Systematic review: the application of ChatGPT on Arabic language text processing

Ali Mousa AlSbou[1,2], Fadzli Syed Abdullah[1], Ashanira Mat Deris[1]
[1]Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu (UMT), Terengganu, Malaysia
[2]Department of Computer Information Systems, College of Information Technology, Al Hussein Bin Talal University (AHU), Maan, Jordan

Article Info

ABSTRACT

Over 420 million people speak Arabic, and it is the official language of 22 countries. Its complex morphology and dialectal diversity present unique challenges for natural language processing (NLP) models like ChatGPT. This systematic review investigates the application of ChatGPT in Arabic language text processing, examining its potential uses, accuracy, and limitations. Covering literature published between 2021 and 2024, this review synthesizes findings from 21 articles, addressing four key research questions: ChatGPT's applications in Arabic text processing, its performance in terms of accuracy and reliability, the challenges and limitations encountered, and future directions to enhance its utilization. Results indicate that ChatGPT has potential in several applications, including educational tools, machine translation, text generation, and sentiment analysis. Despite current limitations, ChatGPT's potential in Arabic text processing is promising. While it shows high accuracy in structured tasks, it struggles with dialectal variations and cultural nuances, especially in complex text types. Primary limitations include a lack of high-quality Arabic datasets, difficulty handling dialects, and a need for more nuanced contextual understanding. Future research should focus on improving data quality, expanding dialectal coverage, fine-tuning models for specific linguistic tasks, and integrating AI with human teaching methods. Addressing these areas will enhance ChatGPT's accuracy and reliability for Arabic NLP.

Corresponding Author:

Fadzli Syed Abdullah
Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu (UMT)
21030 Kuala Nerus, Terengganu Darul Iman, Malaysia
Email: efadzli@umt.edu.my

1. INTRODUCTION

Arabic language is an official language for 22 Middle Eastern and North Africa countries. Over 420 million people speak Arabic worldwide. Due to its historical, cultural, and geopolitical functions, Arabic is still essential today. It is the liturgical language of Islam and has a rich literary history. Arabic plays a significant role in global business, regional politics, and international diplomacy, due to the economic significance of the region. It is crucial for communication and negotiation in these fields. Arabic has a distinct alphabet, complex morphology with a root-and-pattern system, a sophisticated phonological system, and adaptable syntax. Modern standard Arabic (MSA), classical Arabic (CA), and other regional dialects are all included in it, making it a crucial and sophisticated language in many fields today. Natural language processing (NLP) and artificial intelligence (AI) have particular hurdles and opportunities when dealing with Arabic's intricate morphology, syntax, and vocabulary [1].

AI technologies and applications are incredibly versatile, covering various fields, including NLP, robotics, computer vision, ML, and others. This versatility is what makes AI so fascinating, as it seeks to replicate human intelligence and decision-making in a wide range of contexts. NLP is a prime example of this, studying natural language generation, comprehension, and analysis. NLP's versatility is evident in its various applications, such as information extraction, sentiment analysis, machine translation, text summarization, and speech recognition. An example of this versatility is ChatGPT from OpenAI, a conversational AI built on the GPT architecture. Text generation models from OpenAI are taught to comprehend natural language, code, and pictures, and they are often called substantial language models or generative pre-trained transformers (GPTs). These models generate text outputs (in multiple languages) in response to inputs referred to as "prompts." [2].

Arabic natural language processing (ANLP) has seen remarkable growth as a research domain, driven by the intricate characteristics of Arabic such as its complex morphology, rich syntax, and diverse dialects. Modern ANLP systems heavily leverage machine learning (ML) techniques, which have proven effective despite the language's inherent ambiguities, including diglossia and unique script features. Significant advancements include the development of specialized tools like corpora and lexicons tailored for Arabic, supporting tasks such as parsing and part-of-speech tagging. However, challenges persist, such as the absence of standardized formal grammar for MSA, which hinders the evolution of more sophisticated systems. Additionally, addressing Arabic's sociolinguistic complexities, particularly diglossia, remains a nascent area of research. Given Arabic's global significance with over 400 million speakers, enhancing ANLP not only supports linguistic studies but also facilitates practical applications in variant domains. Continued innovation in ML methodologies and adapting existing NLP frameworks for Arabic are crucial to surmounting these obstacles and advancing the field [3], [4].

This systematic review aims to evaluate the current state of ChatGPT's application with Arabic text processing, addressing critical questions about its potential uses, accuracy, and main challenges and limitations. While several generative AI (GenAI) models and large language models (LLMs) have recently emerged, this review focuses on ChatGPT due to its widespread adoption and the unique linguistic and cultural challenges it presents when applied to Arabic language processing. The review will explore practical applications, including educational tools and language learning, machine translation, generating text, and sentiment analysis, to demonstrate the relevance and usefulness of the research. The paper is guided by four research questions that explore the applications, performance, challenges, and future directions of ChatGPT in Arabic text processing. The Method section then details the systematic review process, including the search strategy, inclusion and exclusion criteria, and the data extraction and synthesis approach. In the Results section, findings related to the research questions are presented, covering the applications of ChatGPT for Arabic language text, evaluating its accuracy and reliability, and discussing the challenges and limitations identified in literature. The Discussion section follows, exploring the broader implications of the findings for linguistic research, practical applications, and future directions to enhance the use of ChatGPT in Arabic text processing. The paper concludes with a Conclusion that summarizes the key findings, emphasizing the potential and limitations of ChatGPT in Arabic language processing and suggesting areas for future research.

