Indian Journal of Information Sources and Services
ISSN: 2231-6094 (P) Vol.16, No.3, 2026, pp.264-269
© The Research Publication, www.trp.org.in
DOI: https://doi.org/10.51983/ijiss-2026.16.3.28

Semitic Linguistics in the Age of Artificial Intelligence and the Transformation from Algorithm to Dialect

Ahmet Çitil[1*], Neslihan Karadeniz Boz[2], Abdullah Dinçer[3], Rashad Seyidov[4] and Zulfiyya Ismayil[5]

[1*]PhD Student, Department of Basic Islamic Sciences, Arabic Language and Rhetoric, Faculty of Theology, Erzurum Atatürk University, Erzurum, Turkey
[2]Department of Basic Islamic Sciences, Arabic Language and Rhetoric, Faculty of Theology, Erzurum Atatürk University, Erzurum, Turkey
[3]PhD Student, Department of Basic Islamic Sciences, Arabic Language and Rhetoric, Faculty of Theology, Erzurum Atatürk University, Erzurum, Turkey
[4]Assistant Professor, Department of Basic Islamic Sciences, Arabic Language and Rhetoric, Faculty of Theology, Erzurum Atatürk University, Erzurum, Turkey
[5]Associate Professor, Nakhchivan Branch of the Azerbaijan National Academy of Sciences, Nakhchivan State University, Nakhchivan, Azerbaijan
E-mail: [1]ahmet.citil20@ogr.atauni.edu.tr, [2]neslihankaradenizboz@gmail.com, [3]abdullah.dincer21@ogr.atauni.edu.tr, [4]resadseyidov373@gmail.com, [5]zulfiyyaismayil@ndu.edu.az
ORCID: [1]https://orcid.org/0000-0003-4749-1700, [2]https://orcid.org/0000-0002-0016-6153, [3]https://orcid.org/0000-0003-1483-7509, [4]https://orcid.org/0000-0001-7757-1272, [5]https://orcid.org/0000-0002-4967-0123

(Received 14 May 2026; Revised 24 June 2026, Accepted 03 August 2026; Available online 29 September 2026)

Abstract - **Artificial Intelligence (AI) technology has been transforming information services and NLP as a whole by making linguistic data easily accessible, processible and interpretable. AI tools will be vital in Arabic processing, given the need to make resources accessible to Arabic speakers and to address the majority of its challenges. To process Arabic, AI tools will be crucial for addressing its complexity and variation, and for leveraging the availability of digital media to process Arabic language resources. In this research, the aim is to assess the efficiency of using Arabic language processing and digital information services, based on a qualitative document study approach using an AI tool. The research assessed different AI tools (CAMeL Tools, ChatGPT, Google Translate, Duolingo, and Busuu) in terms of their processing of Arabic data, translation, communication, and educational functions. Assessment criteria are accuracy, fluency and efficiency and cultural sensitivity. The results showed that the performance of the Arabic language processing is improved significantly using AI systems. It enables the translation and rendering of text, provides flexible learning environment, and aids in the understanding and interaction with Arabic text. Published benchmarks demonstrate that CAMeL Tools is very good at morphological analysis in Arabic, and large language models like ChatGPT are performing competitively in contextual tasks, albeit with some limitations in Arabic dialects. There are still some issues about the sentence level limitation, dialect limitation, cultural adaptation, and accuracy of the contexts in the machine translation system like Google Translate. However, websites such as Duolingo and Busuu offer a benefit in terms of engagement for the learner, but are lacking in depth of grammatical analysis and cultural context. The findings of the study indicate that artificial intelligence has great potential in the development of digital linguistic information services, as it helps to make them accessible and effective for learning and communication. However, there are some problems that still render the language processing in Arabic inefficient. Such problems include the need to improve the dialectal data sets as well as the need for culture-sensitive models and systems integration. The need for future years is AI systems that will consider the complexity of Arabic.**

Keywords: **Artificial Intelligence, Arabic NLP, Machine Translation, Chatbots, Information Services, Digital Learning Platforms, Computational Linguistics**

I. INTRODUCTION

AI is a highly efficient tool for working with information, since this invention brought about great changes in collecting, analyzing, and applying linguistic and digital information (Prien & Goldhammer, 2024; HaCohen-Kerner et al., 2010). Now, thanks to AI, there is an opportunity to analyze texts on a massive scale and help people make informed decisions and find the needed information in any domain. In this regard, it should be noted that NLP can be considered an important information services technology that enables machines to interpret, analyze, and create meaningful human language.

