The Role Of Natural Language Processing In AI-Based Interpretation Analysis: Opportunities And Limitations

Aan Najib[1*], Mohammad Kurjum[2]
[1]Universitas Islam Negeri Sunan Ampel, Surabaya, Indonesia: aan.najib@uinsa.co.id
[2]Universitas Islam Negeri Sunan Ampel, Surabaya, Indonesia: mkurjum@uinsa.co.id

Abstract: *To identify the opportunities and limitations of NLP implementation in understanding and interpreting Qur'anic verses, especially in the aspect of semantic and syntactic analysis. This research uses the literature review method with a descriptive-analytical qualitative approach. Data were obtained from various academic sources, including Scopus-indexed scientific journals, international conference proceedings, and academic books that discuss NLP, AI, and Qur'anic interpretation. The data collection technique was conducted through a literature study by searching academic databases such as Scopus, IEEE Xplore, SpringerLink, and Google Scholar using related keywords. The collected data were analysed using the content analysis method to identify patterns, trends, and challenges in the application of NLP in the interpretation of the Qur'an. The results show that NLP has great potential in assisting the decoding, translation, and semantic and syntactic analysis of Qur'anic texts. Language models such as Arabert and datasets such as MASAQ have been used to improve interpretation accuracy. However, the research also found significant challenges, including data bias, limitations in capturing historical and cultural contexts, and ethical issues in AI interpretive authority. This study encourages the development of NLP models that are more sensitive to the linguistic and theological context of the Qur'an and the need for active involvement of exegetes in the AI development process to avoid interpretive bias. In addition, this study recommends improving the quality of the dataset by including annotations from various sources of interpretation to ensure diversity of perspectives in interpretation. The originality of this study lies in its in-depth exploration of the NLP approach in Qur'anic exegesis and critical analysis of the challenges and limitations faced in its application. Unlike previous studies that focussed more on technical applications, this research also examines the ethical and methodological implications in the translation and interpretation of sacred verses.*

Keywords: *Natural Language Processing, Tafsir Al-Qur'an, Artificial Intelligence, AI, Text Analysis.*

INTRODUCTION

Advances in artificial intelligence (AI) technology, particularly in the fields of machine learning and natural language processing (NLP), have revolutionised Qur'ānic text analysis by providing innovative tools for deeper understanding and wider accessibility. The use of NLP enables a range of applications, such as syntactic parsing and thematic analysis, that support both academic research and personal study of the Qur'ān (Sawalha et al., 2025). One of the main datasets used in this analysis is MASAQ, which includes more than 131,000 morphological entries and 123,000 syntactic functions, enabling various tasks such as dependency parsing and grammar analysis (Khidhir, 2024).

However, the integration of AI in Qur'anic studies poses various ethical challenges, especially in relation to algorithmic bias and potential distortion of the meaning of the sacred text. Therefore, a framework aligned with Islamic values is needed to ensure that these technologies do not contradict the interpretations that have been agreed upon in the Islamic scholarly tradition (Zuhri et al., 2024). Collaboration between technology experts and scholars is crucial in maintaining interpretative authority and the validity of AI-based analyses (Fitryansyah & Fauziah, 2024). In addition, the transparency and interpretability of AI models are also important factors, especially in their application in broader fields, such as clinical decision support systems (CDSS), where the clarity of the relationship between inputs and outputs determines user confidence (Xu et al., 2023).

In a broader context, NLP plays an important role in a wide range of applications, including sentiment analysis, machine translation, and natural language interfaces (NLI), which are increasingly democratising access to data and information (Singh et al., 2024). However, the challenges faced in the

development of NLP are still considerable, such as semantic ambiguity in language processing and the limited quality of training data that can affect the accuracy of the model (Chen et al., 2024).

This research offers a new contribution to the study of artificial intelligence-based tafsir by exploring in depth the role of Natural Language Processing (NLP) in the analysis of Qur'anic texts. Different from previous studies that focus more on morphological and syntactic aspects (Khidhir, 2024), this research focuses on how NLP can improve the effectiveness of understanding the meaning and context of Qur'anic verses through semantic and deep learning approaches. In addition, this research highlights the limitations and challenges in applying NLP to sacred texts, such as algorithmic biases and potential distortions of meaning, which are still rarely examined in the existing literature (Zuhri et al., 2024). By integrating the perspectives of technology and Islamic studies, this research also proposes a framework that can be used to ensure that interpretations generated by NLP models remain aligned with established principles of tafsir. Furthermore, this study discusses the potential adaptation of deep learning models, such as BERT and other transformers, in improving the interpretation of verses that have metaphorical meanings or complex historical contexts (Reggad et al., 2023). This innovation opens up opportunities to develop a more accurate and contextualised Qur'an-based question-and-answer system that can be utilised by academics, researchers, and the general public. Thus, the novelty of this research lies in the interdisciplinary approach that combines NLP, Qur'anic exegesis, and ethical principles in the utilisation of AI, as well as the development of a more comprehensive methodology in semantic analysis of sacred texts.