2. RELATED WORK

While research on Arabic natural language processing (ANLP) has grown significantly, most previous surveys have focused on general language models or broader Arabic NLP applications. For instance, Seyidov [5] provided an overview of artificial intelligence applications in Arabic NLP, emphasizing prospects but without detailed performance analysis of individual models like ChatGPT. Similarly, Al-Sarayreh *et al.* [3] discussed the challenges of Arabic NLP in social media contexts but did not address generative models. A recent systematic review by Mustafa *et al.* [4] focused on speech emotion recognition but remained limited to prosody and vocal features, excluding large language models. In parallel, Ferdush *et al.* [2] reviewed ChatGPT's application in clinical decision support, demonstrating growing interest in specific domain-based assessments but not Arabic text processing.

To date, no systematic review has comprehensively examined ChatGPT's specific applications, performance, and challenges in Arabic language text processing. This paper fills that gap by synthesizing findings from 21 studies published between 2021 and 2024, addressing four focused research questions related to ChatGPT's use in educational tools, translation, text generation, and sentiment analysis. Our approach follows a PRISMA-guided review protocol and includes peer-reviewed and preprint sources, distinguishing it from prior narrative overviews.

3. METHOD

This paper employed a systematic review method for identifying, aggregating, and synthesizing existing research relevant to a research topic with the aim of synthesizing evidence. In sorting out the relevant articles (records) to be analyzed, researchers used a PRISMA model, see Figure 1. For related articles, a comprehensive literature search until 14/8/2024 was conducted using databases such as ScienceDirect, SpringerLink, Web of Science (WoS), Scopus, IEEE Xplore, ResearchGate, and ACL Anthology, as well as pre-print repositories such as medRxiv, and ArXiv. It utilized search engines like Google Scholar and Semantic Scholar.
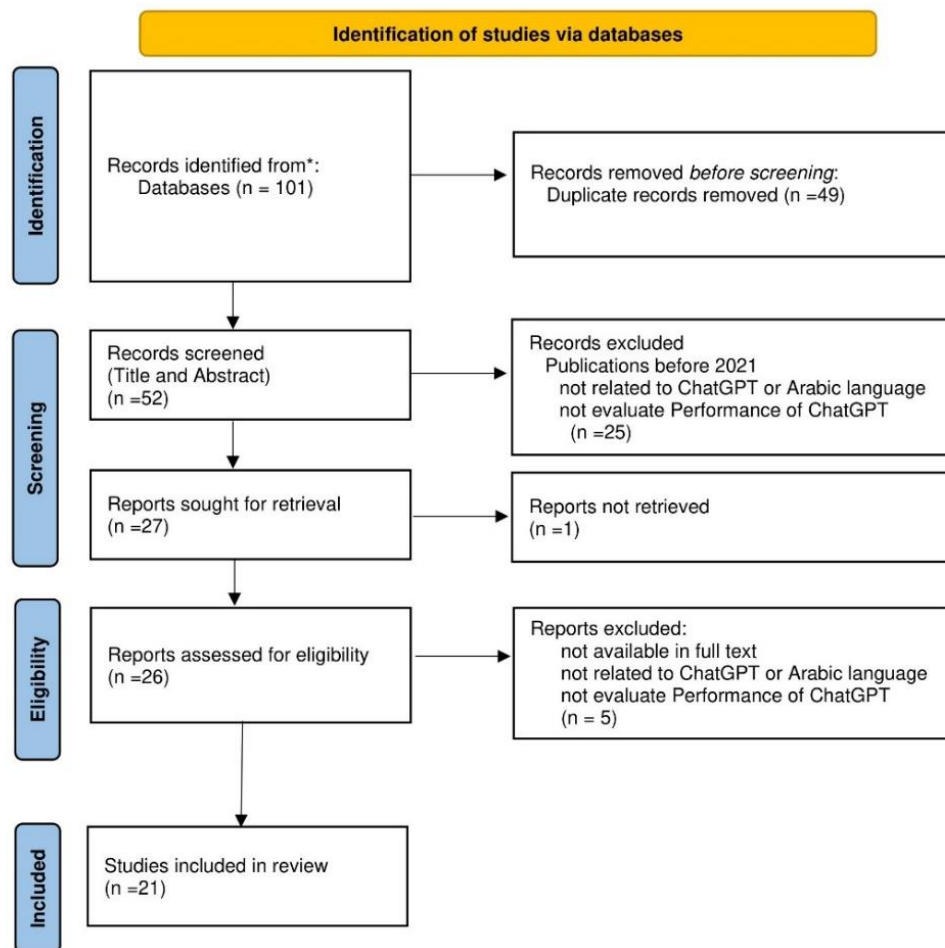

Figure 1. PRISMA flow diagram for systematic reviews

3.1. Definition of research questions

The objective of this study is to investigate the ChatGPT models that have been used for Arabic text processing. Four questions were defined Based on this objective:

a. Research question 1(RQ1): what are the primary applications (potential uses) of ChatGPT for Arabic language text processing?
b. Research question 2(RQ2): how does ChatGPT perform in terms of accuracy and reliability when processing Arabic text processing?
c. Research question 3(RQ3): what are the main challenges and limitations associated with using ChatGPT for Arabic language text processing?
d. Research question 4(RQ4): What future directions can enhance the utilization of ChatGPT for Arabic text processing?

3.2. Search phase

The search query was formulated using Boolean operators, with keywords such as "ChatGPT" "Arabic" and "openAI". In addition, some terms are listed in the table below. This analysis focused on recent

articles that contained relevant keywords in their titles or abstracts to address the research questions. Despite the targeted nature of our keywords specific to the Arabic language, we observed a notable need for more resources directly related to Arabic text, as much of the existing research tends to address broader aspects of text preprocessing and categorization. Thus, this highlights the critical need for more research tailored to Arabic NLP. Table 1 summarizes the number of records retrieved from Scopus and all other combined sources (*e.g.*, SpringerLink, IEEE Xplore, ACL Anthology, ResearchGate, and Google Scholar).

