

Holistic Approach to Artificial Intelligence in Arabic Language Learning: Skill Effectiveness, Implementation Models, and Conceptual Challenges through Systematic Literature Review

Ika Wahyu Susiani

Institut Agama Islam Riyadlotul Mujahidin Ngabar Ponorogo, Indonesia
ikawahyususiani@gmail.com

Received: 15 Oct, 2025

Accepted: March 16, 2026

Published: April 20, 2026

Abstract

This study aims to systematically synthesize the findings of previous research on the use of Artificial Intelligence (AI) in Arabic language learning, focusing on the effectiveness of improving writing, speaking, reading, and listening skills, as well as implementation trends and accompanying conceptual challenges. This study uses the Systematic Literature Review (SLR) approach with the stages of identification, screening, feasibility assessment, and inclusion of studies based on the PRISMA flow. A literature search was conducted through the Scopus database using keywords related to AI and Arabic language learning, which resulted in seven articles that met the inclusion criteria for in-depth analysis. The results of the synthesis show that AI is most effectively used on writing and listening skills through formative feedback and audio-based learning media, while on reading skills AI plays a major role in readability analysis and cognitive load mapping. Conversely, speaking skills still have relatively little empirical support despite having great development potential. The implementation of AI in Arabic language learning is carried out through various approaches, including Generative AI, NLP systems, and adaptive learning, but it is still generally partial and not holistically integrated. The study concludes that AI has a significant contribution to Arabic language learning, but requires further development that is integrated, sensitive to linguistic and cultural contexts, and supported by a clear pedagogical framework.

Keywords: *Artificial Intelligence, Arabic Language Learning, Language Skills, Systematic Literature Review, AI Implementation*

Introduction

The development of Artificial Intelligence (AI) in the field of language education marks a significant pedagogic transformation, especially in the learning of languages with high linguistic complexity such as Arabic. Arabic has unique characteristics that include a root-based morphological system, flexible syntactic structure, a wealth of semantic meaning, as well as dialect variations that affect phonological and pragmatic aspects. This complexity is often a challenge for learners, especially non-native speakers, in mastering the skills of writing, speaking, reading, and listening in a balanced manner. In this context, AI is no longer seen as just a technical tool, but as a pedagogical instrument that has the

potential to reconceptualize the language learning process as a whole. Alkaabi and Almaamari Studies (2025) shows that the emergence of Generative AI has shifted the role of technology from passive media to active teaching assistants in Arabic language learning, which is involved in material planning, training delivery, formative assessment, and personalization of learning. This transformation marks the state-of-the-art use of AI in Arabic pedagogy that has gone beyond the experimental stage. However, the effectiveness of AI is still influenced by the readiness of educators, the pedagogical framework used, and the linguistic and cultural sensitivity inherent in Arabic.

In addition to showing technological advancements, the literature also shows a shift in research focus from the study of language skills separately to the recognition of the potential of AI to reach all four language skills simultaneously. Alkaabi and Almaamari Research (2025) revealed that AI has been used to support writing and reading skills through grammar correction and sentence structure analysis, as well as supporting speaking skills through text-based dialogue simulations. Although listening skills have not been the main focus in the study, the potential of AI in producing audio materials and listening exercises has begun to be identified. This suggests that recent research has recognized the ability of AI to contribute to the development of all four Arabic language skills, although its implementation is still partial. On the other hand, the research of Umi Hijriyah and colleagues (2025) provides strong empirical evidence that audio-based AI and real-time feedback are effective in significantly improving Arabic listening skills. Meanwhile, the research of Baazeem et al. (2025a) expanding the role of AI in reading skills through the integration of linguistic features and eye-tracking data to cognitively assess the readability of Arabic texts. However, each of these studies still focuses on specific skills, so it does not provide a comprehensive picture of the effectiveness of AI on all Arabic skills simultaneously.