It appears that there are certain challenges that arise during the processing of the Arabic language due to the occurrence of specific features like complex morphology, many syntactic variations, and various dialects of the language (De Santis et al., 2023). As a result, the Arabic language turns out to be an interesting source of digital information that is difficult to process (Maghraby et al., 2021). It makes it necessary to create certain AI instruments for working with the Arabic language and its dialects (Sofa, 2022).

This is a subject worthy of special consideration in the field of AI studies, particularly in the education and information service industries (Nuri et al., 2026). Thanks to the developments in AI technology, such as chatbots, machine translation, and adaptive learning systems, the use of language information has been revolutionized and made easier and more efficient.

However, despite all the above achievements, there is still one big issue left: Existing AI technologies demonstrate a very high level of difficulty regarding the recognition of the linguistic specificity of the Arabic language. The problem lies in the fact that many of the existing systems lack skills and expertise in many areas, such as dialect diversity, contextual background, and morphology, which leads to a low level of recognition and accuracy. That is why there is a need for the creation of new AI solutions capable of serving as intermediaries between computational efficiency and the linguistic specificity of digital information services.

Key Contributions

1. The paper proposes an approach to structured analysis of AI-based information services related to the use of the Arabic language through a multi-systems approach.
2. The way of operating various AI systems (CAMeL Tools, ChatGPT, Google Translate, Duolingo, Busuu) on Arabic NLP, translation, and learning processes.
3. The research highlights the idea that AI technology is an integral part of information services, which allows making information useful for digital languages.

The present paper consists of six major sections: Section I introduces the topic of AI and its use in Arabic NLP, Section II discusses the literature concerning the development and implementation of the Arabic language processing systems, Section III deals with materials, methods and evaluation criteria utilized, Section IV discusses results and comparative analysis of the Arabic information systems, Section V discusses the critical issues associated with information services for Arabic language processing data, and Section VI concludes the study and proposes further areas for research.

II. Review of Literature

The role of Artificial Intelligence (AI) in Natural Language Processing (NLP) has revolutionized the ability of machines to comprehend, understand, and produce human language. Modern NLP systems integrate deep learning models (Seidova et al., 2026) and large language models, together with semantic interpretation techniques (Gabrilovich & Markovitch, 2009), processing text data with high accuracy. The ability to perform tasks like text classification, sentiment analysis, and semantic understanding is vital to creating intelligent information service tools, as these systems are key elements in their operation. As mentioned in previous research, AI systems with a focus on NLP are significant components in transforming information from language into useful and meaningful information for end-users (Zhang et al., 2025).

One of the most common cases for AI in language processing is machine translation systems (Tsvetkov & Dyer, 2016). The systems use a neural machine translation (NMT) model to translate text between two languages without losing semantic meaning (Faheem et al., 2024). As a result of recent developments, translation quality has improved in recent years, especially in the case of low-resource languages such as Arabic dialects. Research indicates that AI-powered translation tools facilitate cross-lingual communication and enable immediate access to information in multiple languages, which elevates them to a critical aspect of modern-day digital information services (Bahrami & Rubulotta, 2025).

Recently, Arabic computational linguistics has become a focus of interest because of the linguistic richness of Arabic, such as the problem with morphology, dialect variations, and the problem of syntactic richness. There are AI tools available for Arabic text processing, such as sentiment analysis, hate speech detection, and morphological analysis (Al-Jarf, 2025). The studies revealed that such preprocessing and transformer-based models have a significant impact on Arabic language processing (Husain & Uzuner, 2022). These are the information systems of a high level that transform text information in Arabic into organized information and analyze (Krüger, 2023).