Table 1. Boolean query syntax

Boolean query syntax	Scopus	Others/Database
TITLE-ABS-KEY (OpenAI OR ChatGPT) AND Arabic	57	27
Application AND (OpenAI OR ChatGPT) AND Arabic	11	6

3.3. Inclusion and exclusion criteria (Eligibility criteria)

In line with standard systematic review methodology, clear inclusion and exclusion criteria were established to guide the selection of studies. These criteria set the boundaries of the review and help minimize selection bias by filtering the literature in a consistent, transparent manner. The purpose of defining strict criteria is to ensure that the review remains focused on pertinent, high-quality evidence while excluding out-of-scope or low-relevance studies. Below are the specific inclusion and exclusion criteria used in this review.

3.3.1. Inclusion criteria

Four (4) inclusion criteria were established:
a. Articles focused on the application of ChatGPT to Arabic language text.
b. Articles evaluating the performance and accuracy of ChatGPT on Arabic text.
c. Publications discussing the challenges and limitations of ChatGPT with Arabic language text.
d. Peer-reviewed articles, conference papers, and technical reports.

3.3.2. Exclusion criteria

Four (4) exclusion criteria were established:
a. Articles not related to ChatGPT or Arabic language.
b. Articles do not evaluate the performance and accuracy of ChatGPT on Arabic text.
c. Articles not available in full text.
d. Publications before 2021 were excluded. The review focuses on literature published between 2021 and 2024, aligning with the most recent developments in AI and NLP technologies to ensure the relevance of the findings.

3.4. Data extraction and synthesis

Initially, a total of 101 articles were selected from all databases, including 68 from SCOPUS, and 33 from other databases. The article selection process involved the following stages:
a. Removing duplications: Articles were imported into a Reference Manager System (EndNote) and 49 duplicate records were removed, reducing the number of articles to 52.
b. Title and Abstract Screening: This process involves reviewing the titles and abstracts to determine whether they are relevant to the topic based on the inclusion and exclusion criteria. As the result, 27 records were included for further analysis, with 25 being excluded.
c. Information gathering: Full text is reviewed in detail for eligibility assessment to ensure that the articles meet all the study questions. Data were extracted from selected articles using a standardized form, capturing information on study design, methods, key findings, and conclusions. As the result, 1 article was considered irrelevant to the study and excluded, leaving 26 articles.
d. Quality check: Finally, after assessing the eligibility of the articles, another 5 articles were being excluded for having poor paper quality, leaving only 21 articles to be included in this review study.

4. RESULTS AND DISCUSSION

After data extraction from selected articles, the narrative format will synthesize data. We present the results of the reviewed articles, focusing on our 4 research questions, particularly with regard to ChatGPT. To extract data from the included articles, a pre-defined data extraction model was used. The model contains the following variables: article title, authors, year, type, and other information. The distribution of articles by

publication year and publication type is shown in Figure 2. As can be seen, there has been a sharp increase in articles using ChatGPT for Arabic text in the past three years. While there was only one article in 2022, the number increased to 8 in 2023 and 12 in 2024. The distribution of articles by publication type shows that most of the articles included in this study (6 articles, 29%) are conference articles. Fourteen articles (67%) are workshop articles. There is also one journal letter (5%).
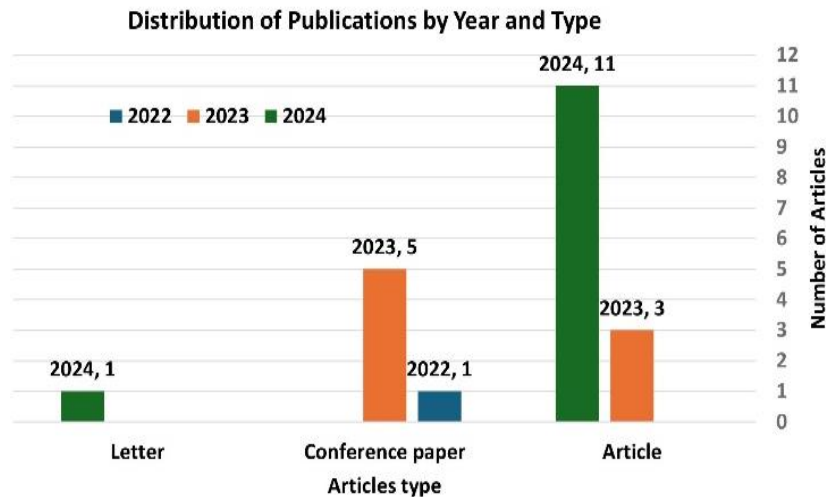

Figure 2. Distribution of publications by year and type

4.1. RQ1: What are some potential applications of GPT for Arabic language text processing?

Five predetermined themes emerged from RQ1 and were used in the synthesis. Those themes are: i) educational tools and language learning, ii) machine translation, iii) text generation, and iv) sentiment analysis. As illustrated in Table 2, 5 articles (23%) indicate that ChatGPT is both an educational tool and a language learning model. In comparison, ten articles (47% of the articles) highlight the potential applications of ChatGPT in machine translation. Specifically, four articles (19%) suggest that ChatGPT is a machine translation tool for converting English text to Arabic and vice versa. Additionally, two articles (0.9%) report its use in sentiment analysis. The Distribution of articles based on potential applications is shown in Figure 3. As can be seen, applications include educational tools (5 studies), machine translation (10 studies), text generation (4 studies), and sentiment analysis (2 studies).