THEORETICAL FRAMEWORK

Recent research has shown that the application of neural networks in the syntactic parsing of Quranic sentences can achieve an accuracy rate of up to 96%, indicating the effectiveness of NLP in understanding the complexity of Arabic grammar (Khidhir, 2024). In addition, an NLP-based semantic approach has been applied to interpret metaphorical language in the Quran, which is a crucial aspect in understanding the text in contemporary social and cultural contexts (Chukhanov & Kairbekov, 2024). Furthermore, deep learning-based models, such as BERT, have been adapted in Quran-based question-and-answer systems, which show significant improvements in contextual understanding of religious texts (Reggad et al., 2023).

In addition to the analytical aspect, the digitisation of the Quran through NLP has resulted in various tools that increase accessibility to this sacred text, allowing a wider audience, including non-specialists, to interact with the Quran in a more interactive form (Bashir et al., 2023). Interdisciplinary studies in this field also underline the importance of an in-depth understanding of the linguistic and literary aspects of the Quran in order to ensure that its meaning remains in accordance with established rules of tafsir (Wajdi et al., 2023).

METHOD

This research uses the literature review method to examine the role of Natural Language Processing (NLP) in artificial intelligence (AI)-based Qur'anic interpretation analysis. This approach aims to identify opportunities and limitations of NLP implementation in understanding and interpreting Qur'anic verses.

This type of research is qualitative with a descriptive-analytical approach. This research will examine various sources of scientific literature, including Scopus indexed journals, international conference proceedings, and academic books relevant to the topics of NLP, Qur'anic interpretation, and AI. The literature reviewed includes previous research on the implementation of NLP in Arabic text processing, interpretation of Qur'anic verses, and ethical and technical challenges in the use of AI in Islamic studies.

The data sources used in this study were sourced from academic publications, including scholarly journals, reference books, and research reports that discuss NLP, AI, and Qur'anic interpretation. The main literature analysed included studies on Qur'anic morphological and syntactic datasets such as MASAQ, as well as AI models applied in semantic and syntactic analysis of religious texts.

The data collection technique was conducted using the library research method. Data was collected by searching various academic sources using the keywords 'Natural Language Processing,' 'interpretation

of the Qur'an,' 'AI,' and 'Qur'an' in journal databases such as Scopus, IEEE Xplore, SpringerLink, and Google Scholar. Source selection was based on relevance to the research topic as well as the citation rate and academic quality of the published journals.

The data were analysed using the content analysis method, in which the results of previous studies were categorised and compared to identify patterns, trends and challenges in the application of NLP to Qur'anic exegesis. The analysis was conducted through several systematic stages. First, the literature was categorised based on the aspects under study, such as the application of Natural Language Processing (NLP) in syntactic and semantic analysis of the Qur'ān, the effectiveness of artificial intelligence models in understanding texts, and the technical and ethical challenges faced. Next, a thematic analysis was conducted by classifying the results of the study based on the main themes, such as the accuracy of the NLP model in understanding the structure of Qur'anic Arabic, the limitations of the available dataset, and the implications of the application of artificial intelligence in Islamic studies. The final stage involved a critique and synthesis of the analysed literature, comparing the various studies to identify gaps in existing research and provide recommendations regarding the development of NLP in exegetical studies.

FINDINGS AND DISCUSSIONS

Natural Language Processing (NLP) applied in AI-based Qur'anic interpretation analysis

Machine learning techniques have undergone many advancements in the last ten years (Septiyandini et al., 2024). These advancements include the development of more sophisticated algorithms, increased computing capacity, and wider availability of data. One area that has been greatly impacted by these developments is natural language processing (NLP). AI technologies, particularly NLP, enable the parsing, translation and thematic analysis of Qur'anic texts, making them more accessible and interpretable. However, the integration of AI in this domain presents ethical and interpretive challenges that require special attention.

One of the main benefits of AI in Qur'ānic interpretation is its ability to parse and analyse texts (Yatri et al., 2023). An AI model has been developed to parse sentences in the Qur'an, with an accuracy rate of 96% in understanding complex Arabic structures (Khidhir, 2024). In addition, AI plays a role in translating complex linguistic features, such as wordplay and hidden meanings, although human intervention is still required to capture more complex depths of meaning (Albalawi & Ghafour, 2024). It also improves accessibility and learning, enabling personalised learning experiences as well as AI-based interactive education (Mauluddin, 2024).

However, the application of AI in Qur'anic interpretation raises questions about interpretive authority. AI's judgement in interpreting religious texts may pose a risk of distortion of meaning if not carefully monitored (Fitryansyah & Fauziah, 2024). In addition, AI often has limitations in understanding the cultural sensitivities and historical contexts inherent in Qur'anic verses, requiring collaboration between technologists and religious scholars to ensure accuracy of interpretation (Fitryansyah & Fauziah, 2024). Thus, while AI offers promising tools in this field, human supervision is still necessary to maintain the integrity of religious knowledge.

One important aspect of Qur'ānic interpretation is the concept of *munasabah*, the inter-verse and inter-surah relationships that form a unity of meaning. NLP can enhance the understanding of munasabah by analysing the complex relationships between verses and surahs, which is very important in Qur'anic interpretation. Munasabah includes various aspects, such as logical relationships, cause-and-effect dynamics, and thematic connections that support a unified understanding of the Qur'ān (Elkarimah, 2023).