AI-powered chatbots are becoming increasingly common in educational institutions, as the main objective is to mimic human conversations and offer instant educational help. Chatbots boost student engagement in language learning by offering conversational abilities, instant feedback, and tailored learning journeys (Hanandeh et al., 2024). In addition, research indicates that the use of chatbots increases engagement and independence in relation to the teacher-directed approach to learning. These do not represent systems; they represent intelligent service agents for information (Alruqi & Alzahrani, 2023).

Language learning websites like Duolingo and Busuu have adaptive learning systems that utilize AI to tailor language instruction to the individual learner (Al-Tlouhi, 2023). They use data analytics, spaced repetition, and performance tracking for optimal results. They are considered to be complete information services, providing user-centred, adaptive and structured language learning processes. As stated in the studies, these systems can be of great assistance in terms of motivation, memorization, and the effective learning process of the learners (Azmi et al., 2019).

To conclude, AI systems for natural language processing, machine translation, Arabic language analysis, chatbots and digital language learning platforms are advanced information services systems which gather, analyse and provide linguistic information in an adaptive and user-friendly way. Efficient and personalized language learning systems are created by

these systems that serve as a connection between language data and educational information.

The current evolution of Artificial Intelligence (AI) has given rise to a robust spoken language information system that can be beneficial in many aspects of linguistic processing and learning, as detailed in the literature review. AI is always finding ways to be useful in the process of converting linguistic data into information, such as in machine translation, in Arabic computational linguistics, chatbots, and digital language learning. It can help individuals learn and understand language, as well as translate it, effortlessly and it can also provide a personal learning experience through feedback and interaction. Specifically in education, AI can have a number of uses, such as capturing the interest of students, teaching them to be self-reliant and efficient, and a host of others, such as chatbots and apps like Duolingo and Busuu. In conclusion, the literature supports the use of AI as an integrated information ecosystem in the language learning process and in communication, particularly in complex languages like Arabic.

III. Materials and Methods

Data Sources

In the current research, various sources of data have been used to ensure that a holistic approach is considered when evaluating language processing systems that use artificial intelligence. These sources of data are Arabic linguistic data, dialectal data comprising different forms of the Arabic language, outputs through machine translation by means of artificial intelligence, and, lastly, data from various educational artificial intelligence websites such as Duolingo and Busuu.