From the perspective of adoption and implementation, the literature shows that the effectiveness of AI is not only determined by technological sophistication, but also by behavioral, pedagogical, and institutional factors. Al-Jamali and Abdalla's research (2025) highlighting that performance expectancy and social influence are the main determinants in the acceptance of AI by students in Arabic language learning, as explained through the Extended UTAUT2 framework. College students tend to use AI when they believe that the technology provides real benefits to improving their language skills. On

the other hand, Alenezi and Alenezi (2025) emphasizes the importance of AI-based formative assessments combined with a human-in-the-loop approach to maintain linguistic accuracy and cultural sensitivity. The study of Hanandeh et al. (2024) also reminded that the limitations of the Arabic corpus, the lack of dialect data, and the dominance of English-language data in the development of global AI are still significant structural obstacles. These challenges show that while AI has great potential, its implementation in Arabic language learning still faces various conceptual and technical limitations. Thus, there is an urgent need to conduct a systematic study that maps approaches, models, and forms of AI implementation as a whole.

Based on these conditions, this research was conducted to answer the conceptual and empirical gaps that are still contained in the literature related to the use of AI in Arabic language learning. Specifically, this study aims to synthesize the findings of previous research on the effectiveness of AI in improving writing, speaking, reading, and listening skills. In addition, this study also aims to identify approaches, models, and forms of AI implementation in Arabic language learning and its relationship with language skill development. Furthermore, this study examines the challenges, limitations, and conceptual issues identified in the literature, as well as their implications for the direction of further research development. Using the Systematic Literature Review approach, this study seeks to integrate previously fragmented findings into a holistic analytical framework. The relationship of this research with the existing literature lies in efforts to go beyond partial studies towards a comprehensive synthesis across language skills and AI implementation trends. Thus, this research is expected to make a significant theoretical and practical contribution to the development of Arabic language learning based on Artificial Intelligence.

