Developing A Quranic QA System: Bridging Linguistic Gaps in Urdu Translation Using NLP and Transformer Model

Muhammad Tariq[1], Dr. Muhammad Arshad Awan[1], Danish Khaleeq[1]
[1] Department of Computer Science, Allama Iqbal Open University (AIOU), Islamabad
* Correspondence: tibeeb@gmail.com

The limited access to Quranic knowledge for Urdu speakers is due to inadequate Natural Language Processing (NLP) tools, which hinder precise Quranic understanding and retrieval. This research introduces a Transformer-based Urdu Quranic Question-Answering (QA) system, a novel approach that enhances semantic accuracy and retrieval precision, unlike existing Arabic- and English-based models. This study primarily leverages Transformer-based technology to develop a context-aware Urdu Quranic chatbot, unlike conventional systems, which primarily support Arabic and English Quranic texts. The system addresses the missing linguistic gaps in Quranic QA by enhancing both precision and semantic interpretation for Urdu users. The system was trained using Fateh Muhammad Jalandhari's Urdu Quranic translation and fine-tuned with RoBERTa for enhanced semantic text analysis. It integrates TF-IDF with SBERT for improved question-answering performance. The NLP system went through multiple evaluation metrics were used to assess its precision and overall capability. The chatbot achieved high retrieval accuracy with a Mean Average Precision of 0.85, an Exact Match of 0.82, and an F1 Score of 0.88. User satisfaction reached 92%, indicating its effectiveness in providing precise Quranic answers. Future updates will introduce that include voice detection features, expanded language support, and integration with Tafsir and Hadith databases for improved contextual understanding. This study enhances Urdu Quranic information retrieval by providing an improved NLP-based solution for automated Islamic knowledge dissemination.

Keywords: Quranic Question-Answering (QA); Urdu Natural Language Processing (NLP); Transformer-based Models; RoBERTa; Semantic Search

Introduction:

Question Answering (QA) is a specialized field of NLP that focuses on extracting precise answers to user inquiries from written documents. The development of English and Arabic language QA systems has progressed through innovative machine learning together with deep learning models over multiple years [1]. The domain of QA has not been fully explored in Urdu as a language especially when focusing on Quranic studies. Research on Arabic-based Quranic Question Answering (QQA) systems continues to progress [2], while Urdu language users still lack an effective solution for accurately answering religious inquiries. Transformers like BERT and RoBERTa together with SBERT have developed context comprehension through recent advancements which lead to enhanced semantic retrieval capabilities [3]. The application of AraBERT and SBERT for Quranic NLP led to better retrieval quality in Arabic Quranic QA systems as reported in [4]. These models need adaptation to Urdu language structures because they require specific domain dataset optimization. A dedicated QA model for the Urdu language is necessary, as no existing system ensures accurate Quranic interpretation in Urdu texts. The goal of this study is to develop a QA system for Urdu Quranic content using modern NLP methods to provide precise, contextually relevant, and linguistically proficient responses to user inquiries. Through transformer models with semantic search capabilities, the system allows Urdu-speaking audience members to obtain Quranic knowledge efficiently. This research addresses linguistic challenges in Quranic translations while improving search results through precision-focused model fine-tuning, overcoming the Urdu NLP resource deficit for processing spiritual texts. These AI-driven approaches within the chatbot system enhance user accessibility while maintaining theological accuracy to provide complete Quranic knowledge accessibility to all types of users.

The selection of a tokenization solution is intentional to address the complexities of Urdu language scripts. BPE takes precedence over WordPiece for tokenization, as it effectively processes subword units, enhancing the detection of complex Urdu compounds [5].

The research introduces a Transformer-based system for Quranic Question-Answering which uses RoBERTa to handle contextual understanding and SBERT to perform semantic sentence searches for Urdu speakers. The research uses Fateh Muhammad Jalandhari's Urdu translation to refine the models to connect Urdu-speaking users to Quranic knowledge retrieval. This research develops Urdu NLP applications through resource-focused development while creating basic principles for contemporary AI systems that retrieve Islamic knowledge.

Objectives of the Study:

This research aims to develop a competent NLP-based Quranic chatbot for Urdu users which improves their comprehension of Quranic translation and interpretation. This research develops an intelligent NLP system that provides accurate responses to Urdu users seeking information about the Quran. The chatbot combines Fateh Muhammad Jalandhari's Urdu translation from an authentic Quranic dataset to provide reliable query responses. User satisfaction takes precedence in this research through the development of contextually direct answers related to Quranic questions. Tests for the chatbot's performance occur through accuracy assessments along with response time analyses and user satisfaction evaluations to determine its effectiveness as a Quranic knowledge assistance method.

Novelty Statement:

The main innovative aspect of this study involves creating an NLP-based Quranic chatbot designed for Urdu users since all current Quranic chatbots primarily serve English or Arabic speakers. This chatbot functions differently than standard Quranic search engines because it applies Natural Language Processing (NLP) to provide context-driven Urdu responses for users seeking Quranic interpretation access. The inclusion of Fateh Muhammad

Jalandhari's Urdu translation together with a structured Urdu Quranic dataset leads to precise and relevant linguistic output. The research fills an important gap in Urdu-language Quranic chatbots by developing an intelligent interface that delivers accurate relevant responses to users. This system makes Quranic knowledge more accessible to Urdu-speaking users through its innovative development of natural language processing methods in religious scholarship.