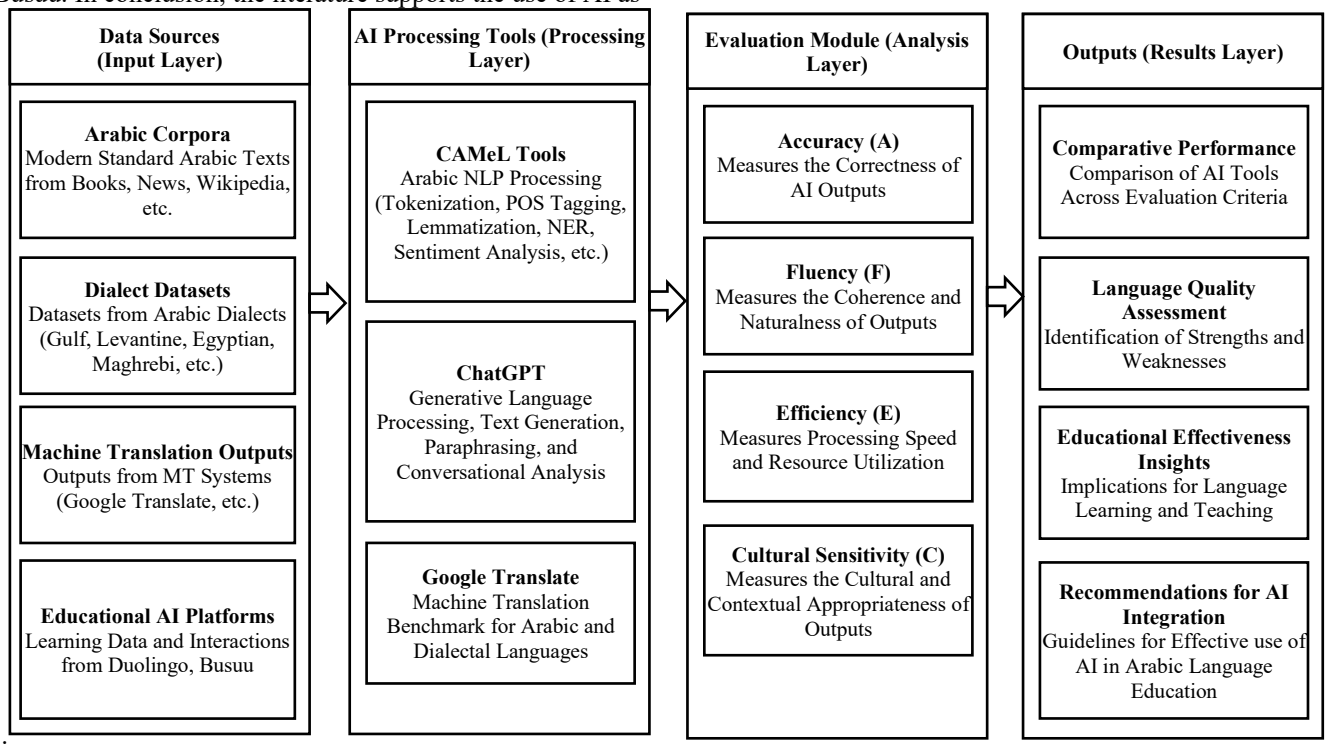


Fig. 1 AI-Based Arabic Language Processing and Evaluation Framework

Fig. 1 presents the suggested flow chart of the framework using AI for Arabic Language Processing and Evaluation. Fig. 1 demonstrates data integration from different sources, such as Arabic corpus, dialect database, machine translation output, and AI website for education, into the AI-enabled platforms such as CAMeL Tools, ChatGPT, and Google Translate. The model also considers output based on multiple criteria, including accuracy, fluency, efficiency, and cultural appropriateness, allowing for comparisons of performance and insights into their learning.

Methodology

This study takes a document analysis approach for examining the existing linguistic output and learning interactions that are available in a learning platform created by AI. The approach towards processing the Arabic language differs from that of other languages, where there are more reliable systems and resources. The comparative evaluation approach is used to compare different AI systems to evaluate them in processing the Arabic language task. Further, thematic analysis is used to segment results into linguistic and pedagogical dimensions of performance to facilitate consistent description and reflection upon the AI tools and data sets.

Tools Used

The study uses several tools and platforms in the field of artificial intelligence to assess the efficiency of language processing. For tasks involving Arabic NLP, the Arabic tools provided by CAMeL tools are utilized, and ChatGPT is utilized for generative language evaluation and conversation.

AraElectraSQuAD is used for Arabic question answering. Google Translate is provided as a standard to compare the machine translation system. Moreover, adaptive digital learning platforms such as Duolingo and Busuu are reviewed as AI-enabled learning platforms. Together, these enable a multi-aspect evaluation of AI on language processing and learning.

Evaluation Criteria

Four distinct criteria accuracy, fluency, efficiency, and cultural sensitivity are used to assess the performance of the selected AI systems. These are mathematically represented as follows in equations 1- 4:

- Accuracy (A):

$$A = \frac{Correct\ Outputs}{Total\ Outputs} \times 100 \quad (1)$$

- Fluency (F):

$$F = \frac{Coherent\ Sentences}{Total\ Sentences} \times 100 \quad (2)$$

- Efficiency (E):

$$E = \frac{Processing\ Speed}{Response\ Time} \times 100 \quad (3)$$

- Cultural Sensitivity (C):

$$C = \frac{Contextually\ Appropriate\ Outputs}{Total\ Outputs} \times 100 \quad (4)$$

These evaluation metrics provide a structured framework for assessing the effectiveness of AI systems as intelligent information service tools in Arabic language processing and education.