Method

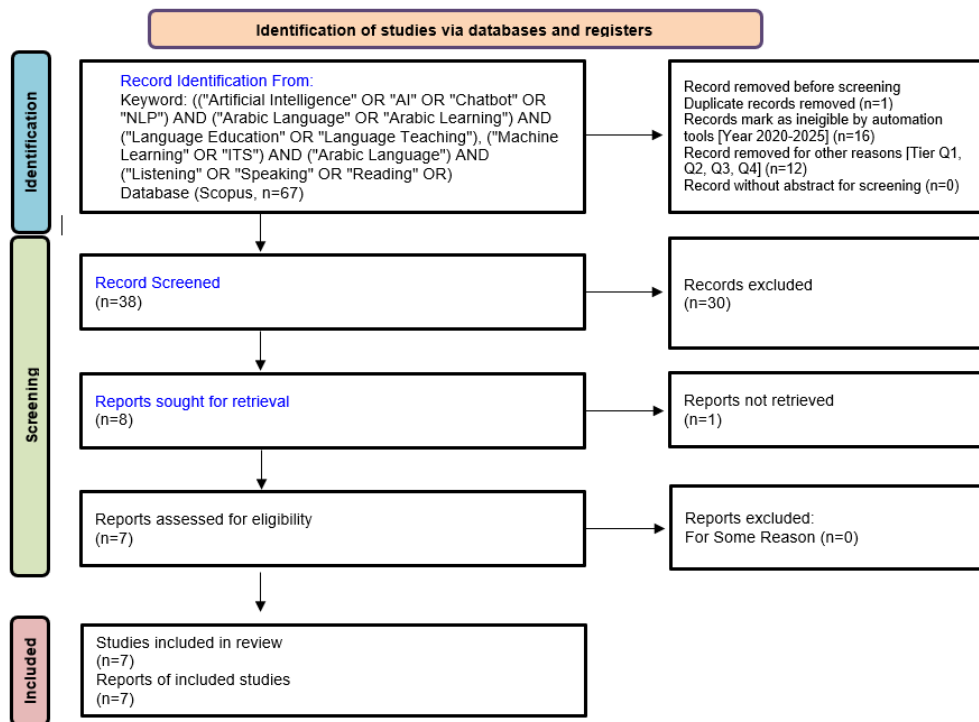


Figure 1. Literature Identification Process Using the PRISMA Procedure

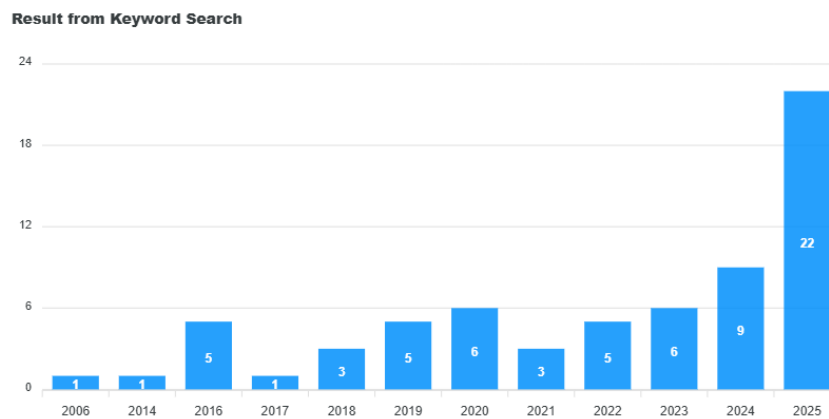


Figure 2. Graph of Related Literature from 2006-2025

This study uses the design of Systematic Literature Review (SLR) with the aim of systematically identifying, evaluating, and synthesizing previous research findings related to the use of Artificial Intelligence (AI) in Arabic language learning. The SLR approach was chosen because this study focuses on the conceptual and theoretical exploration of the effectiveness of AI, implementation models, and its challenges and

limitations, rather than on primary data collection. This design allows researchers to obtain a comprehensive picture of the development of AI research in Arabic language learning over time. In addition, SLR is used to integrate research results that were previously fragmented based on language skills or methodological approaches. The SLR process in this study follows the stages of identification, screening, feasibility, and inclusion of the study, as reflected in the PRISMA flow chart used. Thus, the design of this research ensures transparency, repeatability, and traceability of every step taken. This approach is also in line with the research objectives that want to map the state of the art, research gap, and novelty based on the available literature.

The characteristics of the sample in this study are not individuals or groups of respondents, but in the form of scientific articles that are relevant to the research topic. A sample of literature was obtained through a systematic search of the Scopus database with a range of years of publication 2006 to 2025, as shown in the graph of keyword search results. The keywords used include a combination of the terms "Artificial Intelligence", "AI", "Chatbot", "NLP", "Arabic Language", "Arabic Learning", "Language Education", "Machine Learning", as well as language skills such as "Listening", "Speaking", and "Reading". The initial identification process resulted in 67 thematically relevant articles. Furthermore, duplication and automatic filtering are carried out based on the criteria of the year of publication, so that a number of articles are eliminated at the initial stage. After the title and abstract screening process, 38 articles were obtained that were considered relevant for further study. Of these, only articles that meet the inclusion criteria are retained as final samples.

The inclusion criteria in this study include articles that explicitly discuss the use of AI in Arabic language learning, both in terms of language skill effectiveness, implementation approaches, and conceptual challenges. The selected article must be the result of empirical research or theoretical studies published in reputable scientific journals. Meanwhile, exclusion criteria include articles that are not relevant to the Arabic language, do not discuss AI substantively, or do not have an abstract that can be analyzed. In addition, articles that are only opinions without a clear methodological basis are also excluded from the sample. Based on the screening process, eight research reports were obtained that were searched in full text, but one report could not be accessed. At the feasibility assessment stage, seven articles were declared to meet all the criteria and

included in the final synthesis process. Thus, the number of studies analyzed in depth in this study is seven articles.

The main instrument in this study is a data extraction sheet that is used to collect important information from each selected article. This extraction sheet contains components such as research objectives, methodological design, types of AI used, language skills studied, key findings, and reported challenges and limitations. This instrument is prepared based on the need to answer the three research questions that have been formulated, namely related to the effectiveness of AI, implementation models, and conceptual issues. In addition, this instrument allows researchers to systematically compare findings between studies. The use of data extraction instruments also helps to minimize subjective bias in the literature synthesis process. All the extracted data were then grouped based on the main themes relevant to the research objectives. Thus, this instrument plays an important role in ensuring the consistency and reliability of the analysis process.