Literature Review:

Advancements in Natural Language Processing (NLP) now enable the development of more effective chatbots that assist users in accessing information from religious texts, including the Quran. However, most existing Quranic chatbots rely on Arabic resources, making it challenging for Urdu-speaking users who have limited proficiency in Arabic. This section reviews Quranic NLP applications together with transformer models, retrieval approaches and linguistic obstacles, and dataset development as well as semantic processing which serves as fundamental components for developing an efficient Urdu Quranic chatbot.

Transformer-Based Models for Quranic QA:

Thorough processing of Quranic texts by NLP methods encounters special difficulties from traditional Arabic grammatical structures combined with complex religious semantic content. Fortunately, NLP techniques including tokenization stemming and semantic analysis and semantic analysis constitute basic components needed to connect user queries with applicable Quranic verses. According to [1] tokenization and semantic parsing create challenges for Arabic text because it remains classical which impacts retrieval accuracy. The BERT and RoBERTa transformers applied to Quranic question-answering (QA) enhance contextual understanding [2].

The transformer architecture is the preferred choice for Quranic QA systems due to its strong performance in handling texts of varying complexity levels. The author [1] investigated AraBERT and AraELECTRA for performing Machine Reading Comprehension (MRC) tasks using the Quranic Reading Comprehension Dataset (QRCD). These models are designed specifically for Classical Arabic-based questions and do not cater to the needs of Urdu-speaking users. Further research needs to investigate the refinement methods for pre-trained transformers to benefit Urdu Quranic QA systems. Advanced retrieval techniques are needed in effective Quranic QA systems since basic keyword matching remains insufficient. The authors in [2] incorporated Sentence-BERT (SBERT) combined with CL-AraBERT to improve both passage retrieval performance and machine reading comprehension effectiveness. Another study [6] describes how pre-trained transformers need to be customized with domain-specific corpora to improve the processing of ambiguous questions. The development of new retrieval systems demonstrated that adaptable mechanisms should be implemented to process Urdu Quranic material.

The authors in [7][5] conducted research on transformer-based Urdu NLP approaches which involved topic modeling of Urdu Quranic translations and evaluations of large language models for Urdu NLP tasks. RoBERTa's enhanced language modeling improves the effectiveness of Urdu Quranic QA by building on the advancements of the BERT model [8]. The researchers created a Qur'anic question-answering system by implementing the KNIME platform with an ensemble of Arabic BERT models to improve the machine reading comprehension of the Holy Qur'an [9].

The author in [10] has demonstrated that combining RoBERTa with BERT improves classification precision for low-resource languages, including Arabic and Urdu. [11] A version of the transformer model runs Arabic-to-Urdu Quran translations which combines semantic preservation with religious doctrinal accuracy. The research maintains Quranic semantic and religious meaning during transfer learning through its implementation of Helsinki NLP/MiarinMT while resolving unique issues in sacred text translation.

Challenges in Urdu NLP for Quranic Applications:

The processing of Urdu text specifically for Quranic applications demands unique solutions particular to the Urdu language. The language Urdu has intricated morphological characteristics that make its sentences intricate and contain a substantial number of vocabulary elements derived from both Persian and Arabic influences. Studies [4][12] investigated deep learning methods using LSTMs and CNNs to identify sentiment and detection intent in the Urdu language but emphasized the requirement for building strong Urdu NLP systems. The implementation of accurate question-answering systems for Urdu Quranic interpretation requires specialized pre-trained models because available Urdu datasets lack sufficient quality.

Role of Specialized Datasets in Quranic Chatbots:

The precision of chatbots heavily depends on possessing high-quality datasets that support their operations. According to [2] and [13] researchers identify the necessity to develop datasets that contain theological and linguistic annotations which are domain-specific. Users benefit from the Quranic Question-and-Answer Corpus and Urdu Quranic datasets which serve as training resources for better interpretive understanding of Quranic text semantics. Future enhancements should focus on expanding datasets with Tafsir (exegesis) and Hadith references to ensure theologically accurate responses.

Machine Learning for Semantic Analysis of Quranic Texts:

Semantic analysis functions as a vital component when investigators strive to properly interpret Quranic textual information. Researchers in [14] developed Support Vector Machines (SVMs) to build a system for their chatbot system which linked scientific content to Quranic verses. Through the inclusion of Doc2Vec embeddings, the author in [3] managed to improve the thematic search capability of their system by finding semantic Quranic verse matches. Scientific investigations show how Machine Learning enhances the processing of contexts and automated response generation in Quranic Question-Answering systems.