Table 2. Distribution of articles based on potential applications

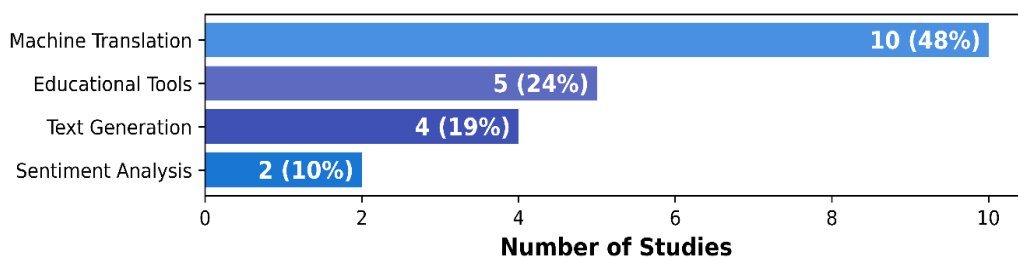| Application	List of articles	Total
Educational tools and language learning	[5]–[9]	5
Machine translation	[10]–[19]	10
Text generation	[20]–[23]	4
Sentiment analysis	[24], [25]	2

Figure 3. Distribution of articles based on ChatGPT potential applications

4.1.1. Educational tools and language learning

In recent studies, ChatGPT is highlighted as a versatile educational tool with significant applications for Arabic language teaching. Mohamed *et al.* [6] note that ChatGPT enhances Arabic language teaching, student performance analysis, and content-learner interaction by facilitating research, task completion, and engaging activities. Seyidov [5] points out its potential in developing intelligent tutoring systems, language learning platforms, and educational games tailored for Arabic learners through personalized learning. Nasaruddin [7] emphasizes ChatGPT's role in supporting Arabic language teachers with customized educational materials. Additionally, Butgereit *et al.* [8] demonstrate ChatGPT's effectiveness in the Prof Pi mathematics tutoring system for Arabic-speaking students, showing high user satisfaction and improved math skills. Lelepary *et al.* [9] find that ChatGPT significantly enhances university-level Arabic language learning by improving reading skills, boosting motivation, and easing assignment completion. This information underscores ChatGPT's growing importance in enhancing educational practices and resources for Arabic language learners.

4.1.2. Machine translation

ChatGPT is a machine translation tool that converts English text to Arabic and vice versa. Because it has advanced NLP capabilities, it provides instant and contextually relevant translations. Several studies have explored the use of ChatGPT for translating Arabic texts. Despite its strong performance in English and other high-resource languages, ChatGPT's effectiveness in Arabic translation faces unique challenges and has mixed results.

Research on ChatGPT's role in machine translation, particularly between Arabic and English, highlights its versatility and effectiveness across various domains. Banimelhem and Amayreh [11] evaluate ChatGPT as a tool for translating English to Arabic, noting its ability to handle diverse text formats, while Alkhawaja [13] emphasizes its potential to enhance translation efficiency and accessibility. Khoshafah [19] finds that ChatGPT generally delivers accurate translations, effectively conveying the intended meaning. In specialized applications, Alghamdi *et al.* [10] focus on fine-tuning ChatGPT-3.5 Turbo for translating financial news from Arabic to English, outperforming other neural machine translation models, while Alkhawaja [13] explore its use in film translation, demonstrating its ability to maintain quality across various dimensions. Obeidat and Jaradat [15] assess their effectiveness in translating resistance literature, particularly its ability to capture literary essence.

Alafnan [12] investigates its performance in translating high-stakes speeches, and Kadaoui *et al.* [26] examine ChatGPT's application across different Arabic dialects, including CA and MSA. Mohsen [16] demonstrates ChatGPT's precision in translating academic abstracts, particularly in specialized contexts, while Shahin and Ismail [17] explore their potential for translating Arabic sign language (ArSL) and other sign languages. Finally, AlKaabi *et al.* [18] evaluate its ability to translate culture-bound terms and idiomatic expressions in literary texts, specifically in Naguib Mahfouz's novel Zuqāq al-Midaqq. These studies collectively demonstrate ChatGPT's capability in various machine translation tasks, ranging from technical documents and financial news to literary works and sign languages.

4.1.3. Text generation

Research demonstrates ChatGPT's powerful potential in various linguistic text generation tasks, particularly in Arabic. El-Shangiti *et al.* [20] highlight its ability to generate coherent and fluent Arabic stories, emphasizing the tool's usefulness in creative writing and cultural storytelling tailored to Arab regions. Similarly, Beheitt and Hmida [21] show that GPT-2 can be effectively utilized to generate high-quality Arabic poems, reinforcing its capability in producing accurate cultural and linguistic content. Antar [22] discusses the application of ChatGPT and other large language models (LLMs) in creative writing, content generation, and educational tools specifically for Arabic-speaking audiences, with a focus on story generation. Additionally, Amin [23] explores the use of ChatGPT for automated Arabic text summarization, which is particularly valuable in academic research, content management, and information retrieval, offering fast and accurate summarization solutions. These studies underscore ChatGPT's versatility and effectiveness in generating and processing Arabic content across various creative and academic contexts.

4.1.4. Sentiment analysis

ChatGPT has demonstrated effectiveness in sentiment analysis and opinion extraction for Arabic texts, yielding promising results across various applications. It has been utilized to analyze social media data, customer characteristics, political opinions, and services. These capabilities suggest that ChatGPT can be a valuable tool for understanding public sentiment in Arabic-speaking regions, especially when large volumes of unstructured text data are involved.

Al-Thubaity *et al.* [24] suggest that advanced generative models, particularly GPT-4, perform relatively well on Arabic sentiment analysis tasks, even in low-shot settings, outperforming some fully supervised models. Similarly, Alderazi *et al.* [25] indicate that ChatGPT can classify sentiment and topics in Arabic social media, functioning alongside traditional machine learning and deep learning models. These studies highlight ChatGPT's potential in handling Arabic sentiment analysis tasks efficiently.