The stages of data collection are carried out in stages and structured. The first stage is the identification of the study through keyword searches on the Scopus database. The second stage is the screening of titles and abstracts to assess the initial relevance of the article to the research topic. The third stage is a full-text search to assess the feasibility of an article based on inclusion and exclusion criteria. The fourth stage is data extraction from articles that have passed the selection using the instruments that have been prepared. Each stage is carried out systematically to ensure that only truly relevant and quality articles are analyzed. This process is also documented in detail through PRISMA diagrams to maintain methodological transparency. By following these stages, this research can be replicated by other researchers by following the same procedure.

The data analysis technique used in this study is qualitative thematic analysis, which aims to identify patterns, themes, and trends in the analyzed literature. The data extracted from each article was analyzed by grouping the findings based on the language skills studied, the type of AI used, and the implementation approach applied. Next, a narrative synthesis was carried out to integrate the findings into a complete conceptual framework. The analysis also includes the identification of the challenges and limitations reported in the literature, as well as their implications for the future direction of the research. This analytical approach allows researchers to answer research questions comprehensively

without conducting quantitative meta-analysis. Thus, the analysis techniques applied are in line with the exploratory objectives and conceptual synthesis of this study.

Results



Figure 3. Literature Review

The results of the research synthesis show that publications related to the use of Artificial Intelligence (AI) in Arabic language learning have increased significantly in the last two decades. Based on the results of keyword searches on the Scopus database, the number of publications has increased gradually since 2006 and experienced a very clear spike in the period 2023 to 2025. The peak in the number of publications is recorded in 2025 with the highest number of articles compared to previous years. This pattern shows that the topic of AI in Arabic learning is a relatively new but rapidly evolving field of research. This temporal distribution reflects the increasing academic attention to the integration of intelligent technology in language education. In addition, the increase in publications also goes hand in hand with advances in Generative AI and Natural Language Processing technology which are increasingly accessible. Thus, the publication data shows a consistent and sustainable growth trend in AI and Arabic studies.

Based on the literature selection process using the PRISMA flow, the results of the initial identification resulted in 67 articles that were relevant to the research keywords. After

duplication removal and automatic filtering based on the criteria of the year of publication and initial relevance, a number of articles are eliminated before the screening stage. At the title and abstract screening stage, there are 38 articles that are considered relevant to the topic of using AI in Arabic learning. Of these, eight research reports were searched in the form of complete text to assess their feasibility. One report is inaccessible so it is not proceeded to the next stage. Furthermore, seven articles were declared to meet all inclusion criteria and did not meet the exclusion criteria. Thus, the final number of studies included in the synthesis is seven research articles. These results show that although AI and Arabic topics are growing in popularity, the number of studies that are truly relevant and meet the strict criteria of SLR is still limited.

The results of the synthesis also show that the studies analyzed include a variety of methodological approaches and different language skill focuses. Some of the research uses a qualitative approach to explore educators' experiences and perceptions of the use of AI in Arabic language learning. Other studies used an experimental quantitative approach to measure the improvement in specific language skills after the use of AI. In addition, there is research that uses a research and development approach to design and test AI-based learning media. This diversity of methodologies shows that the study of AI in Arabic language learning has not been standardized in one particular methodological approach. The focus of language skills studied also varies, ranging from writing, reading, listening skills, to aspects of evaluating the readability of texts. However, no studies were found that simultaneously measured all four language skills in a single research design. This is an important finding in mapping the results of the research.

Regarding Arabic writing skills, the results of the synthesis show that AI is most widely used in the form of grammar correction, sentence structure, and providing formative feedback. Research by Alkaabi and Almaamari (2025) reported that Generative AI is used by lecturers as teaching assistants to assist students in compiling Arabic written texts. AI is used to provide vocabulary suggestions, correct grammatical errors, and improve writing coherence. This finding is also strengthened by the research of Alenezi and Alenezi (2025) which shows that AI-based formative assessments are able to improve the quality of student writing revisions. In the study, students received quick and personalized feedback during the writing process. In addition, AI is also used to monitor the development of students' writing skills on an ongoing basis. Thus, the results of the

synthesis show that writing skills are the most dominant aspect in the use of AI in Arabic language learning.