To support these analyses, various NLP tools and comprehensive datasets have been developed. The Morphologically Analysed and Syntactically Annotated Quran (MASAQ) dataset provides extensive syntactic and morphological annotations, enabling more in-depth analysis of Qur'anic texts (Sawalha et al., 2025). In addition, language models such as Arabert have shown high accuracy in classifying inter-verse semantic relations, which further strengthens the potential of NLP in understanding verse interrelationships (Alsaleh et al., 2021). Nonetheless, challenges remain, especially in ensuring that AI

systems capture not only linguistic relationships but also the deep meanings inherent in the sacred texts. Hence, the involvement of exegetes remains necessary to interpret the more complex nuances of munasabah.

AI also has the potential to distinguish between literal and metaphorical meanings in the Qur'ān through advanced NLP techniques. The integration of syntactic and semantic analysis allows AI to recognise multiple layers of meaning in the Qur'ānic language. Datasets such as MASAQ provide over 131K morphological entries and 123K syntactic functions, supporting more in-depth linguistic analyses (Sawalha et al., 2025). This enables dependency parsing as well as more accurate grammar checking in understanding sacred texts.

In addition, an NLP-based semantic search has been developed to identify verses relevant to the user's question, thus enhancing contextual understanding of a verse (Sawalha et al., 2025). Models like Arabert's have also shown success in classifying semantic links between verses, which is helpful in recognising metaphorical language in the Qur'ān (Alsaleh et al., 2021). Nonetheless, the role of human expertise remains necessary, as AI can only provide basic interpretation. Post-editing by expert interpreters is crucial in capturing the depth of meaning, especially in complex linguistic features such as wordplay (Albalawi & Ghafour, 2024). The synergy between AI and human translators is key to achieving a balance between the accuracy of the analysis and the aesthetic richness of the sacred text.

However, data bias in the training of NLP models poses a significant challenge in the interpretation of the Qur'ān by AI. When models are trained on biased data sets, they can perpetuate those biases in their output, potentially resulting in inaccurate interpretations. One common form of bias is cultural bias, where language models tend to be more attuned to Western cultural contexts compared to non-Western texts such as the Qur'ān (Naous & Xu, 2025). This bias arises due to the dominance of Western culture-focussed data in the training set. In addition, data bias can also lead to the propagation of false stereotypes regarding Islamic teachings, as the model can reproduce the social biases present in its training data (Raza et al., 2024).

Furthermore, political bias in pre-training data can lead to distorted interpretations, where AI models reflect the polarised views of their training sources (Feng et al., 2023). This poses serious challenges in ensuring that AI remains neutral and does not provide interpretations influenced by specific political agendas (Afandi et al., 2023). While there is a view that AI can provide more diverse interpretations of the Qur'ān by integrating various scientific perspectives, the risk of bias remains a major concern. Therefore, strict curation of the training dataset is required to ensure fair and accurate representation in AI-based Qur'ānic interpretations. Thus, while NLP and AI offer advances in Qur'ānic analysis and interpretation, ethical, linguistic and data bias challenges remain to be addressed. The synergy between AI and human expertise is key in ensuring that the resulting Qur'ānic interpretations remain accurate, meaningful, and in line with established principles of tafsir in the Islamic tradition.

Table 1. Application of NLP in AI-based Quranic Tafsir Analysis

Aspects	Methods/Technology	Implementation & Results	References
Crowdsourcing for Linguistic Annotation	Crowdsourcing platform (Amazon Mechanical Turk, a dedicated Arabic NLP platform)	Improved Arabic annotation accuracy, but complex tasks such as grammar are still challenging	(Zaghouani & Dukes, 2014); (Alahmadi et al., 2020)
Audio Dataset Collection	Crowdsourcing Quran recitation recordings from non-Arabic speakers	Generated 7,000+ recordings to train AI models in Quran recitation learning	(Salameh et al., 2024)
Quran Translation Memory	Crowdsourcing term base and translation memory	Assisted in Quran-specific translation tools	(Mollanazar & Tayyebi, 2019)

Aspects	Methods/Technology	Implementation & Results	References
Voice Recognition for Quran Recitation Verification	Wavelet Packet Decomposition (WPD) & MFCC	Used to automatically assess Quran recitation	(Alvindo et al., 2024)
Blockchain for Quran Text Authenticity	Hyperledger Fabric Framework	Ensures Quran text integrity, prevents digital manipulation	(Hanif et al., 2024)
Deep Learning in Quran Authentication	LSTM & Word Embedding	Verse verification accuracy up to 99.98%	(Touati-Hamad et al., 2021); (Touati-Hamad et al., 2022)
Quran Morphological and Syntactic Analysis	MASAQ dataset (131K morphological entries, 123K syntactic functions)	Supports dependency parsing and grammar analysis	(Sawalha et al., 2025)
Quran Interpretation Semantic Search	Cosine Similarity	Achieves 0.97 search accuracy in finding relevant commentaries	(Shohoud et al., 2023)
Quran Sentence Parsing	Neural networks in syntax analysis	Parsing accuracy reaches 96%	(Khidhir, 2024)
Verse Classification Based on Difficulty	SVM (Support Vector Machine)	Effective in classifying verses by difficulty level	(Pambudi & Lhaksmana, 2024)
Ontology in Quran Semantic Search	Ontology-based query expansion	Improves search results by considering semantic relationships between terms	(Beirade et al., 2021)
Text Summarisation and QA of Interpretation System	TextRank, Cosine Similarity, AyaTEC Dataset	Helps users get a summary of tafsir and a Quran-based question-and-answer system	(Fakhrezi et al., 2021); (Malhas et al., 2022)

