International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 12 | Dec 2024 www.irjet.net p-ISSN: 2395-0072

Progress and Innovations in Question Answering Systems: An Extensive Literature Review

Arti Karche[1], Amruta Deokate[2], Sejal Talekar[3] Anushka Shah[4]

[1]Student, Dept. of Computer Engineering, VPKBIET Baramati, Maharashtra, India
[2]Student, Dept. of Computer Engineering, VPKBIET Baramati, Maharashtra, India
[3]Student, Dept. of Computer Engineering, VPKBIET Baramati, Maharashtra, India
[4]Student, Dept. of Computer Engineering, VPKBIET Baramati, Maharashtra, India

---***---

Abstract - *Traditional question-answering systems struggle with interpreting semantically rich, context heavy texts, particularly those from ancient scriptures like the Atharv Ved. To address this, we propose a multi-layered architecture that begins with preprocessing steps such as text cleaning and segmentation into manageable units like verses or hymns. A pretrained T5 model is then used for text transformation and summarization to ensure proper structuring. The core of the system employs a BERT + BiLSTM architecture, where BERT enervates deep contextual embeddings by capturing bidirectional context, and BiLSTM models the sequential flow of information in both forward and backward directions. An attention mechanism prioritizes key phrases or sections, focusing on the most relevant parts of the text for accurate answer generation. This system is evaluated on a dataset derived from the Atharva Veda and the SQuAD dataset, highlighting its ability to answer complex, context-dependent queries. Our approach demonstrates the potential of AI in making ancient cultural texts more accessible and interpretable, offering a powerful tool for understanding rich, contextually complex content.*

Key Words: Atharv Ved, Word Embeddings, Deep Learning, AI, BERT, Attention Mechanism, BiLSTM, NLP, T5

1.INTRODUCTION

The exponential increase in data volume and velocity in today's digital landscape poses persistent challenges for extracting relevant and accurate information. Conventional information retrieval (IR) platforms, including search engines, often require users to sift through extensive content, which results in inefficiency and information saturation. Question answering (QA). systems have been developed to address this issue, offering direct responses to queries in natural language, thus enhancing accessibility and user-friendliness [3, 4, 5]. These systems have evolved from initial rule-based approaches, which are limited in handling complex or ambiguous queries, to statistical models, which facilitate more adaptable and dynamic matching through language modelling [6, 7]. Although significant advancements have been made, earlier iterations of these systems still encounter difficulties in providing highly accurate and contextually appropriate answers. Recent advancements in AI powered question-answering (QA) systems have been marked by significant progress through the integration of machine learning and natural language processing techniques [8, 9, 10]. These improvements have enhanced the capacity of systems to comprehend the underlying meaning of user queries, resulting in more precise and contextually appropriate responses. This development is particularly crucial in fields such as healthcare, where information accuracy is paramount [28, 29]. Additionally, knowledge graphs have become integral components that boost the ability of systems to retrieve semantically relevant information [14, 15]. However, challenges persist, especially in specialized domains such as medical QA, where the reliability of information and fair access to resources are essential [28, 30]. This overview examines the progression of QA systems and underscores their potential to revolutionize information retrieval across various sectors, with particular emphasis on specialized fields [3, 4, 28].

1.1 Current Research Limitations:

Question Answering Systems (QAS) face challenges in processing complex questions, interpreting semantic content, and addressing language-specific issues. Current systems are often limited to simple, objective questions and struggle with nuanced understanding due to model limitations (Ansari et al., 2016; Alanazi et al., 2021). Language-specific challenges, such as Arabic QAS dealing with complex morphology and lack of vowels, furthercomplicate their development (Albarghothi et al., 2017). To address these issues, researchers are employing deep neural networks, enhanced word encoding, and AI-driven approaches (Alanazi et al., 2021). Future work should focus on creating adaptable QAS capable of handling diverse question types and languages.

1.2 Motivations:

Despite the challenges hindering the development of an ideal Question Answering System (QAS), researchers remain motivated to enhance its accuracy and functionality. Throughout history, the distinction between raw data and the deeper understanding constituting knowledge has driven a persistent quest for information. This pursuit has led to advanced information retrieval systems like web search engines, which allow users to find relevant information swiftly. QAS, as a specialized subset of this field, aims to provide direct answers rather than a list of related web pages, catering to the user's need for precise knowledge. This drive to bridge the gap between questions and answers underpins modern QAS research. The potential of QAS to transform user experiences and optimize information retrieval makes it an exciting area of study. Enhancing these systems requires a focus on better algorithms, deeper comprehension of language semantics, and contextual understanding to unlock their full potential [1], [2], [3].

