh multiple, partially independent dimensions simultaneously, which is considerably more demanding than scoring a written multiple-choice item.

2.2. Human Rater Reliability and Its Costs

Even trained human raters show measurable inter-rater variability, rater drift over time, and susceptibility to halo effects, in which a strong first impression colours judgments of subsequent performance dimensions [9]. Addressing these threats to reliability requires rater training, double-marking or moderation procedures, and periodic recalibration — all of which impose recurring costs that are difficult to sustain outside large-scale standardised testing organisations. In everyday classroom contexts, teachers rarely have the time to assess each learner's spontaneous speech individually and provide detailed, rubric-referenced feedback on a regular basis, which limits how much spoken practice students receive with meaningful feedback attached [10].

2.3. From ASR-Based Scoring to Generative LLMs

Early automated speech-scoring systems combined ASR transcription with hand-engineered acoustic and linguistic features — speaking rate, pause frequency, pitch range, word-error-rate-based fluency proxies — feeding these into statistical regression models trained to predict human rater scores [3]. These systems achieved respectable score agreement with human raters on constrained, task-specific speech samples but offered little interpretive feedback: a learner received a number, not an explanation. Litman [11] noted that this first generation of educational speech and language processing systems was designed primarily for summative scoring rather than formative dialogue. Generative LLMs mark a qualitative shift: because they are trained to produce fluent, context-sensitive natural-language output, they can generate rubric-referenced explanations (“your response would benefit from more varied linking words such as *however* or *in addition*”), simulate extended conversational exchanges to elicit spontaneous speech, and adjust the linguistic complexity of their own output to the learner's level, functions that earlier feature-based systems could not perform [12].

2.4. Review Methodology

This article adopts the format of a critical narrative review rather than a systematic review with meta-analytic aggregation, since the object of inquiry — generative LLM applications to speaking assessment specifically — is too recent and methodologically heterogeneous for meta-analysis to be meaningful at this stage. The review draws on three overlapping bodies of literature, identified through targeted searches of Google Scholar, ERIC, and the *Language Learning & Technology* and *Language Assessment Quarterly* archives: (a) the established language-testing literature on the constructs and reliability challenges of L2 speaking assessment; (b) the CALL literature on ASR-based and dialogue-based systems for spoken language practice, predating the generative-AI era; and (c) the rapidly growing literature, largely published since 2022, on generative LLMs in education generally and language learning specifically. Sources were prioritised for inclusion where they addressed speaking or oral interaction directly, or where findings from the closely related domain of AI-assisted writing were judged to generalise plausibly to spoken-language contexts, a judgment made explicit wherever it occurs in the discussion below.

3. Opportunities

3.1. Scalable Formative Feedback

The most immediate opportunity LLM-based tools offer is scale. Where a teacher with thirty students cannot individually listen to and comment on each learner's spoken response every lesson, an LLM-based application can, in principle, provide every learner with individualised feedback on every speaking attempt, at any time of day. Warschauer et al. [13], writing about AI-generated text in second-language writing instruction, describe a comparable shift toward what they call “just-in-time” feedback that is available at the moment of need rather than delayed until a teacher can mark a batch of submissions; the same logic applies directly to spoken practice, where immediacy of feedback is known to support noticing and uptake [14].

3.2. *On-Demand Conversational Practice*

Conversational LLM applications can function as tireless, endlessly patient interlocutors, allowing learners to rehearse extended spoken interaction — booking a hotel room, debating a topic, describing an image — outside class time and without the social risk of making mistakes in front of peers. Bibauw, François, and Desmet [15], in their systematic review of dialogue-based CALL systems, found consistent evidence that task-based conversational practice with an artificial interlocutor increased learners' willingness to communicate, even where the interlocutor's linguistic naturalness was limited; the considerably more fluent and context-aware conversational behaviour of current LLMs plausibly strengthens this effect further, though this remains to be verified empirically for the newest generation of tools.