Effectiveness of NLP in understanding the meaning and context of Qur'anic verses
The utilisation of Natural Language Processing (NLP) in AI-based exegetical analysis has shown rapid development, especially in the aspects of semantic search, ontology processing, and the application of machine learning in the understanding of Qur'anic texts. In this study, various methods have been developed to overcome the limitations of keyword-based search systems and improve the understanding of Qur'anic texts through NLP approaches.

Table 2. Application of NLP and AI in Understanding and Searching Qur'anic Texts

No	Researchers	Issues Discussed	Techniques Applied/Contribution
1	(Sherif & Ngonga Ngomo, 2015)	Semantics of words in the Qur'ān	Development of Qur'ānic semantic database and integration with online platforms

No	Researchers	Issues Discussed	Techniques Applied/Contribution
2	(Afzal & Mukhtar, 2019)	Limitations of keyword-based search in the Qur'ān	Development of QEWN database and Qur'ānic vocabulary-based VQC search system
3	(Alqahtani & Atwell, 2015)	Weaknesses of existing Qur'ānic search systems	Development of QSST for semantic + keyword-based search
4	(E. H. Mohamed & Shokry, 2022)	Improved semantic search results	Embedding matrix based search system
5	(Alshammeri et al., 2021)	Richer semantic search results	Verse vectorisation with Doc2Vec
6	(Fakhrezi et al., 2021)	Lack of summary answers in Qur'ānic topics	Use of text rank algorithm with cosine similarity based search
7	(Malhas & Elsayed, 2022)	Absence of benchmark datasets for Qur'anic QA systems	Development of AyaTEC dataset and evaluation of verse-based question-answer system
8	(Beirade et al., 2021)	Semantic search in the Qur'an	Ontology-based query expansion technique
9	(Khazani et al., 2021)	Semantic search in the Qur'an	Development of semantic graph framework
10	(Alargrami & Eljazzar, 2020)	Semantic system for Islamic knowledge	Use of word2vec for embedding Islamic vocabulary
11	(ELAffendi et al., 2021)	POS prediction in Qur'anic text	Use of GPOW2 with neural network
12	(R. Mohamed et al., 2015)	Machine learning in Qur'anic text processing	Supervised learning with decision tree
13	(Ahmed & Atwell, 2016)	Ontology-based knowledge extraction	Fusion of various Qur'anic ontologies for knowledge base development

Several studies highlighted the limitations of keyword-based search systems that often fail to capture contextual meanings in the Qur'ān. Sherif and Ngonga (2015) focused on developing a semantic database to improve the accessibility of tafsir with the integration of online platforms. A similar concept was applied by Afzal and Mukhtar (2019) who proposed the QEWN database and VQC search system, allowing users to perform Qur'anic vocabulary-based searches with more accurate results. However, keyword-based search still has limitations in capturing semantic relationships between verses. To address this, Alqahtani and Atwell (2015) developed QSST, a system that combines semantic and keyword search, enriching the search results with broader contextual understanding. Meanwhile, Ensaf and Eyad (2022) and Menwa et al. (2021) proposed a method based on matrix embedding and Doc2Vec vectorisation, which enables search by considering the vector representation of Qur'anic verses. With this approach, the system not only matches words literally, but is also able to identify meaning relationships between verses, a step forward in AI-based text search. Nonetheless, the challenges faced are the dependence on the size of the training dataset and the embedding accuracy in capturing complex theological contexts.

In addition to text search, NLP also plays a role in the automatic generation of commentary summaries and Qur'anic text-based question-and-answer systems (Fakhrezi et al., 2021) applying text rank and cosine similarity algorithms, allows the system to extract relevant information and present it as a more structured summary. This provides efficiency for users seeking a quick interpretation of a particular verse. However, the main problem with this system is the limited understanding of contextual meaning and

adapting to various interpretations. To address this, Malhas and Elsayed (2022) developed the AyaTEC dataset, which acts as a benchmark Qur'anic question-and-answer system. This dataset allows for a more systematic evaluation of the effectiveness of AI-based search and query systems. This approach provides great opportunities in improving the accessibility of tafsir, but also poses challenges in terms of the accuracy of the answers generated. Given that Qur'anic commentaries have various viewpoints and methods of interpretation, AI systems must be able to accommodate various commentaries without losing their theological nuances.

The application of ontologies in semantic search is increasingly being developed to capture conceptual relationships in Qur'anic texts. Faiza et al. (2021) proposed an ontology-based query expansion technique, which enriches search results by considering semantic relationships between terms in the Qur'an. Meanwhile, Muhammad et al. (2021) developed a semantic graph framework, which helps the system in connecting Qur'anic concepts more accurately.