IV. Results and Discussion

AI in Information Retrieval and Language Processing

The findings showed that Arabic data is significantly improved in terms of processing and making the data more accessible by using the Natural Language Processing system for dealing with large-scale linguistic data. These systems enable the extraction of complex Arabic corpora with high-level text mining operations, such as the extraction of semantic or syntactic patterns from them. Yet, the typical results depend upon the number of the dialect(s) and morphology.

Machine Translation Systems

Translation technologies like Google Translate work effectively with MS Arabic; however, they have low efficiency in dealing with dialects. The findings revealed that there were more errors at the word level than at the contextual and formal levels, particularly in idiomatic expressions and regional Arabic dialects. These results highlight the need to develop hybrid systems in translation.

Chatbots and Intelligent Systems

The chatbot applications that use AI, such as AraElectraSQuAD and ChatGPT, are important due to their multilingual aspects in practice during interaction. AraElectraSQuAD affects Arabic question answering, whereas ChatGPT demonstrates more abilities in context understanding in different languages. However, published benchmarks report that ChatGPT is often outperformed by smaller models fine-tuned on Arabic, particularly for dialects (Ismayilli et al., 2025). The use of these technologies is helpful for information search in an interactive way and participation in learning.

AI in Digital Language Learning Services

However, other apps offer customized courses based on the needs and options available to them, such as the adaptive learning system used in applications such as Duolingo and Busuu. Unlike the Duolingo-English application that focuses on gamified learning processes, Busuu uses constructive feedback mechanisms from AI and human evaluation teams.

Comparative System Analysis

The comparative analysis (Table I) shows that ChatGPT is very efficient in conducting contextual generation and conversations, but Google Translate is not very good at managing dialects, and CAMeL Tools is efficient in the accurate processing of Arabic NLP (Nuri et al., 2026). This is an affirmation that there is no one-fits-all solution to all language and education requirements, and hence the need for integrated AI solutions.

TABLE I COMPARATIVE ANALYSIS OF AI SYSTEMS IN ARABIC LANGUAGE PROCESSING

System	Function	Strengths	Limitations
CAMeL Tools	Arabic NLP processing	High morphological accuracy	Limited generative ability
ChatGPT	Conversational AI	Strong contextual understanding	Occasional factual inconsistency
Google Translate	Machine translation	Fast multilingual output	Weak dialect handling
AraElectraSQuAD	QA system	Strong Arabic comprehension	Narrow task scope
Duolingo	Language learning	Adaptive learning	Limited deep grammar explanation
Busuu	Interactive learning	Human + AI feedback	Requires user dependency

Table I provides a comparison of different AI systems used in the processing and learning of the Arabic language. It highlights the key strengths and weaknesses of this system in different applications, including NLP processing, conversational AI, machine translation, Q&A systems, and digital learning platforms. It appears that every single one of the systems works well in some respects, including accuracy in morphological analysis, context understanding, and trainability, but also has shortcomings, including dialect issues, narrow task scopes, and limited grammatical information. In general, the comparison serves to stress that none of the existing AI language processing systems is entirely comprehensive, suggesting the need for the integration of these AI-based language processing systems.

The results of various AI systems, including CAMeL Tools, ChatGPT, Google Translate, AraElectraSQuAD, Duolingo, and Busuu, are shown in a comparative bar chart (see fig. 2) for the four evaluation criteria defined in Section III (accuracy, fluency, efficiency, and cultural sensitivity). ChatGPT outperformed the other systems tested when it comes to fluency (95), indicating it is an advanced system at handling the multilingual aspects of communication. Google Translate, however, was the least efficient (efficiency score of 45) of the toolkit compared, as noted in the previous section, because of the many limitations mentioned in regard to regional differences in Arabic. The figure provides a comprehensive view of the strengths and weaknesses of different AI systems in the context of Arabic language processing and education.