On Arabic reading skills, the synthesis results show that AI is used mainly for text readability analysis and reading difficulty level mapping. Research by Baazeem et al. (2025) report on the use of Machine Learning and Deep Learning to predict the readability level of Arabic text by integrating linguistic features and eye-tracking data. The developed AI model is able to classify text based on difficulty with higher accuracy than text-based approaches alone. In addition, the development of the AraEyebility corpus provides cognitive data that represents the reading behavior of native Arabic speakers. This data includes the duration of fixation, regression, and eye movement patterns related to the reader's cognitive load. These results show that AI is used not only as a reading learning tool, but also as a cognitive evaluation tool. However, these studies did not directly measure the improvement in learners' reading ability after AI interventions.

For listening or listening skills, the synthesis results show that AI is used in the form of audio learning media equipped with real-time feedback. Research by Umi Hijriyah and colleagues (2025) reported on the use of the SUNO platform. AI in learning to listen to Arabic. This study shows a significant increase in students' listening skills scores after the use of AI-based media. AI is used to automatically present audio material, listening exercises, and comprehension evaluations. In addition, the platform allows learners to repeat the material as per individual needs. The results of this study provide quantitative empirical evidence regarding the effectiveness of AI in listening skills. However, this study only focused on one language skill and did not associate it with another.

The results of the synthesis show that Arabic speaking skills are the least empirically studied aspect in the literature analyzed. AI is reportedly used for dialogue simulations and text-based conversation exercises in several qualitative studies. Alkaabi and Almaamari (2025) report that lecturers use AI to help students practice speaking through chatbot-based conversations. However, the study did not present a quantitative measurement of speech skill improvement. Additionally, speaking skills are often only mentioned as potential uses of AI, rather than as a directly measured variable. No studies were found that specifically tested the effectiveness of AI on phonology, fluency, or pragmatic aspects of speaking Arabic. Thus, the results of the synthesis show the limitations of empirical data on speaking skills.

In terms of approach and implementation model, the synthesis results show that AI is applied in various forms, ranging from Generative AI, chatbots, NLP systems, to Machine Learning and Deep Learning. Generative AI is used primarily for text production, writing feedback, and dialogue simulation. NLP systems are used for linguistic analysis and Arabic text processing. Machine Learning and Deep Learning are used for readability evaluation and cognitive modeling tasks. Additionally, some studies emphasize a human-in-the-loop approach, where educators remain instrumental in verifying and adjusting AI outputs. This approach is reportedly important for maintaining linguistic accuracy and cultural sensitivity. Thus, the results of the synthesis show that the implementation of AI is diverse and contextual.

The results of the synthesis also show that the user acceptance factor is part of the research findings related to AI implementation. Research by Al-Jamali and Abdalla (2025) reported that performance expectancy and social influence influence students' intention to use AI in learning Arabic. Students are more likely to use AI if they feel the direct benefit of language skills. In addition, effort expectancy and facilitating conditions are also reported to affect the sustainability of AI use. These findings show that the implementation of AI is not only technical, but also influenced by psychological and institutional factors. However, this study did not directly link user acceptance to objective improvement of language skills.

Regarding challenges and limitations, the results of the synthesis show several issues that consistently arise in the literature. The limitations of the Arabic corpus, particularly for dialect variation, are reported to be a major obstacle in the development of AI. In addition, the dominance of English-language data in the development of global AI models has an impact on the accuracy of the system for the Arabic language. Some studies have also reported concerns over cultural sensitivity and religious values in the use of AI. Ethical issues and academic integrity also arise, especially related to the potential for over-reliance on AI. In addition, the limitations of previous research include a focus on a single language skill, a limited sample, and a narrow institutional context. These findings suggest a pattern of recurrent limitations in the literature.

Overall, the results of the research synthesis show that AI has been used in Arabic language learning with a primary focus on writing, reading, and listening skills, while speaking skills are still lacking in empirical attention. The implementation of AI is carried

out through various approaches and technologies, with varying levels of pedagogical integration. The number of studies that meet the SLR criteria is still limited, although publication trends show a significant increase. These key findings provide a comprehensive overview of the AI research landscape in Arabic language learning as reported in the literature analyzed. This part of the results only describes the findings without providing further interpretation. Interpretation and discussion of findings will be presented in the discussion section.