4.2. RQ3: How does ChatGPT perform in terms of accuracy and reliability when processing Arabic text?

Studies assessing ChatGPT's performance on Arabic text generally report high accuracy in generating grammatically correct and contextually relevant responses. However, the complexity of Arabic morphology and syntax poses challenges, sometimes leading to errors in word agreement and context interpretation. These findings suggest that while ChatGPT demonstrates strong linguistic capabilities, it still requires enhancement to manage the intricacies of Arabic grammar, particularly in dialectal and literary contexts.

4.2.1. Accuracy

ChatGPT shows great potential in processing Arabic, but its accuracy varies greatly depending on the task and context. Studies by [16], [20] show that ChatGPT, especially GPT-4, performs well on specific tasks such as story generation, sentiment analysis, and academic translations, often outperforming traditional models such as Google Translate in maintaining semantic integrity and coherence. However, it needs help with more subtle or complex tasks, such as accurately translating literary works, cultural nuances, expressions, and dialects.

Focusing on these struggles, especially in specialized fields such as literary translation, research by Banimelhem and Amayreh [11] Ali and Afzal [14] show that ChatGPT fails to capture cultural and emotional depth fully. In addition, the accuracy of the output often depends on the quality of the input, as highlighted by Nasaruddin [7], who noted that clear and structured instructions can improve ChatGPT's performance in educational contexts. Despite these limitations, studies such as those by Antar [22] and Al-Thubaity *et al.* [24] show that fine-tuning models such as GPT-4 improve accuracy, especially in sentiment analysis and creative writing tasks.

Although GPT-3.5 and GPT-4 outperform commercial MT systems when dealing with Arabic dialects, Kadaoui *et al.* [26] showed that they performed worse in CA and MSA. While ChatGPT shows promise, Alkhawaja [13] and others point out that its performance still needs to be improved for human translation, especially in complex or culturally rich texts. ChatGPT's accuracy is task-dependent, with better results in structured and fine-tuned applications, while challenges persist in more complex and nuanced language tasks.

4.2.2. Reliability

The results' synthesis highlights the varying reliability of ChatGPT across different tasks and contexts, particularly in Arabic language processing. ChatGPT generally performs well in well-defined, structured tasks, but its reliability decreases significantly in more complex, nuanced, and culture-dependent scenarios. This underscores the need for further development and improvement. ChatGPT shows high reliability when supported by clear instructions and guidance from teachers in educational settings, as demonstrated by Nasaruddin [7] and Butgereit *et al.* [8], with positive student feedback confirming its effectiveness. Similarly, Lelepary *et al.* [9] found that ChatGPT reliably supports independent language learning. These findings suggest that while ChatGPT may not fully replace traditional instruction, it can effectively complement it when thoughtfully integrated into educational strategies. For translation tasks, although there are minor shortcomings in complex language pairs. AlAfnan and Alkhawaja [12], [13] note that ChatGPT performs reliably for general and formal translations, but reliability declines in more nuanced and culturally sensitive tasks, such as film translations [14] and resistance literature Obeidat and Jaradat [15], where ChatGPT struggles to maintain cultural and emotional depth.

Studies Antar [22] and Al-Thubaity *et al.* [24] indicate that ChatGPT handles MSA reliably but becomes less reliable with dialectal variations due to limited training data. Seyidov [5] also emphasizes the impact of dialects and cultural contexts on the model's reliability in real-world applications. In sentiment analysis and academic translation, Alghamdi *et al.* [10] and Mohsen [16] highlight that ChatGPT, especially GPT-4, shows strong reliability and outperforms other models in these domains. However, its reliability declines when faced with complex or specialized texts that require deeper contextual understanding, as in [11], [25].

4.3. RQ3: What are the main challenges and limitations associated with using ChatGPT for Arabic language text?

The studies collectively highlight several key challenges and limitations of AI models like ChatGPT in processing Arabic language and related tasks. A significant issue is the need for more high-quality, annotated Arabic data and comprehensive datasets encompassing the full range of dialectal and linguistic nuances, affecting these models' performance [5], [6]. The vast dialectal diversity within Arabic also poses a challenge, as AI models struggle to process and generate text across different dialects accurately [5], [6], [20], [22], [24]. This limitation highlights the urgency of developing standardized datasets and training protocols to ensure fair and effective performance across the Arabic-speaking world. In addition, the ability to understand and capture cultural nuances is limited, often resulting in translation inaccuracies, especially in culturally sensitive contexts [5], [13], [15], [18], [19].

Another prominent issue is the dependence on user input; the quality of AI-generated outputs is highly influenced by how well users craft their prompts. This dependence underscores the need for user education and system improvement [7], [11]. Furthermore, AI-driven language learning tools reduce essential human interaction, which is necessary for developing nuanced understanding and critical thinking skills [9]. AI's performance in translation tasks remains suboptimal, particularly in domain-specific fields such as legal, medical, and scientific texts, with translation quality often depending on prompt sensitivity [11], [13]–[15]. AI models also struggle with complex linguistic structures, idiomatic expressions, and specialized vocabulary, especially in diplomatic and academic contexts [12], [16], [18], [19], [21].

The use of synthetic data introduces errors and unnatural sentence structures, requiring careful curation to maintain performance. This need for careful curation highlights the importance of data quality in AI model training [10]. Discrepancies between evaluation metrics also complicate the assessment of AI models [10]. Some studies identified limitations in AI systems' adaptation to non-English-speaking contexts, such as Arabic, due to differences in cultural norms and linguistic forms [8]. Computational constraints further limit the exploration of larger, more powerful models, especially in dialectal Arabic text generation [20], [22]. In terms of Arabic Sign Language, limited online resources hinder accurate translations between Arabic Sign Language and spoken Arabic [17].