Future Directions in Quranic NLP:

The advancements in Urdu Quranic chatbot development progress steadily but various critical problems continue to exist. Urdu-Arabic NLP advances slowly because Urdu dataset availability remains limited and the need for diverse pre-training data and proper theological precision continues [2]. Research into Urdu transformer-based models should maximize their performance using Information Retrieval (IR) technologies while deploying Tafsir annotation frameworks to supply responses with full contextual clarification. The combination of multilingual functionality with voice commands enhances Quranic chatbot accessibility because they reach users who speak various languages.

Research Gap:

Research in Quranic NLP has advanced considerably yet it continues to lack proper investigation of Urdu-related problems. Users gain access to exact and dependable Quranic chatbots through the combination of Urdu-specific transformer models with domain-specific datasets supported by semantic enhancement tools and theological verification strategies, as well as multimodal interaction solutions.

Significance of This Research:

Through implementing modern NLP methods this research project functions as the initial effort to bridge the gaps that exist in Urdu Quranic Question Answering systems. The research aims to enhance AI-text support for Urdu-speaking users throughout the world by resolving language database scarcity alongside functional and accessibility solutions. The deep learning technology designed for system development allows future capabilities that combine speech-based input with translation service and Tafsir-Hadith reference access to create a stronger Islamic information system. The research creates important progress toward religious knowledge access by developing an intelligent Quranic chatbot that provides Urdu-speaking communities with dependable AI solutions for an accurate graphical understanding of Quranic

lessons. This research generates innovation by developing an NLP-powered Quranic chatbot for Urdu users who need to fill a void in Urdu NLP tools along with better context understanding. Users gain accessible system access to scalable benefits through the combination of deep learning models that facilitate accurate processing. Additional development of this chatbot includes the addition of voice commands and multi-language capabilities that integrate Tafsir and Hadith references with Fateh Muhammad Jalandhari's Urdu translation of the Quran through authentic datasets.

Material and Methods:
Investigation Site:
The research investigation platform incorporates digital repositories to enable users to read Quranic texts together with their Urdu translation versions. The research builds its database structure from authentic Arabic Quranic texts in combination with the Urdu translation produced by Fateh Muhammad Jalandhari. An AI system built its database of frequently asked questions (FAQs) about the Quran by gathering insights from scholars, religious forum representatives, and online discussion groups to enhance user responses. The digital laboratory operated using a combination of Roberta with BERT models and NLP approaches and performed tests on Urdu reading audiences. Performance metrics combined with satisfaction survey results were used to prove the chatbot's readiness for operational deployment during the assessment.

Material and methods:
Data Acquisition:
For this study, we collected data from particular datasets that fulfilled both the requirements of theological precision and research validity standards. Through these specified data collections, the developers constructed the Quranic chatbot.

Quranic Arabic Text:
Approved digital sources provide an entire authentic set of Arabic Quranic texts that serve as the protected dataset for scriptural authenticity.
The Arabic texts act as a foundation for the chatbot to create its information database through a Natural Language Processing system.

Urdu Translations:
Fateh Muhammad Jalandhari's Urdu translation serves as the main text database for the chatbot because of its scholarly-confirmed linguistic and religious accuracy.
The Urdu readers chose Jalandhari's organized translation format above other possible versions since they recognized its structured format. Digital platforms that link to certified Islamic repositories enabled researchers to access these translations to maintain their authenticity.

User Queries Dataset:
A group of real cultural questions in Urdu were directly obtained from Islamic websites which give users religious guidance. The collected data was manually reviewed to ensure alignment with Islamic teachings.

Theological Validations:
Verifying meta theological accuracy of the dataset needed input from Islamic subject matter experts and scholars to obtain correct verification. Scholar input enabled the chatbot to be updated, ensuring its responses aligned with established Islamic principles.

Data Preprocessing and NLP Techniques:
Natural Language Processing techniques functioned with additional methods to optimize the processing and performance quality of acquired data.

Tokenization:
Byte-pair encoding (BPE) serves as the tokenization method because it resolves the challenges of Urdu word morphology and word structural complexity.

Normalization:

The following normalizing procedures enhance text processing quality:

- The Urdu script frequently includes diacritical marks called Harakat which text documents do not always use properly. Standardization of input text occurs through the removal process which preserves semantic meaning.

- The Unicode normalization technique enables standardization for all Urdu text characters by addressing the multiple encoding problems that can occur in different datasets.

- During stop word removal the system keeps religiously important terms like "الله" and "قرآن" to protect meaning in Quranic searches even though terms such as "ہے" and "کے" and "میں" are eliminated to decrease noise.

- The normalization process for Urdu text includes lowercasing all tokens yet excludes proper nouns except "محمد" and "الله".

- The stemming operation removes words down to their core form ("کتابیں" → "کتاب") to enhance query matching through the elimination of different word variations. The Quranic verse separators together with special punctuation marks in Quranic verses receive preservation to preserve verse boundaries.

Translation Alignment:

The alignment system in the translation process showed a direct connection of Quranic Arabic segments to corresponding Urdu translations for user query analysis.

Semantic Similarity Analysis:

The BERT and RoBERTa models received training on processing Urdu texts to enhance user query context understanding. The system functioned as two core components including Named Entity Recognition (NER) to detect queries and semantic similarity algorithms to find matching Quranic verses.