Discussion

The Effectiveness of Artificial Intelligence in Improving Arabic Language Skills

The findings of this study show that Artificial Intelligence (AI) has varying effectiveness in improving Arabic language skills, depending on the type of skills being studied and the form of technology implementation used. The literature that has been consistently analyzed shows that writing skills are the aspect that benefits the most from the use of AI. This is in line with the findings of Alkaabi and Almaamari who show that Generative AI is widely used by lecturers as teaching assistants to assist students in compiling Arabic texts, providing grammatical corrections, and improving coherence and fluency in writing. These findings are also reinforced by Alenezi and Alenezi who emphasize that AI-based formative assessments allow students to obtain quick and personalized feedback during the writing process. This effectiveness is significant because Arabic writing skills are known to have high complexity due to their morphological and syntactic systems. Thus, AI functions as a tool that is able to reduce the cognitive burden of learners while improving the quality of written results. The significance of these findings lies in the paradigm shift of writing learning from a post-production corrective approach to a reflective and sustainable writing process.

In the context of reading skills, research findings show that AI is more widely used as a cognitive evaluation and modeling tool than as a medium of direct learning intervention. Baazeem and colleagues' research shows that the integration of linguistic features and eye-tracking data allows AI to predict the readability level of Arabic text more accurately. This finding is important because reading Arabic involves not only lexical comprehension, but also complex morphological and syntactic processing. By utilizing cognitive data, AI is able to map the cognitive load of readers in more depth. Although

the study did not directly measure the improvement in reading ability after the use of AI, the theoretical implications were significant. These findings suggest that AI has the potential to be the basis for the development of adaptive and individual ability-based reading learning systems. The contribution of this research to the field of Arabic language learning lies in strengthening the cognition-based approach in the design of reading learning.

In contrast to writing and reading skills, the effectiveness of AI in listening skills is supported by relatively strong quantitative empirical evidence. Research by Umi Hijriyah and her colleagues shows that the use of AI-based learning media, especially audio platforms with real-time feedback, significantly improves the listening skills of Arabic language students. These findings are important because listening skills are often a major challenge for foreign language learners due to limited exposure to authentic inputs. AI allows the provision of audio material that can be tailored to the learner's ability level as well as allowing for repetition according to individual needs. The significance of these findings lies in AI's ability to create a more flexible and personalized listening learning environment. This research contribution reinforces the argument that AI is highly effective when used for language skills based on auditory input. However, the limitation of the study lies in its focus which covers only one language skill, so it does not provide an integrative picture across skills.

Meanwhile, speaking skills emerged as the least empirical support aspect in the literature analyzed. Although some studies, such as those conducted by Alkaabi and Almaamari, report on the use of AI for dialogue simulations and chatbot-based conversational exercises, no quantitative measurements were found that directly assessed the improvement of speaking skills. These findings show that there is a significant gap between the potential of the technology and the available empirical evidence. The significance of these findings lies in the disclosure of areas that are still underdeveloped in AI research and Arabic language learning. The contribution of this research is to affirm that speaking skills require a more complex approach, as they involve phonological, pragmatic, and social interaction aspects that are difficult to model computationally. Thus, the results of this study provide a solid basis for the development of future research agendas that focus more on speaking skills.

Approaches, Models, and Forms of AI Implementation in Arabic Language Learning

The research findings show that the approach and model of AI implementation in Arabic language learning are diverse and reflect the evolution of educational technology globally. One of the key approaches identified is the use of Generative AI as teaching assistants, as reported by Alkaabi and Almaamari. In this approach, AI is used to support learning planning, the creation of teaching materials, and the provision of exercises and feedback. This approach is significant because it marks a shift in the role of AI from a mere passive aid to an active pedagogical partner. The contribution of this research lies in the affirmation that AI can function as a support for the student-centered learning process. However, the literature also suggests that the success of this approach is highly dependent on the AI literacy of educators and the pedagogical framework used.