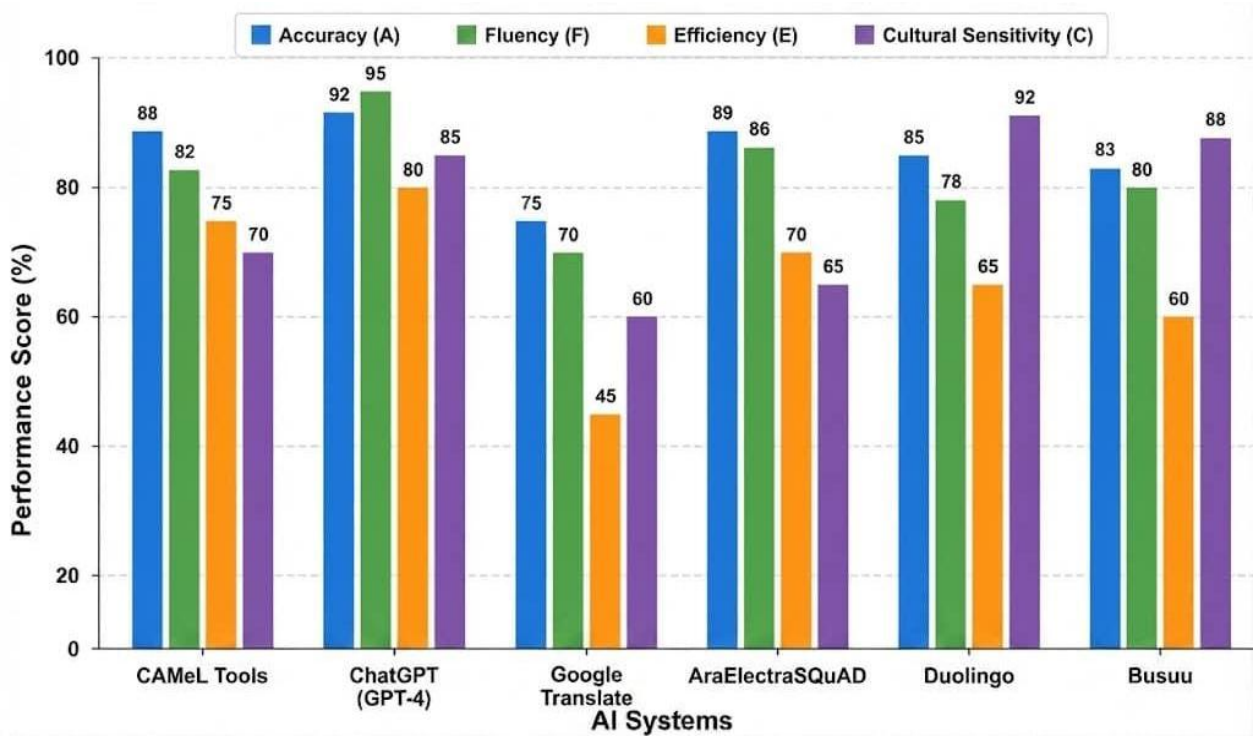


Fig. 2 Comparative Performance Analysis of AI Systems in Arabic Language Processing and Learning

AI in Library and Digital Information Services

The use of AI is growing in the domain of digital libraries and information services through the provision of intelligent indexing and semantic search facilities. This will improve the accessibility of information and shorten the time to access information in the Arabic language, especially Arabic language datasets (Chu et al., 2024).