Besides ontology-based approaches, machine learning also plays an important role in Qur'anic text analysis. Ali and Maged (2020) applied word2vec embedding to develop an AI-based Islamic knowledge system. This model enables the recognition of interayat concept similarity and meaning relationships based on word distribution in the Qur'anic text. In the linguistic aspect, ElAffenid et al. (2021) examined part-of-speech (POS) prediction in Qur'anic text using GPOW2-based neural network, which improved the accuracy of syntactic analysis of Arabic Qur'anic text. Meanwhile, Mohamed et al. (2015) applied supervised learning with decision tree, which helped in classifying verses based on their thematic categories.

In addition, Ahmed and Atwell (2016) proposed an approach of combining Qur'anic ontologies to build a richer knowledge base. With this approach, the system can connect concepts in the Qur'an in a more systematic and structured way. However, the main challenge of ontology-based and machine learning approaches is the need for high-quality datasets for model training, as well as the customisation of algorithms to the classical Arabic language rules used in the Qur'an.

Challenges in the application of Natural Language Processing (NLP) in Qur'ānic exegetical analysis include the unique characteristics of Qur'ānic texts, the need for more specific models, and limitations in semantic understanding. The linguistic complexity of the Qur'ānic text is a major obstacle, as the syntactic and morphological structures of classical Arabic differ from modern Arabic. Al Moaiad et al. (2024) highlighted that a single lemma in Arabic can have many forms with different meanings, which makes it difficult for NLP in semantic parsing and processing. In addition, the use of classical Arabic in the Qur'ān requires specialised models that can capture nuances of grammar and structure not found in modern dialects (Sawalha et al., 2025).

One of the main challenges in NLP for Qur'ānic texts is the problem of diacritisation. The Qur'ān is often written without diacritics, even though these marks play an important role in determining the pronunciation and meaning of words. The absence of diacritics in the text leads to ambiguity in interpretation, which results in errors in machine translation and automatic summarisation (ZubeiriA et al., 2025). This challenge is further compounded by the limited annotated datasets required to train high-accuracy NLP models. For example ElKoumy and Sarhan (2024) revealed that Qur'anic reading comprehension datasets are often under-resourced, which results in limited performance of NLP models in tasks such as question answering and semantic search.

In addition to linguistic limitations, NLP in Qur'ānic analysis also faces challenges from the aspect of semantic modelling. Traditional NLP techniques tend to focus on syntactic analysis without considering the depth of meaning of the text. Bashir et al. (2023) highlighted that NLP systems still often have difficulty in capturing complex semantic relationships in Qur'anic texts, leading to potential misinterpretations. In this context, Large Language Model (LLM)-based models also still have limitations due to the tendency to perpetuate biases and inaccuracies in religious contexts (Patel et al., 2023). Therefore, it is necessary to develop domain-specific models that are faithful to Islamic teachings and based on high-quality datasets in order to represent the meaning of texts more accurately.

Another challenge in NLP for Qur'anic analysis is the interpretive gap in text understanding. Bakr et al. (2023) noted that existing NLP tools are often unable to capture deep semantic aspects, which can

result in an inaccurate understanding of the Qur'ānic content. To address this, various new semantic paradigms have been proposed to tailor interpretations to the intrinsic and extrinsic meanings of the text (Ismail, 2024). This effort reflects the importance of collaboration between NLP experts and Islamic scholars to ensure that the system developed is not only technically accurate, but also in accordance with the principles of Qur'anic interpretation.

Although challenges remain, the development of datasets and NLP models specifically for Qur'anic analysis continues to make progress (Sawalha et al., 2025) highlighted that morphologically analysed datasets and syntactic annotations can contribute to improving NLP capabilities in this domain. With interdisciplinary collaboration and the development of more advanced technologies, NLP has the potential to become a more effective tool in understanding and interpreting the Qur'ān more accurately.

CONSLUSION

Based on the research results, it can be concluded that Natural Language Processing (NLP) plays an important role in artificial intelligence (AI)-based tafsir analysis by offering various opportunities while facing a number of limitations. NLP enables the decoding of complex Arabic structures in the Qur'ān with a high degree of accuracy, increasing the accessibility of the sacred text through more precise translation and semantic search. In addition, it supports the understanding of inter-verse relationships (munasabah) as well as distinguishing literal and metaphorical meanings in texts, thus enriching the study of Qur'ānic exegesis.

However, this research also revealed some key challenges in applying NLP to Qur'anic exegesis. One of these is the AI's limitation in understanding cultural sensitivities and deep historical context, which still requires the intervention of an exegete to ensure accurate interpretation. In addition, biases in the training data can affect the analysis results, potentially resulting in interpretations that are not neutral or deviate from established theological understandings. Other issues include the limitations of NLP in handling the diacritisation of classical Arabic texts as well as challenges in capturing more complex semantic relationships. Thus, while NLP offers innovations in AI-based tafsir analysis, this research confirms that a hybrid approach that combines artificial intelligence with human expertise is still needed. The synergy between NLP and traditional exegetical sciences will be key in ensuring that Qur'ānic interpretation remains accurate, contextualised and in line with established religious principles.