Another approach that stands out is the use of AI as a formative assessment tool, as described by Alenezi and Alenezi. In this model, AI is used to provide continuous feedback during the learning process, particularly in writing skills. These findings are significant because they show that AI can support process-based learning, not just the evaluation of the final outcome. Its theoretical contribution lies in strengthening the concept of formative assessment in the context of technology-based Arabic learning. However, the literature also emphasizes the importance of a human-in-the-loop approach to ensure linguistic accuracy and cultural sensitivity. Thus, effective AI implementation requires collaboration between humans and machines.

In the context of reading skills, the approach to AI implementation is characterized by the integration of technology with cognitive processing theory. Baazeem and colleagues' research shows that the use of eye-tracking data allows AI to understand the reading process more deeply. These findings are significant because they expand the role of AI from just a text analysis tool to a cognitive modeling tool. The contribution of this research to the development of knowledge lies in the incorporation of linguistic and cognitive approaches in Arabic language learning. However, the limitation of this approach is the high need for technical resources, making it difficult to implement widely in different educational contexts.

The approach to AI implementation is also influenced by user acceptance factors, as Al-Jamali and Abdalla's research shows. Using the Extended UTAUT2 framework, the study

showed that performance expectancy and social influence influence students' intention to use AI. These findings are significant because they confirm that the implementation of AI depends not only on the quality of the technology, but also on user perceptions and attitudes. The contribution of this research is to enrich understanding of the social and psychological dimensions in the adoption of AI for Arabic language learning. However, the limitation is the absence of a direct measurement between user acceptance and improvement of language skills.

Challenges, Limitations, and Implications of AI Development in Arabic Language Learning

The research findings suggest that the main challenge in using AI for Arabic learning lies in the linguistic complexity of the language itself. Hanandeh and colleagues emphasize that Arabic's complex morphological system, dialect variation, and pragmatic richness constitute significant obstacles to the development of accurate AI. These findings are significant because they explain why many global AI models are less than optimal when applied to Arabic. The contribution of this research is to underline the importance of developing AI that is sensitive to the linguistic characteristics of the Arabic language. These challenges also demonstrate the need for greater investment in the development of a representative Arabic corpus.

Data and corpus limitations also emerged as a major issue in the literature. Baazeem and colleagues' research shows that although the development of the AraEyebility corpus is a step forward, the scope of the data is still limited. These findings are significant because the quality of the data greatly determines the accuracy and generalization of AI models. The contribution of this research lies in the affirmation that the development of AI for Arabic must be supported by adequate linguistic and cognitive data. However, resource limitations and data collection costs are significant practical barriers.

Ethical issues and cultural sensitivity are also important concerns in the literature. Alenezi and Alenezi as well as Hanandeh and colleagues emphasized that the use of AI in Arabic language learning should take into account cultural and religious values. This finding is significant because Arabic is closely linked to cultural and religious identity. The contribution of this research is to affirm that AI development cannot be separated from

social and cultural contexts. However, the existing literature is still limited in providing an operational ethical framework.

Theoretically, this research contributes to the development of AI studies and Arabic language learning by providing a holistic synthesis across language skills. This research strengthens adaptive learning theory and formative assessment in the context of Arabic. In practical terms, the findings of this study provide guidance for educators and institutions in implementing AI effectively and responsibly. However, this study also has limitations, including the relatively limited number of studies analyzed and the focus of previous research that is still fragmented. However, these limitations actually open up opportunities for further research to develop a more integrative and empirical approach.

Conclusion

This study concludes that the use of Artificial Intelligence (AI) in Arabic language learning has shown significant development and has made a real contribution to improving the quality of learning, especially in writing, reading, and listening skills. The results of the literature synthesis show that AI is most effectively applied to writing skills through the provision of formative feedback, grammatical correction, and ongoing revision process support. In reading skills, AI plays an important role in the analysis of readability and mapping the cognitive load of readers, although its use is still more dominant as an evaluation tool than direct learning interventions. Meanwhile, listening skills gain strong empirical support through the use of adaptive and interactive AI-based audio learning media. Conversely, speaking skills still receive relatively little empirical attention, so the effectiveness of AI in this aspect cannot be strongly concluded. These findings confirm that the effectiveness of AI in Arabic language learning is differential and is strongly influenced by the characteristics of language skills and the form of technology implementation used.