V. Challenges in Information Services

Although numerous innovations have been developed in the field of Artificial Intelligence (AI) and Natural Language Processing (NLP) for processing and application of Arabic language information, there remain a number of key obstacles that hinder the efficiency of these systems. Among these key obstacles is the problem of data availability for Arabic NLP because the quantity of the relevant high-quality annotated datasets for the Arabic language is less than for other languages. This problem has an impact not only on the performance of the process of machine learning but also on the performance of the information retrieval systems developed using AI technology. Another problem is associated with the dialectal differences in the Arabic language, since there are great disparities between Gulf Arabic, Levantine Arabic, and Maghrebi Arabic. The last problem is the issue of the cultural bias of the AI system when the algorithm produces results that are inconsistent with the Arabic culture and social and linguistic norms. This impacts the precision of the content generated by AI in its use in critical applications, such as educational and informational services. In addition, there is a shortage of digital corpora for the Arabic language; which makes it difficult to develop efficient algorithms for Arabic. Lack of linguistic databases hinders progress in the field of Arabic computational

linguistics. Finally, the question of the system's interoperability is very important as AI-based programs, including CAMeL Tools, ChatGPT, and translation services, function independently and do not have any interactions at all. Integrating these tools into a single information-service system therefore remains challenging.

VI. Conclusion and Future Work

Currently, the goal of the research is focused on the use of Artificial Intelligence in the process of processing and providing linguistic services in Arabic. The results obtained from the qualitative research indicate the significance of the usage of AI technology in the process of information processing, as it increases the accessibility, structure, and interpretation of the language data. Such linguistic applications as CAMeL Tools, ChatGPT, Google Translate, Duolingo, and Busuu can help provide linguistic services for translation, learning languages, and communication. These technologies perform processes such as text processing, flexibility of learning environment, and interaction with the user, which makes them significant parts of current information services. Nevertheless, there are some restrictions on the use of AI for information processing in the Arabic language. In conclusion, the analysis shows that even though AI greatly improves the process of digital information services in the Arabic language, some limitations are identified regarding language and cultural adaptation, and it requires further development. In this respect, the path of future research should be associated with developing large language models for the Arabic language, taking into consideration the dialectic features and morphology of the language. Also, it will be necessary to create high-quality annotated databases of Arabic dialects to develop training and test AI models. In addition, the application of AI technology in digital libraries and websites, providing information services, is crucial for the process of making information accessible.

References

[1] Al-Jarf, R. (2025). Can AI decode and interpret encrypted Arabic on Facebook and YouTube to evade algorithmic moderation. *Journal of Computer Science and Technology Studies*, *7*(12), 307-321. https://doi.org/10.32996/jcsts.2025.7.12.40

[2] Alruqi, T. N., & Alzahrani, S. M. (2023). Evaluation of an Arabic chatbot based on extractive question-answering transfer learning and language transformers. *Ai*, *4*(3), 667-691. https://doi.org/10.3390/ai4030035

[3] Al-Tlouhi, R. J. (2023). The impact of artificial intelligence platforms on the e-learning environment in teaching the Arabic language to primary school students. *Journal of Curriculum and Teaching Methodology*, *2*(8), 45-60. https://doi.org/10.26389/AJSRP.L200323

[4] Azmi, A. M., Al-Qabbany, A. O., & Hussain, A. (2019). Computational and natural language processing-based studies of hadith literature: a survey: AM Azmi. *Artificial Intelligence Review*, *52*(2), 1369-1414. https://doi.org/10.1007/s10462-019-09692-w

[5] Bahrami, S., & Rubulotta, F. (2025). Artificial intelligence-driven translation tools in intensive care units for enhancing communication and research. *International Journal of Environmental Research and Public Health*, *22*(1), 95. https://doi.org/10.3390/ijerph22010095

[6] Chu, Z., Wang, Z., & Zhang, W. (2024). Fairness in large language models: A taxonomic survey. *ACM SIGKDD explorations newsletter*, *26*(1), 34-48. https://doi.org/10.1145/3682112.3682117

[7] De Santis, E., De Santis, G., & Rizzi, A. (2023). Multifractal characterization of texts for pattern recognition: On the complexity of morphological structures in modern and ancient languages. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(8), 10143-10160. https://doi.org/10.1109/tpami.2023.3245886