This research has significant implications in the development of NLP for AI-based tafsir analysis, both in terms of academic, practical, ethical, and policy. Academically, this research enriches the study of Islamic computational linguistics and emphasises the need to develop NLP models that are more suited to classical Arabic. Practically, the integration of AI in tafsir requires the supervision of scholars and linguistic experts to ensure accurate and contextualised interpretations. Ethically, the risk of algorithm bias and distortion of meaning demands regulation and ethical standards in the use of AI for religious texts. In addition, this research encourages the development of more representative datasets and policies that govern the use of AI in tafsir to remain in line with Islamic values. Therefore, the synergy between NLP and human expertise is key to ensure that AI acts as a tool that enhances the understanding of the Qur'an without replacing the authority of traditional interpretation.

REFERENCES

1. Afandi, Y., Maskur, M., Fiernaningsih, N., & Fauzi, A. (2023). Semantic And Natural Language Processing Development Applications For Chatbots To Enhance Online Store Customer Service. *INTERNATIONAL JOURNAL ON ADVANCED TECHNOLOGY ENGINEERING AND INFORMATION SYSTEM (IJATEIS)*, *2*(4). https://doi.org/10.55047/ijateis.v2i4.986
2. Afzal, H., & Mukhtar, T. (2019). Semantically enhanced concept search of the Holy Quran: Qur'anic English WordNet. *Arabian Journal for Science and Engineering*, *44*, 3953–3966. https://doi.org/10.1007/s13369-018-03709-2
3. Ahmed, R., & Atwell, E. S. (2016). Developing an ontology of concepts in the Qur'an. *International Journal on Islamic Applications in Computer Science And Technology*, *4*(4), 1–8. https://doi.org/https://eprints.whiterose.ac.uk/112967/1/ahmed16ijasat.pdf
4. Al Moaiad, Y., Alobed, M., Alsakhnini, M., & Momani, A. M. (2024). Challenges in natural Arabic language

processing. *Edelweiss Applied Science and Technology*, *8*(6), 4700–4705.
https://doi.org/10.55214/25768484.v8i6.3018

5. Alahmadi, D., Babour, A., Saeedi, K., & Visvizi, A. (2020). Ensuring inclusion and diversity in research and research output: A case for a language-sensitive NLP crowdsourcing platform. *Applied Sciences*, *10*(18), 6216. https://doi.org/10.3390/APP10186216

6. Alargrami, A. M., & Eljazzar, M. M. (2020). Imam: Word Embedding Model for Islamic Arabic NLP. *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, 520–524. https://doi.org/10.1109/NILES50944.2020.9257931

7. Albalawi, A. F., & Ghafour, A. Q. K. A. (2024). Unravelling Qur'anic Arabic Puns Through AI English Translation And The Art Of Human Post-Editing. *Ijaz Arabi Journal of Arabic Learning*, *7*(2), 680–695. https://doi.org/10.18860/ijazarabi.v7i2.27819

8. Alqahtani, M. M. A., & Atwell, E. (2015). *A review of semantic search methods to retrieve information from the Qur'an corpus*. eprints.whiterose.ac.uk.

9. Alsaleh, A. N., Atwell, E., & Altahhan, A. (2021). Quranic verses semantic relatedness using arabert. *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, 185–190.

10. Alshammeri, M., Atwell, E., & Alsalka, M. ammar. (2021). Detecting semantic-based similarity between verses of the quran with doc2vec. *Procedia Computer Science*, *189*, 351–358. https://doi.org/10.1016/j.procs.2021.05.104

11. Alvindo, F., Setiawan, N. A., & Nugroho, H. A. (2024). Al-Quran Recitation or Sound Processing Analysis: A Systematic Literature Review on Methods. *2024 International Conference on Electrical Engineering and Computer Science (ICECOS)*, 54–58. https://doi.org/10.1109/icecos63900.2024.10791189

12. Bakr, A., Yousef, A. H., & Arafa, T. (2023). Specialized Syntactic Quran Search Engines: Evaluation and Limitations. *2023 Intelligent Methods, Systems, and Applications (IMSA)*, 269–275. https://doi.org/10.1109/imsa58542.2023.10217550

13. Bashir, M. H., Azmi, A. M., Nawaz, H., Zaghouani, W., Diab, M., Al-Fuqaha, A., & Qadir, J. (2023). Arabic natural language processing for Qur'anic research: a systematic review. *Artificial Intelligence Review*, *56*(7), 6801–6854. https://doi.org/10.1007/s10462-022-10313-2

14. Beirade, F., Azzoune, H., & Zegour, D. E. (2021). Semantic query for Quranic ontology. *Journal of King Saud University-Computer and Information Sciences*, *33*(6), 753–760. https://doi.org/10.1016/j.jksuci.2019.04.005

15. Chen, Y., Wang, H., Yu, K., & Zhou, R. (2024). Artificial intelligence methods in natural language processing: A comprehensive review. *Highlights in Science Engineering and Technology*, *3*(12), 545–550. https://doi.org/10.57185/jetbis.v3i12.164