In terms of approach and implementation model, this study shows that the use of AI in Arabic learning is carried out through various forms, ranging from Generative AI as a teaching assistant, NLP systems for linguistic analysis, to Machine Learning and Deep Learning for readability evaluation and cognitive modeling. The implementation of AI reported in the literature is generally still partial and focused on a specific skill or function, not yet thoroughly integrated in a single holistic pedagogical framework. In

addition, user acceptance factors, educator readiness, and human-in-the-loop approach have proven to be important in determining the success of AI implementation. The research also identifies a range of conceptual challenges, including the linguistic complexity of the Arabic language, the limitations of the data corpus, cultural and religious sensitivities, and ethical and academic integrity issues. By systematically mapping these challenges, this study provides a realistic picture of the limitations of the use of AI in Arabic language learning. Overall, these findings enrich understanding of the actual conditions and potential of AI development in the field of Arabic language education.

The main contribution of this research to the development of science lies in the provision of a holistic synthesis across language skills, implementation approaches, and conceptual issues that were previously studied in a fragmented manner. This research strengthens the theoretical foundation for the study of AI in Arabic language learning by placing technology as a pedagogical tool that is augmentative, not a substitute for the role of educators. In addition, this research makes a practical contribution by providing a reference for educators, researchers, and policymakers in designing and implementing AI-based Arabic learning more effectively and responsibly. The suggestions for further research include the need for empirical studies that examine all four language skills simultaneously in a single research design, in particular by paying greater attention to speaking skills. Future research is also suggested to develop an AI implementation model that is more integrated with the language learning theory and cultural context of the Arabic language. Thus, the future direction of research is expected to be able to produce a more comprehensive, contextual, and sustainable use of AI in Arabic language learning.

References

- Alenezi, A., & Alenezi, A. (2025). AI Formative Assessment in Saudi Education: A Study Across Universities. *Journal of Teaching and Learning*, 19(4). <https://doi.org/10.22329/jtl.v19i4.10012>
- Al-Jamali, S., & Suliman Abdalla, S. Z. (2025). Behavioral Determinants of AI-Driven Arabic Language Learning: Insights from the Extended UTAUT2 Model. *Educational Process International Journal*, 16(1). <https://doi.org/10.22521/edupij.2025.16.282>
- Alkaabi, M. H., & Almaamari, A. S. (2025). Generative AI Implementation and Assessment in Arabic Language Teaching: *International Journal of Online*

- Pedagogy and Course Design*, 15(1), 1–18.
<https://doi.org/10.4018/IJOPCD.368037>
- Baazeem, I., Al-Khalifa, H., & Al-Salman, A. (2025a). AraEyebility: Eye-Tracking Data for Arabic Text Readability. *Computation*, 13(5), 108.
<https://doi.org/10.3390/computation13050108>
- Baazeem, I., Al-Khalifa, H., & Al-Salman, A. (2025b). Integrating Linguistic and Eye Movements Features for Arabic Text Readability Assessment Using ML and DL Models. *Computation*, 13(11), 258.
<https://doi.org/10.3390/computation13110258>
- Department of Arabic Language Education, Faculty of Education and Teacher Training, Universitas Islam Negeri Raden Intan Lampung, Indonesia, & Hijriyah, U. (2025). How Effective Is SUNO.AI in Enhancing Arabic Listening Skills? An Evaluation of AI-Based Personalized Learning. *International Journal of Information and Education Technology*, 15(2), 391–407.
<https://doi.org/10.18178/ijiet.2025.15.2.2251>
- Hanandeh, A., Ayasrah, S., Kofahi, I., & Qudah, S. (2024). Artificial Intelligence in Arabic Linguistic Landscape: Opportunities, Challenges, and Future Directions. *TEM Journal*, 3137–3145. <https://doi.org/10.18421/TEM134-48>