3.3. *Rubric-Referenced Scoring Support*

Rather than replacing the teacher's judgment, LLM-based tools can support it by generating a first-pass, rubric-referenced draft assessment — flagging, for instance, that a response showed limited range of tenses or frequent filler words — which the teacher can then confirm, adjust, or override. Used this way, the LLM functions analogously to the “second rater” recommended in high-stakes testing moderation procedures [9], potentially improving consistency without removing the teacher from the final decision, a division of labour consistent with the broader literature on AI-assisted (rather than AI-automated) assessment in higher education [16].

3.4. *Fine-Grained Pronunciation and Fluency Diagnostics*

Because LLM-based speech applications are frequently paired with modern ASR pipelines capable of phoneme-level alignment, they can generate diagnostic feedback that pinpoints specific segmental errors (for example, a learner's systematic substitution of /θ/ with /s/) or prosodic issues (flat intonation on *wh*-questions) that would require considerable phonetic training for a classroom teacher to identify reliably by ear in real time. This granularity extends what earlier ASR-based systems already did reasonably well [3] but couples it, for the first time, with natural-language explanation that a non-specialist learner can actually act on.

3.5. *Reduced Speaking Anxiety Through Low-Stakes Practice*

Foreign language speaking anxiety is one of the most robust predictors of underperformance in oral assessment, independent of underlying competence [1]. Because interaction with an AI system carries no social judgment and can be repeated privately as many times as the learner wishes, several classroom-based studies of AI conversational practice report reduced self-reported anxiety and increased willingness to attempt spontaneous speech compared with peer-to-peer practice [15], suggesting a plausible affective benefit that operates independently of any purely psychometric gain in scoring accuracy.

3.6. *Cost-Effectiveness and Broadened Access*

A further, largely institutional opportunity concerns cost. Commercial LLM-based applications typically operate on subscription or per-use pricing that is orders of magnitude cheaper than employing additional trained oral examiners, which may make regular, individualised speaking practice and feedback financially feasible for institutions — including under-resourced state schools and universities in developing economies — that could not otherwise afford it. Van Moere and Downey [10] note that this cost argument was already a central motivation for the earlier generation of ASR-based scoring systems in large-scale testing organisations; the argument applies with even greater force to everyday classroom formative assessment, where the alternative is often no individualised spoken feedback at all rather than a human-rated alternative.

4. Limitations

4.1. *Construct Underrepresentation*

A central psychometric concern is construct underrepresentation: the risk that an automated system, however sophisticated, systematically fails to capture dimensions of the speaking construct that matter for real-world communicative competence, such as interactional competence — the ability to manage turn-taking, repair misunderstandings, and co-construct meaning with a human interlocutor [17]. A learner practising with an infinitely patient, endlessly accommodating AI partner may never be assessed on, or even exposed to, the negotiation-of-meaning skills required when a real conversation partner is impatient, ambiguous, or uses regional variation and informal slang [7], meaning that high performance on an LLM-mediated task may not generalise to human-interlocutor performance.

4.2. *Accent- and Dialect-Related Bias*

ASR systems, which remain the front end of most LLM-based speech applications, are well documented to perform less accurately on speech from speakers of underrepresented English varieties and non-native accents than on speech resembling the standard varieties dominant in their training data [10]. Because a mis-transcription propagates directly into whatever downstream LLM scoring or feedback is generated, learners whose accent diverges

most from the training distribution — often precisely those from lower-resource language-learning contexts — risk being systematically under-scored or given feedback based on an inaccurate transcript of what they actually said, raising a serious fairness concern that mirrors long-standing critiques of standardised testing bias [17].

4.3. Hallucination and Inconsistent Reliability

Generative LLMs are known to occasionally produce fluent but factually or pedagogically incorrect output — a phenomenon commonly termed hallucination [4]. Applied to speaking feedback, this could mean an LLM confidently identifying a grammatical error that was not actually present in the learner's utterance, or praising a response that in fact contained a significant error, undermining the pedagogical value and, in assessment contexts, the validity of the feedback. Unlike the earlier generation of statistically trained ASR-based scoring engines, whose error patterns were relatively stable and could be characterised through standard reliability studies, the non-deterministic and rapidly updated nature of commercial LLMs makes reliability harder to establish and to guarantee will remain stable over time [12].