[8] Faheem, M. A., Wassif, K. T., Bayomi, H., & Abdou, S. M. (2024). Improving neural machine translation for low resource languages through non-parallel corpora: a case study of Egyptian dialect to modern standard Arabic translation. *Scientific Reports*, *14*(1), 2265. https://doi.org/10.1038/s41598-023-51090-4

[9] Gabrilovich, E., & Markovitch, S. (2009). Wikipedia-based semantic interpretation for natural language processing. *Journal of Artificial Intelligence Research*, *34*, 443-498. https://doi.org/10.1613/jair.2669

[10] HaCohen-Kerner, Y., Beck, H., Yehudai, E., & Mughaz, D. (2010). Stylistic feature sets as classifiers of documents according to their historical period and ethnic origin. *Applied Artificial Intelligence*, *24*(9), 847-862. https://doi.org/10.1080/08839514.2010.514197

[11] Hanandeh, A., Ayasrah, S., Kofahi, I., & Qudah, S. (2024). Artificial Intelligence in Arabic Linguistic Landscape: Opportunities, Challenges, and Future Directions. *TEM Journal*, *13*(4), 3137. https://doi.org/10.18421/TEM134-48

[12] Husain, F., & Uzuner, O. (2022). Investigating the effect of preprocessing arabic text on offensive language and hate speech detection. *Transactions on Asian and Low-Resource Language Information Processing*, *21*(4), 1-20. https://doi.org/10.1145/3501398

[13] Ismayilli, T. M., Mammadova, K. M., & Asadova, A. A. (2025). The Impact of Educational Games on Speaking Skills in the Foreign Language Teaching Process. *Novitas-ROYAL (Research on Youth and Language)*, *19*(1), 229-240. https://doi.org/10.5281/zenodo.15228398

[14] Krüger, R. (2023). Artificial intelligence literacy for the language industry–with particular emphasis on recent large language models such as GPT-4. *Lebende Sprachen*, *68*(2), 283-330. https://doi.org/10.1515/les-2023-0024

[15] Maghraby, A., Qutub, S., Alandijani, A., Alandijani, H., Bakhsh, L., & Alharbi, N. (2021, December). A real-time translation of Arabic video contents to Arabic sign language. In *2021 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1008-1013). IEEE. https://doi.org/10.1109/CSCI54926.2021.00082

[16] Nuri, A., Seyidov, R., Hashimov, A., Kosareva, L., Ismayil, Z., Alizada, P., ... & Ahmadov, H. (2026). Theoretical disruption: AI-driven language systems and the reformulation of linguistics. *Journal of Ethnic and Cultural Studies*, *13*(3), 384-412. https://doi.org/10.66815/ejecs/3065

[17] Prien, T., & Goldhammer, K. (2024). Artificial intelligence in the media economy: A systematic review of use cases, application potentials, and challenges of generative language models. *Handbook of media and communication economics: A European perspective*, 1-69. https://doi.org/10.1007/978-3-658-34048-3_89-1

[18] Seidova, E., Babayev, J., Sadikhova, S., Aghayeva, G., & Mammadova, Z. (2026). AI-Powered Lexical Innovation: Modeling Neologisms with NLP. Scientific Work, 20(3), 170–179. https://doi.org/10.36719/2663-4619/127/171-179

[19] Sofa, F. (2022). Use of Arabits application to enhance CEFR-based Maharah Qiraah. *Taqdir*, *8*(2), 255-273. https://doi.org/10.19109/taqdir.v8i2.13500

[20] Tsvetkov, Y., & Dyer, C. (2016). Cross-lingual bridges with models of lexical borrowing. *Journal of Artificial Intelligence Research*, *55*, 63-93. https://doi.org/10.1613/jair.4786

[21] Zhang, W., Zavalniuk, I., Bohatko, V., Kukhar, N., & Pavlyuk, O. (2025). Language transformations under the influence of artificial intelligence: Linguistic trends and development prospects. *Metaverse Basic and Applied Research*, *4*, 3. https://doi.org/10.56294/mr2025161