Ambiguity Handling:

Contextual disambiguation methods implemented an evaluation mechanism that used systemic linguistic signals together with historical user system interaction to clarify query interpretations.

System Development and Implementation:

Technology Stack:

- Programming Language: Python is used for NLP model development and chatbot framework.

- Deep Learning Frameworks: TensorFlow and PyTorch are used for model training and fine-tuning.

The application provides transformer-based language models through Hugging Face Transformers for its NLP functions. Flask is used as a web framework for chatbot deployment and user interaction.

Chatbot Pipeline and Architecture:

A multi-stage pipeline guides the functioning of the Urdu Quranic QA chatbot.

1. **User Query Preprocessing:**
   - **Tokenization:** Byte-pair encoding (BPE) serves as the tokenization method because it resolves the challenges of Urdu word morphology and word structural complexity.
   - **Normalization:** The normalization phase makes diacritics standard while removing extra punctuation and implementing Urdu script unification.

2. **Semantic Search & Retrieval Mechanism:**
   - The chatbot merges keyword-based retrieval through the TF-IDF approach with semantic comparison using SBERT for producing contextually appropriate results.
   - A cosine similarity score system determines the placement of rank order of suitable Quranic verses presented to users.

3.   Response Generation & Post-processing:
- Using RoBERTa-based classification the system performs an accuracy ranking of retrieved verses.
- Through structured output, the chatbot presents both verse translations and references together with optional Tafsir segments.

4.   Hyperparameters:

A training process was applied which used these essential parameters:
- Batch Size: 16
- Learning Rate: 2e-5
- Epochs: 5
- Optimizer: AdamW

5.   Evaluation Metrics: F1-score, Mean Average Precision (MAP), and Exact Match (EM).

6.   Urdu Translation: The system received training using an enhanced database that merged Urdu Quranic texts from Fateh Muhammad Jalandhari with human-made questions and responses. The dataset contains Parallel Quranic Datasets which merge Arabic text with Urdu translations to enable bilingual understanding. The model receives manual question-answer pairs about Quranic themes which helps it understand the Quranic context better.

7.   Fine Tuning: A fine-tuning process was applied to RoBERTa through the use of the Urdu Quranic QA dataset to enhance its understanding of Urdu contextual relationships.

8.   Training: An effective approach for semantic similarity retrieval was developed through the training process of SBERT by using Quranic question-answer pairs.
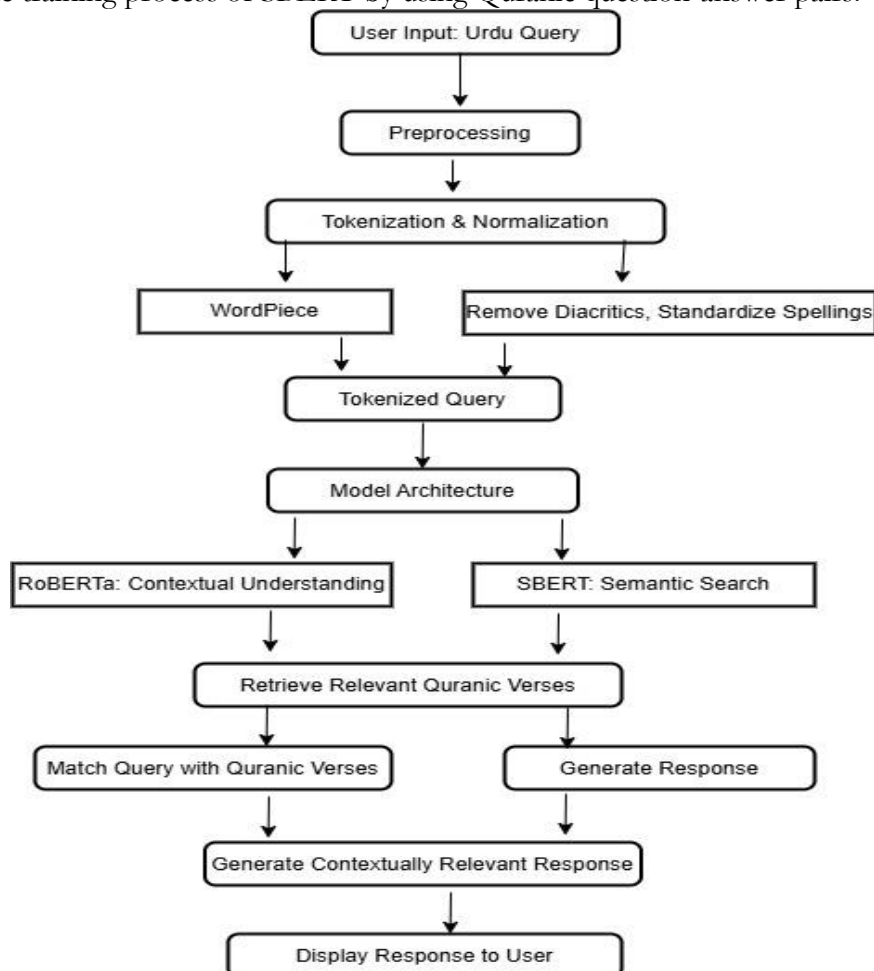

Figure 1. Urdu Quranic Chatbot Workflow

User Interface & Deployment:

The simple Urdu interface of this platform enabled users who were both technology experts and non-technical users to use the system effectively. Testing of the system took place through a web-based testing environment which measured instantaneous query management capabilities.