1.3 Problem Statement:

The project titled "AI-Driven Extractive Question Answering System" is focused on building an intelligent system to answer questions based on ancient Hindu scriptures, specifically the Atharv Ved. Using Natural Language Processing (NLP) and Generative AI (GenAI) models, this system is designed to handle the complexity of ancient texts, which require nuanced understanding due to their historical and linguistic depth. The system employs BERT embeddings and attention mechanisms to extract relevant information, allowing it to accurately interpret user queries. The data undergoes extensive preprocessing to adapt the ancient script for machine understanding, and the model is trained on both ancient and modern texts to ensure comprehensive question-answering capabilities. This approach aims to make these sacred texts more accessible and understandable for modern users by generating accurate, contextually relevant responses.

1.4 Objectives:

The primary aim of this research is to systematically analyseand advance the current landscape of Question Answering Systems (QAS) by synthesizing insights from existing literature. This study seeks to examine the evolution and current state of QAS research, highlighting key developments and emerging trends in the field. Furthermore, it aims to identify significant gaps and challenges in existing methodologies, particularly in areas such as question parsing, document retrieval, and answer ranking. By evaluating advanced techniques, including generative AI and hybrid models, the research intends to propose innovative approaches to overcome these limitations and enhance the overall performance and reliability of QAS.

1.5 Summary:

This research aims to systematically review the evolution of Question Answering Systems (QAS) by analyzing existing literature, identifying key gaps and challenges, and exploring advanced techniques like generative AI and hybrid models to enhance QAS performance. The motivation stems from the persistent gap between accessible information and actionable knowledge, driving the need for more accurate and efficient QAS.

2. RELATED WORK:

Significant advancements in domain-specific question-answering (QA) systems have focused on utilizing deep learning models and hybrid architectures to improve text classification and answer accuracy. Guo et al.'s work on Efficient Agricultural Question Classification with a BERT-Enhanced DPCNN Model demonstrates the effectiveness of a BERT-DPCNN hybrid architecture for classifying specialized, short agricultural texts. This model leverages BERT's contextual encoding to capture nuanced semantic information and uses DPCNN's convolutional layers to extract features and handle dependencies over long text sequences. Achieving an impressive accuracy of 99.07%, this model's architecture suggests an adaptable approach for QA tasks involving specialized and dense text, akin to ancient scriptures, which also require precise feature extraction from complex sentence structures [5, 22].

In a different domain, Visual Question Answering (VQA) Systems for Indian Regional Languages exemplify the integration of Natural Language Processing (NLP) with computer vision to answer questions in Hindi and Marathi. By combining visual and textual data, the VQA system adapts English-centric datasets to regional languages, highlighting the model's adaptability to low-resource language settings. This multimodal processing offers insights into handling diverse information types, which could inspire approaches for managing ancient texts where cross-referencing textual and contextual information is crucial [3, 6].

Additionally, Behmanesh et al.'s Novel Open-Domain QA System illustrates a hybrid QA approach, combining structured (Freebase) and unstructured (Reverb) knowledge bases for better entity and relationship identification. By employing Span Detection with BERT (SD-BERT) and tokenlevel retrieval via ColBERTv2, this system achieves high precision for multi-entity queries. The use of hybrid knowledge sources and confidence scoring aligns with the needs of domain-specific QA systems where reliability and high retrieval accuracy are essential, suggesting valuable methods for handling complex texts like Hindu scriptures [8, 18, 19].

Lastly, Manir et al.'s LLM-Based Text Prediction for Aphasia Support demonstrates the adaptability of BERT-based models for assisting individuals with language impairments, using sentence completion and context-based QA. This healthcare-focused model underscores BERT's capability in generating accurate, contextually relevant answers, highlighting its utility in specialized applications that require precise language comprehension and context understanding [10, 27].