4.4. Washback Effects and Learner Overreliance

Washback — the influence that a test or assessment tool exerts on teaching and learning behaviour — is a well-established concern in language testing [17]. If learners and teachers come to treat LLM-generated scores as authoritative, instruction risks narrowing toward whatever the LLM measures well (for instance, lexical range and grammatical accuracy, which are relatively easy for a language model to evaluate) at the expense of dimensions it measures poorly (interactional competence, pragmatic appropriateness). A related risk is learner overreliance: Warschauer et al. [13] report, in the context of AI-assisted writing, that some learners begin to defer to AI-generated suggestions uncritically rather than developing independent self-monitoring skills, a concern that plausibly transfers to spoken practice if learners come to depend on AI feedback rather than developing their own error-detection strategies.

4.5. Data Privacy and Ethical Concerns

Speech is biometric data, and many LLM-based conversational tools operate through cloud-based processing that transmits learners' voice recordings to third-party servers, raising data-protection questions that are especially acute for learners under the age of eighteen [16]. Institutions adopting these tools must weigh pedagogical benefit against data-governance obligations, informed consent, and the possibility that voice data could be repurposed for model training or other undisclosed uses beyond the immediate educational context.

4.6. The Digital Divide

The cost-effectiveness argument advanced in Section 3.6 has a corresponding limitation: the benefits of LLM-based speaking practice accrue only to learners who have reliable internet access, a suitable device, and often a paid subscription, none of which can be assumed uniformly even within a single institution. Where adoption is uneven, LLM-based tools risk widening rather than narrowing the gap between learners who already have the most resources for independent language practice and those who do not, a concern that echoes long-standing critiques of educational technology adoption more broadly and that institutional decision-makers should weigh alongside the cost savings described above.

5. Toward a Framework for Responsible Integration

Synthesising the opportunities and limitations reviewed above, this article proposes five evaluative criteria that ESL teachers and institutions can apply when appraising an LLM-based speaking-assessment tool before adoption. Table 1 summarises the framework, together with an illustrative guiding question for each criterion.

Table 1 – A five-criterion framework for evaluating LLM-based ESL speaking-assessment tools.

Criterion	Definition	Guiding Question
<i>Validity</i>	The tool measures the dimensions of speaking proficiency it claims to measure, including interactional and pragmatic competence, not only fluency and grammar.	Does the tool's scoring rubric align with an established framework such as the CEFR, and has it been validated against human ratings on comparable tasks?
<i>Reliability</i>	The tool produces consistent scores and feedback across repeated administrations and over time, including across model updates.	Has consistency been checked across sessions, and is there a process for re-verifying reliability after the underlying model is updated?

Criterion	Definition	Guiding Question
<i>Fairness</i>	The tool performs comparably across learners with different first languages, accents, and dialectal backgrounds.	Has the tool been tested on speech from learners with the accents and L1 backgrounds represented in this institution's classrooms?
<i>Transparency</i>	Teachers and learners can understand the basis for a given score or piece of feedback, and errors can be identified and corrected.	Can a teacher trace a specific score back to the transcript and rubric criteria it was based on?
<i>Pedagogical alignment</i>	The tool is used to complement, not replace, teacher judgment and human interaction practice, and its use is matched to formative rather than exclusively summative purposes.	Is the tool positioned as one input among several, with clear boundaries on which decisions remain with the teacher?

The framework is deliberately not a checklist that a tool either passes or fails as a whole; rather, each criterion admits degrees, and institutions may reasonably accept a lower level of one criterion (for instance, imperfect fairness across accents) for low-stakes formative use while requiring a considerably higher standard before allowing any LLM-based tool to contribute to summative, high-stakes decisions such as course grades or placement. This graduated view mirrors the general principle in language-testing ethics that the burden of validity evidence should scale with the stakes of the decision being made [17].