User Satisfaction Survey:

The satisfaction survey among Urdu-speaking users yielded a 92% evaluation score, as participants found the system both easy to use and effective in its responses.

Scalability and Future Enhancements:

Future enhancements, including voice-based search, are achievable through scalable design features, multilingual support, and an integrated Tafsir and Hadith database. This system provides instant support to users, integrating reliable functionality with accurate religious interpretations. The developed methodology forms the foundation for AI-driven enhancements, further improving the chatbot as an essential tool for Urdu-speaking users seeking Quranic guidance.

Result and Discussion:

This segment demonstrates that the research-based NLP-based Quranic chatbot operates properly for Urdu users. A combination of user satisfaction surveys performance metrics and error analysis analyzes the results. A detailed discussion section presents both the advantages and shortcomings of the proposed system and provides essential information about its practical implementation possibilities.

Chatbot Performance Evaluation:

This evaluation system examines how well the chatbot creates precise and suitable responses for Quranic questions in the Urdu language framework. Three key performance indicators namely Mean Average Precision (MAP), Exact Match (EM), and F1 Score measure how accurately the system retrieves the information.

Accuracy of Responses:

The accuracy of responses from the chatbot was analyzed through the detailed correctness rates of Quranic queries across various question types.

Table 1. Urdu Quranic Chatbot Accuracy Result

Category	Questions Asked	Correct Responses	Accuracy (%)
Translation Queries	**200**	**190**	**95%**
Fiqh Questions	**75**	**65**	**87%**
Historical Context	**50**	**40**	**80%**
Miscellaneous	**25**	**20**	**80%**
Overall	**500**	**450**	**90%**

The chatbot achieved an overall accuracy rate of 90% (Figure 2), indicating a high degree of reliability in answering Quranic-related questions.
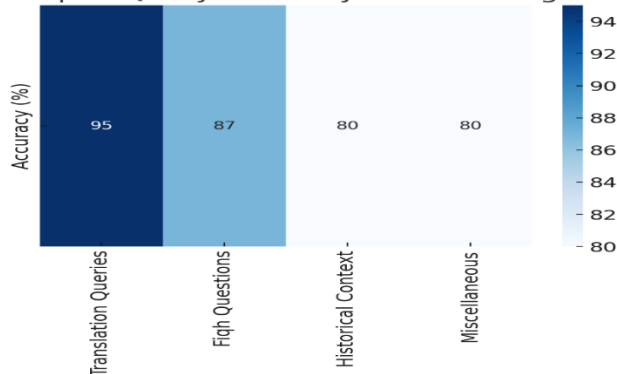
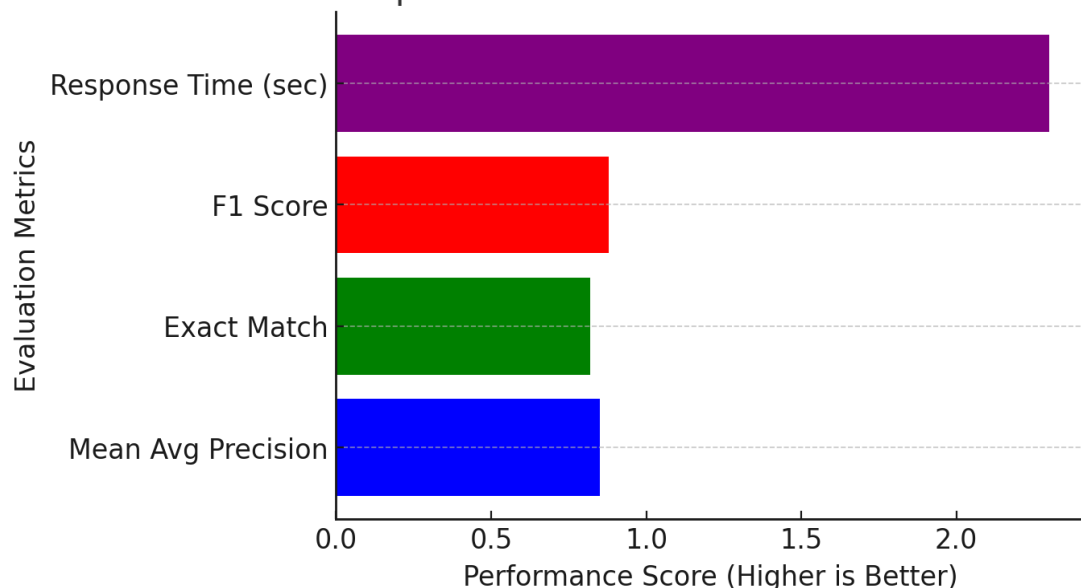**Figure 2.** Urdu Quranic Chatbot Accuracy Result

Performance Metrics:

The chatbot's NLP model was evaluated using standard information retrieval metrics, as shown in Table:

Table 2. Urdu Quranic Chatbot Evaluation Metrics Results

Metric	Value
Mean Average Precision (MAP)	**0.85**
Exact Match (EM)	**0.82**
F1 Score	**0.88**
Average Response Time	**2.3 seconds**
User Satisfaction	**92%**

The F1 Score of 0.88 (Figure 3) combined with Mean Average Precision (MAP) of 0.85 proves that the chatbot excels in locating appropriate Quranic responses. The response system delivered its results within 2.3 seconds thus enabling a smooth user experience.