ChatGPT also faces challenges in generating complex texts, such as Arabic poetry, requiring advanced language modeling to ensure thematic and stylistic coherence [21]. Additionally, omitting grammatical and lexical cohesion elements can affect the coherence of AI-generated summaries in complex texts [23]. Limited datasets in some studies restrict the generalizability of findings, particularly in real-world scenarios involving longer, more complex texts [13], [25]. AI struggles with capturing the cultural depth and ideological elements in resistance literature, often distorting translations, with some exhibiting deforming tendencies like rationalization and the destruction of linguistic patterns [15]. Finally, AI generative models are still underdeveloped in producing high-quality sentiment data in dialectal Arabic, especially for neutral sentiments [24].

Figure 4 summarizes the most frequently reported challenges encountered when applying ChatGPT to Arabic language text processing. Among the most cited issues are dialectal variation (reported in 8 studies), the lack of high-quality and comprehensive Arabic datasets (7 studies), and cultural nuance misinterpretation (6 studies). Other challenges include prompt sensitivity, difficulty translating complex texts—such as literary or domain-specific content—and reduced accuracy in specialized fields like legal and medical translation. These limitations highlight the need for richer datasets, better fine-tuning techniques, and culturally aware model training to improve ChatGPT's performance in Arabic NLP tasks.
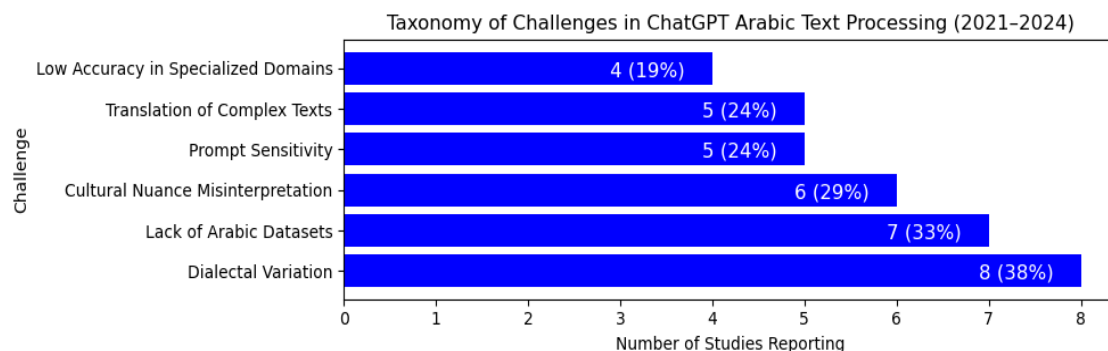

Figure 4. Reported challenges in ChatGPT Arabic text processing across studies

4.4. RQ4: What future directions can enhance the utilization of ChatGPT for Arabic text?

The synthesis of results highlights several future directions that could enhance ChatGPT's effectiveness in Arabic language processing. One of the key areas of focus is improving data quality. Developing more comprehensive datasets that encompass the full range of Arabic dialects and linguistic nuances is essential for enhancing the model's performance. This includes addressing the model's current limitations in recognizing and processing different dialects, which could make it more versatile and reliable for a broader audience. Additionally, fostering AI development that incorporates cultural context, and sensitivity will help models like ChatGPT better understand and process Arabic language with greater accuracy.

Human-AI collaboration is another critical direction for future research. Rather than replacing educators, AI should be seen as a tool that supports and complements human teaching methods. Encouraging this collaboration requires both technological improvements and teacher training to ensure AI systems align with educational goals and cultural expectations. This will require training for Arabic language teachers to effectively use ChatGPT, particularly in formulating precise and clear commands to ensure that AI-generated content aligns with their educational goals. AI should also be balanced with traditional learning methods, maintaining the importance of human interaction and critical thinking in language acquisition.

In terms of translation tasks, improving ChatGPT's contextual understanding and NLP capabilities will be crucial, particularly for complex or high-stakes translations that involve cultural or idiomatic nuances. Future work should focus on fine-tuning models for specific genres, such as literary and economic texts, while also refining AI models to preserve the ideological depth and emotional resonance in complex texts like resistance literature. Another area of focus is expanding datasets for Arabic dialects and sign language, ensuring that AI models can handle a wider variety of linguistic contexts with greater accuracy.

Error analysis and model refinement are also essential for improving ChatGPT's performance in Arabic dialects and complex linguistic tasks. Researchers should conduct thorough error analyses to identify limitations and develop more advanced fine-tuning strategies, which could enhance the model's performance in both text generation and sentiment analysis tasks. Furthermore, continued technical refinements in language models, particularly in handling specialized vocabulary and academic structures, will help bridge the gap between human and machine translation quality.

The publication trend indicates a rapidly growing interest in using ChatGPT models to process Arabic texts, as we found an increase in the number of articles over time, with an increase in 2024 compared to 2023 and 2022, noting that the study applied to publications over the past four years. However, due to the relatively small number of articles included in this review, which can be explained by the need for more research and the fact that the subject of study is relatively new, its concepts have only emerged in recent years. Overall, reviewing the 21 articles helped answer the four research questions.

The results of this review highlight the great potential of ChatGPT in Arabic text processing, especially in educational tools, machine translation, text generation, and sentiment analysis. The combined results of these studies point to an urgent need for improvements in data quality, diversity of training data, and computational resources to enhance ChatGPT's performance in handling Arabic language tasks. Addressing these issues could lead to more accurate and reliable AI models like CHATGPT that understand and generate Arabic better across different dialects and contexts. This comprehensive analysis shows that despite AI models like ChatGPT's tremendous potential, they face several language processing challenges, incredibly complex Arabic dialects, cultural differences, and specialized domains. These limitations highlight the need for further research and development to improve AI's ability to understand and process Arabic.