16. Chukhanov, S., & Kairbekov, N. (2024). The importance of a semantic approach in understanding the texts of the Holy Quran and Sunnah. *Pharos Journal of Theology*, *105*(3), 1–11. https://doi.org/10.46222/pharosjot.105.36

17. ELAffendi, M. A., Abuhaimed, I., & AlRajhi, K. (2021). A simple Galois Power-of-Two real time embedding scheme for performing Arabic morphology deep learning tasks. *Egyptian Informatics Journal*, *22*(1), 35–43. https://doi.org/10.1016/j.eij.2020.03.002

18. Elkarimah, M. F. (2023). Munasabah in the Perspective of Science of the Qur'an: Study of Al-burhan Fi Ulumul Quran Works of Al-zarkasyi (D. 749 H). *Al-Risalah: Jurnal Studi Agama Dan Pemikiran Islam*, *14*(1), 47–61. https://doi.org/10.34005/alrisalah.v14i1.2209

19. ElKoumy, M., & Sarhan, A. (2024). Improving ranking-based question answering with weak supervision for low-resource Qur'anic texts. *Artificial Intelligence Review*, *58*(1), 17. https://doi.org/10.1007/s10462-024-10964-3

20. Fakhrezi, M. F., Bijaksana, M. A., & Huda, A. F. (2021). Implementation of automatic text summarization with TextRank method in the development of Al-qur'an vocabulary encyclopedia. *Procedia Computer Science*, *179*, 391–398. https://doi.org/10.1016/j.procs.2021.01.021

21. Feng, S., Park, C. Y., Liu, Y., & Tsvetkov, Y. (2023). From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair NLP models. *ArXiv Preprint ArXiv:2305.08283*. https://doi.org/10.48550/arXiv.2305.08283

22. Fitryansyah, M. A., & Fauziah, F. N. (2024). Bridging Traditions and Technology: AI in The Interpretation of Nusantara Religious Manuscripts. *Jurnal Lektur Keagamaan*, *22*(2), 317–346. https://doi.org/10.31291/jlka.v22i2.1247

23. Hanif, S., Khan, A., Mustafa, E., Ali, M., Shuja, J., & Hasan, M. H. (2024). Ensuring Integrity of Quranic Verses using Hyperledger Fabric Framework under IBM Blockchain Platform. *2024 IEEE 14th Symposium On Computer Applications & Industrial Electronics (ISCAIE)*, 505–510. https://doi.org/10.1109/iscaie61308.2024.10576345

24. Ismail, A. (2024). Exploring New Arabic Semantic Paradigms For Understanding The Quran Addressing Interpretive Gaps and Hidden Meanings. *Ijaz Arabi Journal of Arabic Learning*, 7(3), 1339–1354. https://doi.org/10.18860/ijazarabi.v7i3.27123

25. Khazani, M. M. M., Mohamed, H., Sembok, T. M. T., Yusop, N. M. M., Wani, S., Gulzar, Y., Halip, M. H. M., Marzukhi, S., & Yunos, Z. (2021). Semantic graph knowledge representation for Al-Quran verses based on word dependencies. *Malaysian Journal of Computer Science*, *Special Issue 2*, 132–153. https://doi.org/10.22452/mjcs.sp2021no2.9

26. Khidhir, A. M. (2024). An AI model for Parsing the Text of Holy Quran Sentences. *Mesopotamian Journal of Quran Studies*, *2024*, 16–23. https://doi.org/10.58496/MJQS/2024/003

27. Malhas, R., & Elsayed, T. (2022). Arabic machine reading comprehension on the Holy Qur'an using CL-AraBERT. *Information Processing & Management*, *59*(6), 103068. https://doi.org/10.1016/j.ipm.2022.103068

28. Malhas, R., Mansour, W., & Elsayed, T. (2022). *Qur'an QA 2022: Overview of The First Shared Task on Question Answering over the Holy Qur'an*. qspace.qu.edu.qa.

29. Mauluddin, M. (2024). Kontribusi Artificial Intelligence (AI) pada Studi Al Quran di Era Digital; Peluang dan Tantangan. *Madinah: Jurnal Studi Islam*, *11*(1), 99–113. https://doi.org/10.58518/madinah.v11i1.2518

30. Mohamed, E. H., & Shokry, E. M. (2022). QSST: A Quranic Semantic Search Tool based on word embedding. *Journal of King Saud University-Computer and Information Sciences*, *34*(3), 934–945. https://doi.org/10.1016/j.jksuci.2020.01.004

31. Mohamed, R., Ragab, M., Abdelnasser, H., El-Makky, N. M., & Torki, M. (2015). *Al-Bayan: A knowledge-based system for Arabic answer selection*. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015).