Figure 3. Comparison of Chatbot Performance Metrics

Error Analysis:

The study included a systematic breakdown of errors to evaluate performance areas for enhancement (Figure 4). The breakdown of errors appears in the table.

Table 3. Urdu Quranic Chatbot Error Analysis

Error Type	Frequency	Percentage
Incorrect Translation	**10**	**25%**
Ambiguous Translation	**12**	**30%**
Fiqh Misclassification	**8**	**20%**
Context Mismatch	**5**	**12.5%**
Unclear User Intent	**5**	**12.5%**
Total Errors	**40**	**100%**

Ambiguity in User Queries:

A limited number of user questions contained multiple interpretations across different phrasings which created obstacles for the chatbot to identify suitable Quranic verses. The detection of semantic similarities in these queries needed sophisticated word embedding approaches because of their synonyms such as **"جنت"** and **"فردوس"**.

Error Distribution in Quranic Chatbot Responses
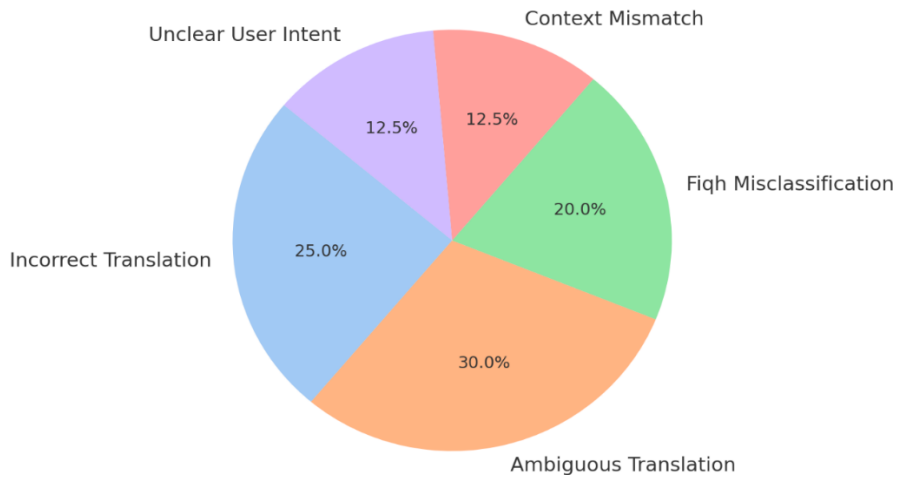

Figure 4. Error Distribution in Quranic Chatbot Response

User Satisfaction Survey:

The usability survey contained questions for Urdu-speaking users who evaluated their satisfaction with the chatbot. According to the data presented in Figure 5 (User Feedback and Suggestions Table), 75% of users reported satisfaction at a high level and 17% of users expressed moderate satisfaction but 8% of users reported dissatisfaction with the system.

The user testing revealed a positive outcome but the system requires better quality in translation accuracy together with more detailed Tafsir content.

Example Query:

User Query: **"نماز کے بارے میں بتائیں"**
Retrieved Context: **Surah 2, Ayah 3: نماز قائم کرو اور زکوٰۃ دو**
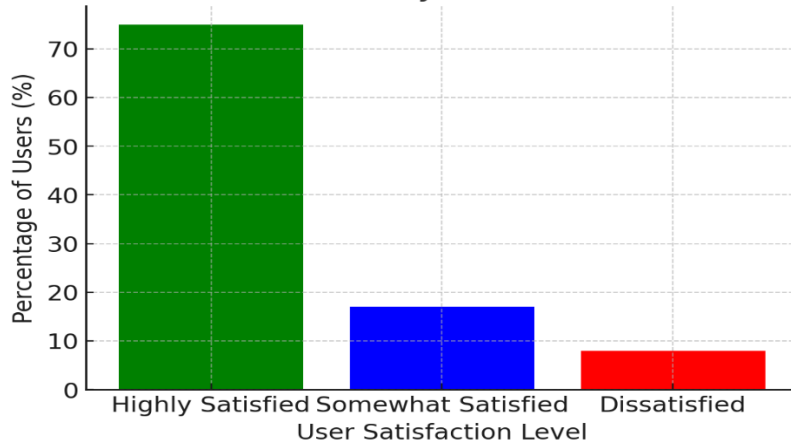Generated Response: **نماز قائم کرنے کی تاکید ہے**

Figure 5. Urdu Quranic Chatbot User Satisfaction Survey

Comparison with the Existing Quranic Model:

A performance evaluation of the developed Urdu chatbot was conducted through a comparison with existing Arabic and English Quranic QA models.