Future work should build on the challenges identified in the reviewed studies by prioritizing several key research directions. First, there is a clear need to develop and publicly share large-scale, high-quality Arabic corpora that include diverse dialects, formal modern standard Arabic (MSA), and underrepresented linguistic varieties such as Arabic Sign Language. Doing so will directly address the current limitations in training data that hinder model accuracy and reliability. Second, fine-tuning and customizing generative models like ChatGPT for domain-specific tasks - such as medical translation, legal terminology, or literary nuance - should be explored through transfer learning and targeted reinforcement techniques. As multiple studies in this review indicate, this would help improve performance in complex or culturally sensitive contexts. Third, future research should also investigate integrating cultural awareness into model training, particularly for sentiment analysis and translation of idiomatic expressions, which are prone to distortion in Arabic NLP. Moreover, scholars should develop standardized evaluation frameworks tailored for Arabic generative tasks to consistently benchmark model outputs, especially across dialects and genres. Finally, future studies may explore human-AI collaboration in educational contexts, emphasizing training educators to formulate precise prompts that improve ChatGPT's classroom utility without compromising pedagogical quality.

5.   CONCLUSION

This systematic review highlights the potential of ChatGPT in Arabic language processing across various applications, including education, translation, and sentiment analysis. While ChatGPT demonstrates high accuracy and reliability in structured tasks, challenges remain in dealing with Arabic's linguistic and cultural complexities, especially in dialectal variations and specialized translations. Addressing these challenges will be crucial for enhancing the utility of AI systems in Arabic NLP, making them more robust and culturally adaptive. The findings of this systematic review highlight new areas of research that need further improvement and future work. Future research should focus on improving data resources and quality, enhancing AI model performance, integrating AI tools with human teaching methods, enhancing cultural and dialectal sensitivity, and exploring new applications in Arabic language education and accessibility. By addressing these gaps, researchers can unleash the full potential of ChatGPT and similar models for Arabic NLP, ultimately contributing to more accurate, reliable, and culturally relevant AI language processing tools.

REFERENCES

[1]   "Arabic speaking countries," *WorldAtlas*. https://www.worldatlas.com/articles/arabic-speaking-countries.html (accessed Aug. 15, 2024).

[2]   J. Ferdush, M. Begum, and S. T. Hossain, "ChatGPT and clinical decision support: scope, application, and limitations," *Annals of Biomedical Engineering*, vol. 52, no. 5, pp. 1119–1124, Jul. 2024, doi: 10.1007/s10439-023-03329-4.

[3]   S. AL-Sarayreh, A. Mohamed, and K. Shaalan, "Challenges and solutions for Arabic natural language processing in social media," in *Smart Innovation, Systems and Technologies*, vol. 358, Springer Nature Singapore, 2023, pp. 293–302.

[4]   H. H. Mustafa, N. R. Darwish, and H. A. Hefny, "Automatic speech emotion recognition: a systematic literature review," *International Journal of Speech Technology*, vol. 27, no. 1, pp. 267–285, Mar. 2024, doi: 10.1007/s10772-024-10096-7.

[5]   R. Seyidov, "Arabic language processing: current status and future prospects of artificial intelligence," *Journal of Namibian Studies : History Politics Culture*, vol. 41, pp. 224–240, Apr. 2024, doi: 10.59670/jwszy037.

[6]   Y. Mohamed, Z. A. Razak, S. Ismail, M. H. Ibrahim, and Z. M. Isa, "ChatGPT-AI technology in teaching Arabic for specific purposes from the Arabic expert's perspective," *Al-Qanatir: International Journal of Islamic Studies*, vol. 32, no. 2, pp. 213–225, 2023.

[7]   N. Nasaruddin, "Using ChatGPT in teaching Arabic as a foreign language," *Arabiyatuna: Jurnal Bahasa Arab*, vol. 8, no. 1, p. 1, May 2024, doi: 10.29240/jba.v8i1.9413.

[8]   L. Butgereit, H. Martinus, and M. M. Abugosseisa, "Prof Pi: tutoring mathematics in Arabic language using GPT-4 and Whatsapp," in *INES 2023 - 27th IEEE International Conference on Intelligent Engineering Systems 2023, Proceedings*, Jul. 2023, pp. 161–164, doi: 10.1109/INES59282.2023.10297824.

[9]   H. L. Lelepary, R. Rachmawati, B. N. Zani, and K. Maharjan, "ChatGPT: opportunities and challenges in the learning process of Arabic language in higher education," *Journal International of Lingua and Technology*, vol. 2, no. 1, pp. 10–22, Jul. 2023, doi: 10.55849/jiltech.v2i1.439.

[10]   E. A. Alghamdi, J. Zakraoui, and F. A. Abanmy, "Domain adaptation for Arabic machine translation: financial texts as a case study," *Applied Sciences (Switzerland)*, vol. 14, no. 16, p. 7088, Aug. 2024, doi: 10.3390/app14167088.

[11]   O. Banimelhem and W. Amayreh, "Is ChatGPT a good English to Arabic machine translation tool?," in *2023 14th International Conference on Information and Communication Systems, ICICS 2023*, Nov. 2023, pp. 1–6, doi: 10.1109/ICICS60529.2023.10330525.

[12]   M. A. Alafnan, "Large language models as computational linguistics tools: a comparative analysis of ChatGPT and Google machine translations," *Journal of Artificial Intelligence and Technology*, vol. 5, pp. 20–32, Jun. 2025, doi: 10.37965/jait.2024.0549.

[13]   L. Alkhawaja, "Unveiling the new frontier: ChatGPT-3 powered translation for Arabic-English language pairs," *Theory and Practice in Language Studies*, vol. 14, no. 2, pp. 347–357, Feb. 2024, doi: 10.17507/tpls.1402.05.