6. Discussion: Implications for EFL Contexts such as Azerbaijan

The opportunities identified in Section 3 are arguably most valuable precisely in contexts where qualified oral examiners and small-group speaking practice are scarce relative to student numbers — a description that applies to many English-language departments in Azerbaijani universities and secondary schools, where large class sizes limit how much individualised spoken practice a teacher can realistically provide within scheduled contact hours. LLM-based conversational practice tools could, in principle, extend spoken-English practice opportunities substantially beyond the classroom without requiring additional staffing, and rubric-referenced draft feedback could help less experienced teachers apply CEFR-aligned speaking criteria more consistently as they gain assessment experience.

At the same time, the fairness concerns raised in Section 4.2 apply with particular force in contexts where learners' English is shaped by transfer from a typologically distant first language: Azerbaijani phonology and prosody differ substantially from English, a contrast also reflected more broadly in comparative Azerbaijani–English linguistic and literary research [18], and ASR systems trained predominantly on speech from learners with more typologically similar L1 backgrounds (for example, other Indo-European languages) may be comparatively less accurate for Azerbaijani-accented English, a hypothesis that would need direct empirical testing before any high-stakes use of these tools could be justified locally. This underscores the importance of the fairness criterion in the proposed framework and suggests a concrete, locally relevant research priority: benchmarking ASR transcription accuracy and downstream LLM feedback quality specifically on Azerbaijani learners' spoken English before any institution-wide adoption decision is made.

6.1. Implications for Teacher Training

Realising the opportunities described in Section 3 while managing the limitations described in Section 4 will depend heavily on teacher preparedness, not merely on tool design. Teachers who understand, at least at a conceptual level, why an ASR system might mis-transcribe accented speech, or why an LLM might occasionally generate confident but incorrect feedback, are far better positioned to use these tools critically — treating LLM output as a draft to be checked rather than an authoritative verdict — than teachers who are simply handed a new application with no orientation to its failure modes. This suggests that pre-service and in-service teacher-training programmes in EFL contexts should begin to incorporate basic AI literacy specifically framed around assessment validity and bias, rather than treating AI tools purely as a technical add-on to existing methods courses.

6.2. A Research Agenda

Several concrete empirical questions follow directly from the analysis above and constitute a research agenda for the field. First, comparative studies are needed that correlate LLM-generated speaking scores with scores from trained human raters on the same learner performances, disaggregated by learners' first-language background, to establish whether the fairness concern raised in Section 4.2 is empirically borne out and how large any gap is. Second, longitudinal classroom studies are needed to establish whether extended use of AI conversational practice measurably improves learners' spoken performance with human interlocutors, or whether gains remain confined to AI-mediated contexts, which would speak directly to the construct-underrepresentation concern raised in Section 4.1. Third, research is needed on optimal task design: which speaking-practice task types benefit most from AI mediation, and which are better reserved for human-to-human interaction. Fourth, hybrid approaches deserve exploration in which LLM-

generated judgments are combined with fuzzy-logic models, which have proved effective for decision-making under uncertainty in domains such as lighting design [19–21], agricultural price regulation [22], and smart environmental management [23]; such models could allow speaking scores to be expressed as graded degrees of proficiency with explicit uncertainty rather than as falsely precise crisp values. Finally, context-specific validation work — of the kind proposed for Azerbaijani learners in Section 6 above — will be needed in each new linguistic and educational setting before confident claims of fairness or effectiveness can be made locally.

7. Conclusion

Large language models have shifted automated speech assessment from narrow, numeric scoring toward something resembling an interactive tutoring partner capable of generating explanatory, rubric-referenced feedback and sustaining extended conversational practice. This shift creates genuine opportunities for ESL speaking instruction, particularly in contexts where access to trained human examiners and abundant speaking practice time is limited. It also introduces new categories of risk — construct underrepresentation, accent-related bias, hallucination, washback, and data-privacy concerns — that are distinct from, and in some respects harder to characterise than, the risks associated with earlier ASR-based scoring systems. The five-criterion framework proposed in this article — validity, reliability, fairness, transparency, and pedagogical alignment — offers one practical starting point for teachers and institutions to evaluate specific tools before adoption, calibrated to the stakes of the decision at hand. The most defensible current position is that LLM-based tools should complement rather than replace human judgment in ESL speaking assessment, particularly for high-stakes decisions, while offering substantial promise for low-stakes formative practice. Future research should prioritise empirical validation of LLM-based speaking feedback against trained human raters on learner populations with diverse L1 backgrounds, including Azerbaijani learners of English, in order to establish whether the opportunities identified here are realised equitably across the learners who stand to benefit most from scalable speaking practice.