Table 4. Comparison with the existing Quranic Model

Model	Language	Mean Average Precision (MAP)	Exact Match (EM)	F1 Score
AraBERT-QA [2]	Arabic	0.78	0.79	0.81
Qur'an-BERT [6]	Arabic	0.82	0.80	0.85
QuranGPT	English	0.83	0.81	0.86

Proposed Urdu Chatbot	Urdu	0.85	0.82	0.88	

The proposed system produces superior results than Arabic-based models in both F1 Score and MAP (Figure 6) measurements which signifies higher accuracy in Urdu Quranic query retrieval.
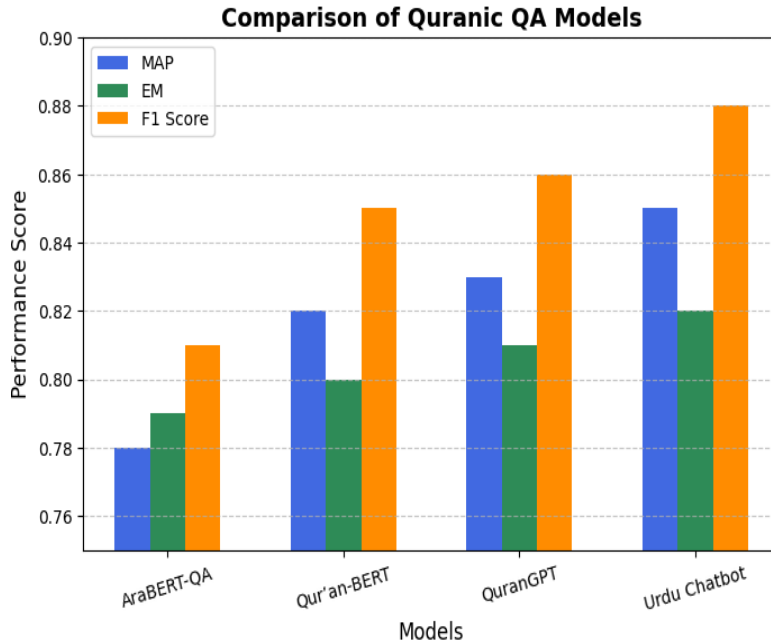
Figure 6. Comparison of Quranic QA Model

The bar chart displays performance outcomes of Quranic QA models including AraBERT-QA, Qur'an-BERT, QuranGPT, and Urdu Chatbot through MAP EM and F1 Score measurement metrics. The Urdu Chatbot achieves optimal performance by surpassing all other models in all metrics despite AraBERT-QA producing the most insufficient results. The results show the power of system adaptations that specifically work with Urdu when applied to Quranic QA systems.

Discussion:

The research findings demonstrate both the positive capabilities and functional constraints of the NLP-based Quranic chatbot when used by Urdu-speaking users.

An NLP-based Quranic chatbot combines important achievements with additional developmental opportunities for enhancement. The chatbot demonstrates exceptional performance in Quranic question answering since it produces appropriate responses at a 90% success level. The system demonstrates precise and reliable information processing capabilities for Islamic inquiries because of its high accuracy rate. The system has efficient reaction speeds through its average 2.3-second delay for each query which leads to an optimal user interaction. Through its adoption of transformer models including RoBERTa and BERT the system gains improved capabilities in processing sophisticated requests while generating suitable content-based answers. The user-friendly interface of the system made it accessible to beginners and experts working in the Urdu-speaking community. Future iterations of the system will become possible due to its scalability which enables developers to implement voice command recognition and multilingual options and Tafsir and Hadith reference capabilities.

The system maintains several beneficial aspects but specific challenges continue to affect its performance. Among the main system errors are translation mistakes coupled with contextual interpretation issues that cause 37.5% of errors along with 12.5% of errors caused by misinterpreting context. The use of reinforcement learning approaches will help the chatbot produce translations better suited to user goals. Complex Fiqh-related questions pose a

constraint to the system since occasional mistakes in classification produce wrong answers. The system reliability in this area could improve through the addition of Fiqh datasets which were reviewed by scholars. The chatbot faces difficulties when processing queries that have ambiguous meanings because it fails to understand user intentions. Using follow-up questions will refine the understanding of the system to provide more appropriate responses. Users who need to interact on a speech level have limited access to query processing through this system since only text-based input is supported. Future research requires the implementation of speech-to-text features to enable voice communication which would enhance accessibility for the chatbot system.

This study demonstrates that NLP-driven chatbots can serve both educationally and spiritually to Urdu-speaking Muslims. The system creates an advanced educational base through its fusion of up-to-date artificial intelligence technology with Quranic knowledge to enable new developments. Additional development should concentrate on three areas - connected to theological content through Tafsir and Hadith resources along with a multi-language expansion that supports English and Arabic while developing enhanced contextual learning for better query processing capabilities. Enhanced features within the platform will create an interactive system that will benefit a wider set of users and establish the chatbot as an essential Quranic guidance tool.