[14]   S. Ali and N. Afzal, "A multidimensional analysis of human and ChatGPT-generated English translations of Arabic film," *World Journal of English Language*, vol. 14, no. 4, pp. 632–641, Jun. 2024, doi: 10.5430/wjel.v14n4p632.

[15]   M. M. Obeidat and M. A. Jaradat, "Artificial intelligence accuracy in translating resistance literature from Arabic into English: Google Translate and ChatGPT as a model," *Research Journal in Advanced Humanities*, vol. 5, no. 2, pp. 45–62, 2024, doi: 10.58256/9yz4dx79.

[16]   M. A. Mohsen, "Artificial intelligence in academic translation: a comparative study of large language models and Google Translate," *Psycholinguistics*, vol. 35, no. 2, pp. 134–156, Apr. 2024, doi: 10.31470/2309-1797-2024-35-2-134-156.

[17]   N. Shahin and L. Ismail, "ChatGPT, let us chat sign language: experiments, architectural elements, challenges and research directions," in *2023 International Symposium on Networks, Computers and Communications, ISNCC 2023*, Oct. 2023, pp. 1–7, doi: 10.1109/ISNCC58260.2023.10323974.

[18]   M. H. AlKaabi, N. M. Alqbailat, A. Badah, I. A. Ismail, and K. B. Hicham, "Examining the cultural connotations in human and machine translations: a corpus study of Naguib Mahfouz's Zuqāq al-Mida," *Journal of Language Teaching and Research*, vol. 15, no. 3, pp. 707–718, May 2024, doi: 10.17507/jltr.1503.03.

[19]   F. Khoshafah, "ChatGPT for Arabic-English translation: evaluating the accuracy," *Research Square (Preprint)*, pp. 1–18, Apr. 2023, doi: 10.21203/rs.3.rs-2814154/v2.

[20] A. O. El-Shangiti, F. Alwajih, and M. Abdul-Mageed, "Arabic automatic story generation with large language models," in *ArabicNLP 2024 - 2nd Arabic Natural Language Processing Conference, Proceedings of the Conference*, 2024, pp. 140–152, doi: 10.18653/v1/2024.arabicnlp-1.13.

[21] M. E. G. Beheitt and M. B. H. Hmida, "Automatic Arabic poem generation with GPT-2," in *International Conference on Agents and Artificial Intelligence*, 2022, vol. 2, pp. 366–374, doi: 10.5220/0010847100003116.

[22] D. Antar, "The effectiveness of using ChatGPT4 in creative writing in Arabic: poetry and short story as a model," *Information Sciences Letters*, vol. 12, no. 12, pp. 2445–2459, Dec. 2023, doi: 10.18576/isl/121207.

[23] M. M. E. Amin, "Evaluation of automated Arabic summaries generated with the AI tool ChatGPT using textual linguistics," *Journal of Scientific Research in Arts*, vol. 25, no. 5, pp. 46–77, Jul. 2024, doi: 10.21608/jssa.2024.280796.1624.

[24] A. Al-Thubaity *et al.*, "Evaluating ChatGPT and Bard AI on Arabic sentiment analysis," in *ArabicNLP 2023 - 1st Arabic Natural Language Processing Conference, Proceedings*, 2023, pp. 335–349, doi: 10.18653/v1/2023.arabicnlp-1.27.

[25] F. Alderazi, A. Algosaibi, M. Alabdullatif, H. F. Ahmad, A. M. Qamar, and A. Albarrak, "Generative artificial intelligence in topicsentiment classification for Arabic text: a comparative study with possible future directions," *PeerJ Computer Science*, vol. 10, p. e2081, Jul. 2024, doi: 10.7717/peerj-cs.2081.

[26] K. Kadaoui *et al.*, "TARJAMAT: evaluation of Bard and ChatGPT on machine translation of ten Arabic varieties," in *ArabicNLP 2023 - 1st Arabic Natural Language Processing Conference, Proceedings*, 2023, pp. 52–75, doi: 10.18653/v1/2023.arabicnlp-1.6.

BIOGRAPHIES OF AUTHORS

Ali Mousa AlSbou Ⓘ 🔍 SC ⓒ received the B.Sc. degree in computer science from the University of Mu'tah, Jordan, and the M.Sc. in information technology from the Universiti Utara Malaysia, Malaysia. He is a lecturer at the College of Information Technology, Al-Hussein Bin Talal University, Jordan. His research area is machine learning and information science. He can be contacted at email: alinaimat@ahu.edu.jo.

Fadzli Syed Abdullah Ⓘ 🔍 SC ⓒ is an associate professor at the Faculty of Science Computer and Mathematics, Universiti Malaysia Terengganu (UMT). He has an M.Sc. in software engineering from Universiti Teknologi Malaysia and a Ph.D. in computational engineering from Cardiff University, United Kingdom. He has made his research contribution in software engineering, embedded programming, robotic manipulations, and internet of things. Most of his current research interests include autonomous controls, smart farming, and IoT-based systems. His latest project is a smart farming management and controlling system based on IoT and artificial intelligence. He can be contacted at email: efadzli@umt.edu.my.

Ashanira Mat Deris Ⓘ 🔍 SC ⓒ is a senior lecturer at Universiti Malaysia Terengganu (UMT), serving in this role since 2021. She holds a Ph.D. in computer science from Universiti Teknologi Malaysia (UTM) and has been actively involved in academia since 2018. Her primary research interests lie in the fields of machine learning, with a focus on regression and classification techniques, as well as modeling and optimization. Throughout her academic career, she has contributed extensively to the scientific community, having published over 20 journal articles, showcasing her expertise and commitment to advancing knowledge in her areas of specialization. She can be contacted at email: ashanira@umt.edu.my.