References

- [1] Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125–132.
- [2] Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.
- [3] Zechner, K., Higgins, D., Xi, X., & Williamson, D. M. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication*, 51(10), 883–895.
- [4] Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.
- [5] Hajiyeve, L. (2025). Advantages and disadvantages of AI in modern ESL classes. *The Scientific Heritage*, (166), 62–64. <https://doi.org/10.5281/zenodo.16763524>
- [6] Hajiyeve, L. (2024). Analysing of the commonly used English idioms and their role in learning English. *Qədim Diyar*, 25, 25.
- [7] Seyidzadeh, L. E. (2019). Semantics of some slangs used in everyday speech in English with cognition.
- [8] Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- [9] Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, 10(2), 135–159.
- [10] Van Moere, A., & Downey, R. (2016). Technology and artificial intelligence in language assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 341–358). De Gruyter Mouton.
- [11] Litman, D. (2016). Natural language processing for enhancing teaching and learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), 4170–4176.
- [12] Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, 26(2), 5–24.
- [13] Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, P., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for second language writing. *Journal of Second Language Writing*, 62, 101071.
- [14] Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- [15] Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning*, 32(8), 827–877.
- [16] Bearman, M., & Luckin, R. (2020). Preparing university assessment for a world with AI: Tasks for human intelligence. In M. Bearman, P. Dawson, R. Ajjawi, J. Tai, & D. Boud (Eds.), *Re-imagining university assessment in a digital world* (pp. 49–63). Springer.

-
- [17] Chapelle, C. A., & Voss, E. (2016). 20 years of technology and language assessment in Language Learning & Technology. *Language Learning & Technology*, 20(2), 116–128.
- [18] Hajiyeva, L. E. (2026). Pseudonym as an integral element in the works of Azerbaijani, English and American writers and poets. *Baltic Journal of Legal and Social Sciences*, (2), 479–485. <https://doi.org/10.30525/2592-8813-2026-2-57>
- [19] Imamguluyev, R. (2020, July). Application of fuzzy logic model for correct lighting in computer aided interior design areas. In *International Conference on Intelligent and Fuzzy Systems* (pp. 1644–1651). Cham: Springer International Publishing.
- [20] Abdullayev, T., Imamguluyev, R., & Umarova, N. (2021, August). Application of fuzzy logic model for optimal solution of light reflection value in lighting calculations. In *International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions* (pp. 384–391). Cham: Springer International Publishing.
- [21] Imamguluyev, R., Mikayilova, R., & Salahli, V. (2022). Application of a fuzzy logic model for optimal assessment of the maintenance factor affecting lighting in interior design. In *Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2022* (pp. 463–474). Singapore: Springer Nature Singapore.
- [22] Imamguluyev, R., Elvin Balakishiyev, & Nihad Agakishiev. (2020). IMPROVING THE MECHANISM OF USING THE PRICE FACTOR IN THE EFFECTIVE REGULATION OF AGRICULTURAL PRODUCTION ON THE BASIS OF FUZZY LOGIC .*International Journal of Innovative Technologies in Economy*, 5(32). https://doi.org/10.31435/rsglobal_ijite/30122020/7325.
- [23] Imamguluyev, R., Sadiqova, L., Aliyeva, A., & Mammadhasanov, E. (2024, April). Fuzzy logic integration in smart environmental management systems: Paving the way for a greener future. In *International Conference on Smart Environment and Green Technologies* (pp. 381–389). Cham: Springer Nature Switzerland.