32. Mollanazar, H., & Tayyebi, A. (2019). *A Model for Crowdsourcing Development of Databases for Qur'anic Studies Sources*. Iranian Journal of Translation Studies. https://journal.translationstudies.ir/ts/article/view/692

33. Naous, T., & Xu, W. (2025). On The Origin of Cultural Biases in Language Models: From Pre-training Data to Linguistic Phenomena. *ArXiv Preprint ArXiv:2501.04662*. https://doi.org/10.48550/arxiv.2501.04662

34. Pambudi, A. A., & Lhaksmana, K. M. (2024). Identifying Difficult Quran Verses Using SVM, LSTM, and CNN. *In 2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD)*, 43–48. https://doi.org/10.1109/icoabcd63526.2024.10704332

35. Patel, S., Kane, H., & Patel, R. (2023). Building domain-specific llms faithful to the islamic worldview: Mirage or technical possibility? *ArXiv Preprint ArXiv:2312.06652*. https://doi.org/10.48550/arxiv.2312.06652

36. Raza, S., Garg, M., Reji, D. J., Bashir, S. R., & Ding, C. (2024). Nbias: A natural language processing framework for BIAS identification in text. *Expert Systems with Applications*, *237*, 121542. https://doi.org/10.1016/j.eswa.2023.121542

37. Reggad, S., Ghadi, A., El Aachak, L., & Samih, A. (2023). Machine Reading Comprehension for the Holy Quran: A Comparative Study. *The Proceedings of the International Conference on Smart City Applications*, 419–428. https://doi.org/10.1007/978-3-031-54376-0_38

38. Salameh, R., Al Mdfaa, M., Askarbekuly, N., & Mazzara, M. (2024). Quranic Audio Dataset: Crowdsourced and Labeled Recitation from Non-Arabic Speakers. *Procedia Computer Science*, *246*, 2684–2693. https://doi.org/10.48550/arxiv.2405.02675

39. Sawalha, M., Al-Shargi, F., Yagi, S., AlShdaifat, A. T., Hammo, B., Belajeed, M., & Al-Ogaili, L. R. (2025). Morphologically-analyzed and syntactically-annotated Quran dataset. *Data in Brief*, *58*, 111211. https://doi.org/10.1016/j.dib.2024.111211

40. Septiyandini, W., Muriman, C., & Mayastinasari, V. (2024). The Impact of Artificial Intelligence (AI) on Human Resources: A Case Study of the Indonesian Police Institution. *Policy, Law, Notary and Regulatory Issues*, *4*(1), 64–74. https://doi.org/10.55047/polri.v4i1.1540

41. Sherif, M. A., & Ngonga Ngomo, A. C. (2015). Semantic Quran. Semantic Web. *Semantic Web*, *6*(4), 339–345.

42. Shohoud, Y., Shoman, M., & Abdelazim, S. (2023). Quranic Conversations: Developing a Semantic Search tool for the Quran using Arabic NLP Techniques. *ArXiv Preprint ArXiv:2311.05120*. https://doi.org/10.48550/arxiv.2311.05120

43. Singh, K. R., Rajarajeswari, S., Kadam, R. S., Kumar, R., Agrawal, T., & Raghavendra, R. (2024). Exploring the Impact of AI-Driven Natural Language Processing in Computational Analysis. *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–6. https://doi.org/10.1109/ICCCNT61001.2024.10724482

44. Touati-Hamad, Z., Laouar, M. R., & Bendib, I. (2021). Authentication of quran verses sequences using deep learning. *2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI)*, 1–4. https://doi.org/10.1109/ICRAMI52622.2021.9585935

45. Touati-Hamad, Z., Laouar, M. R., Bendib, I., & Hakak, S. (2022). Arabic quran verses authentication using

deep learning and word embeddings. *The International Arab Journal of Information Technology*, *19*(4), 681–688. https://doi.org/10.34028/iajit/19/4/13

46. Wajdi, M. B. N., Husniyah, N. I., Ahmad, V. I., Putra, M. I. S., & Syarofi, A. M. (2023). Understanding the Quran Holistically: Interdisciplinary Study of the Language and Linguistics of the Quran. *BASA Journal of Language & Literature*, *3*(1), 11–17. https://doi.org/10.33474/basa.v3i1.19596

47. Xu, Q., Xie, W., Liao, B., Hu, C., Qin, L., Yang, Z., Xiong, H., Lyu, Y., Zhou, Y., & Luo, A. (2023). Interpretability of clinical decision support systems based on artificial intelligence from technological and medical perspective: A systematic review. *Journal of Healthcare Engineering*, *2023*(1), 9919269. https://doi.org/10.1155/2023/9919269

48. Yatri, D., Anugerahwati, M., & Setyowati, L. (2023). Artificial Intelligence (Ai) In Language Learning (English And Arabic Class): Students' and Teachers' experience And Perceptions. *Transformational Language Literature and Technology Overview in Learning (Transtool)*, *3*(1), 1–12. https://doi.org/10.55047/transtool.v3i1.1338

49. Zaghouani, W., & Dukes, K. (2014). *Can crowdsourcing be used for effective annotation of Arabic?* LREC.

50. ZubeiriA, I., Souri, D., & Mohajir, B. E. El. (2025). Arabic text diacritization using transformers: a comparative study. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *14*(1), 702–711. https://doi.org/10.11591/ijai.v14.i1.pp702-711

51. Zuhri, M. T., Sahlani, L., & Munawaroh, N. (2024). The Ethics of Artificial Intelligence (AI) Utilization in Qur'anic Studies: An Islamic Philosophical Perspective. *Asyahid Journal of Islamic and Quranic Studies (AJIQS)*, *6*(2), 1–17. https://doi.org/10.62213/b6hexr21