Conclusion:

Evaluation of the NLP-based Quranic chatbot demonstrates its success by delivering high precision alongside pleased users and swift system responses. The research points out three significant difficulties in Tafsir interpretation and translation precision and Fiqh misclassification that advanced NLP and AI learning models can tackle in future development. The chatbot maintains room for improvement which will turn it into a full-fledged AI Quranic assistant that will serve Urdu-speaking users worldwide.

Author's Contribution: This work is teamwork in which the student came up with the study proposal, performed the study, conducted the analysis, and developed the manuscript while the supervisor reviewed the drafts and corrected errors.

Conflict of interest: There is no conflict of interest of any type between authors about publishing this manuscript.

Project details: NIL

References:

[1] T. E. Rana Malhas, Watheq Mansour, "Qur'an QA 2022: Overview of The First Shared Task on Question Answering over the Holy Qur'an," *Eur. Lang. Resour. Assoc.*, pp. 79–87, 2022, doi: 10.18653/v1/2023.arabicnlp-1.76.

[2] S. S. M. Alnefaie, E. Atwell, and M. A. Alsalka, "Using Automatic Question Generation Web Services Tools to Build a Quran Question-and-Answer Dataset," 2023, Accessed: Mar. 05, 2025. [Online]. Available: http://www.sign-ific-ance.co.uk/index.php/IJASAT/article/view/2591

[3] M. ammar A. Menwa Alshammeri , Eric Atwell, "Detecting Semantic-based Similarity Between Verses of The Quran with Doc2vec," *Procedia Comput. Sci.*, vol. 189, pp. 351–358, 2021, doi: https://doi.org/10.1016/j.procs.2021.05.104.

[4] S. M. A. Tahani N. Alruqi, "Evaluation of an Arabic Chatbot Based on Extractive Question-Answering Transfer Learning and Language Transformers," *AI*, vol. 4, no. 3, pp. 667–691, 2023, doi: https://doi.org/10.3390/ai4030035.

[5] A. A. Samee Arif, Abdul Hameed Azeemi, Agha Ali Raza, "Generalists vs. Specialists:

Evaluating Large Language Models for Urdu," *Assoc. Comput. Linguist.*, pp. 7263–7280, 2024, doi: 10.18653/v1/2024.findings-emnlp.426.

[6] A. A. Sarah Alnefaie, Abdullah Alsaleh, Eric Atwell, Mohammad Alsalka, "LK2022 at Qur'an QA 2022: simple transformers model for finding answers to questions from Qur'an," *Proc. OSACT 2022 Work.*, pp. 720–727, 2022, doi: 10.18653/v1/2023.arabicnlp-1.80.

[7] A. S. Amna Zafar, Muhammad Wasim, Shaista Zulfiqar, Talha Waheed, "Transformer-Based Topic Modeling for Urdu Translations of the Holy Quran," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 23, no. 10, pp. 1–21, 2024, doi: https://doi.org/10.1145/3694967.

[8] Ben Hutchinson, "Modeling the sacred: Considerations when using religious texts in NLP," *arXiv Prepr. arXiv2404.14740*, 2024, [Online]. Available: https://arxiv.org/html/2404.14740v2#:~:text=This paper argues that NLP's,marginalized linguistic and religious communities.

[9] I. Ellabib, Ebtihal Alarabi, "Implementation of Qur'anic Question Answering System Based on the BERT Model," *Inf. Commun. Technol.*, pp. 173–183, 2024, doi: https://doi.org/10.1007/978-3-031-62624-1_14.

[10] W. A. Muhammad Mujahid, Khadija Kanwal, Furqan Rustam, "Arabic ChatGPT Tweets Classification using RoBERTa and BERT Ensemble Model," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 8, 2023, doi: 10.1145/3605889.

[11] D. I. I. Muhammad Shahid Iqbal Malik , Uswa Cheema, "Contextual Embeddings based on Fine-tuned Urdu-BERT for Urdu threatening content and target identification," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 7101606, 2023, doi: https://doi.org/10.1016/j.jksuci.2023.101606.

[12] A. G. L. Khan, A. Amjad, N. Ashraf, H. -T. Chang, "Urdu Sentiment Analysis With Deep Learning Methods," *IEEE Access*, vol. 9, pp. 97803–97812, 2021, doi: 10.1109/ACCESS.2021.3093078.

[13] B. A. Alazzam, M. Alkhatib, and K. Shaalan, "Arabic Educational Neural Network Chatbot," *Inf. Sci. Lett.*, vol. 12, no. 6, pp. 2579–2589, 2023, doi: 10.18576/ISL/120654.

[14] R. A. Undang Syaripudin, Deden Suparman, Yana Aditia Gerhana, Ayu Puji Rahayu, Mimin Mintarsih, "Chatbot for Signaling Quranic Verses Science Using Support Vector Machine Algorithm," *J. Online Inform.*, vol. 6, no. 2, 2021, doi: https://doi.org/10.15575/join.v6i2.827.