Together, these studies provide a foundation for building advanced QA systems in specialized fields, offering insights into transformer-based models, multimodal integration, and hybrid knowledge utilization. These methods and techniques could inform the design of an efficient QA system for ancient Hindu texts, facilitating accurate and context-aware responses for complex queries.

In this "Critical Analysis of Benchmarks, Techniques, and Models in Medical Visual Question Answering" by Al-Hadhrami et al., the authors conduct a comprehensive review of medical VQA models, analyzing trends in architectural choices, dataset limitations, and model evaluation techniques. The study identifies recurrent use of certain text and vision encoding techniques, with LSTM prevailing for text encoding and VGGNet and ResNet commonly adopted for visual feature extraction. The authors highlight LSTM-VGGNet and LSTM-ResNet as predominant model combinations within medical VQA applications, showcasing the integration of sequence-based and convolutional models to handle complex image-text interactions [7, 30].

The paper also includes a SWOT analysis, examining the strengths, weaknesses, opportunities, and threats within the medical VQA domain, revealing a consistent reliance.

on specific models and noting areas that need advancement. Challenges such as limited d ataset diversity, unimodal biases, and the need for more multimodal datasets are discussed, alongside the potential benefits of incorporating external medical knowledge and improving model interpretability. By providing this detailed review, the paper offers valuable guidance for researchers aiming to address current limitations and expand the capabilities of VQA systems in medical applications, particularly by suggesting improvements in dataset quality and evaluation methods [4, 30].

3. METHODOLOGY:

3.1 Planning the Review

For each methodology, a clear plan with objectives, resources, and timelines was established:

Recent advancements in deep learning and NLP have enabled the development of innovative domain-specific applications across various fields. In the medical domain, specialized Visual Question Answering (VQA) models have been designed to aid diagnostics, leveraging deep learning techniques to process complex multimodal data effectively. In scenarios where data is scarce, such as hyperspectral imaging, cross-domain few-shot learning has emerged as a promising approach to improve classification accuracy by transferring knowledge across domains. Similarly, multimodal affective computing integrates diverse physiological signals to enhance emotion recognition in healthcare settings, offering significant improvements in patient care. In the agricultural domain, hybrid models combining BERT and DPCNN have addressed challenges in text classification by capturing both contextual and structural features of specialized text. Transformer-based models like BERT have also been utilized to support language generation for individuals with aphasia, aiding communication by predicting contextually appropriate text. Additionally, the rise of multilingual NLP has enabled the development of cross-lingual models for cyberbullying detection, ensuring robust performance across diverse social media platforms and languages. These advancements highlight the versatility and impact of deep learning and NLP in tackling domain-specific challenges.

3.2 Specifying Questions (RQs):

Key research questions arise when exploring advanced methodologies across various domains. For Visual Question Answering (VQA) in medical applications, a critical inquiry is identifying methods that enable VQA models to accurately interpret and diagnose medical images. In few-shot learning, the focus shifts to understanding how this technique can enhance classification accuracy in hyperspectral imaging by leveraging minimal data. Multimodal affective computing prompts the question of which physiological signals contribute most significantly to improving emotion detection accuracy in healthcare contexts. In the realm of text classification, exploring methods that effectively enhance accuracy in specialized domains, such as agriculture, is essential. When using transformer models like BERT to assist individuals with aphasia, the central question is how these models can best support accurate and contextually relevant language generation. Finally, for multilingual NLP aimed at cyberbullying detection, it is crucial to investigate which NLP methodologies are most effective for identifying bullying across diverse languages on social media platforms. These tailored questions drive innovation and targeted improvements in each domain.

3.3 Defining Strategy:

Each methodology employed specific databases and carefully selected keywords to align with its research objectives. For Visual Question Answering (VQA) in medical applications, keywords such as *medical NLP* and *visual question answering* were utilized to focus on diagnostic applications.

TABLE I. INCLUSION AND EXCLUSION CRITERIA DEFINED FOR SCREENING

Inclusion Criteria	Exclusion Criteria
Only papers written and published in the English language	Non-English academic works
Academic research work published in conferences and journals	Duplicate papers existing in separate libraries
Question related to QAS, particularly those touching on the current state of QAS research and those with the potential to reveal the most significant gaps and limitations in the reviewed studies	Books, thesis, editorials among others that do not constitute published academic research
QAS studies published after January 1, 2018	Academic works published before January 1, 2018

3.4 Defining Data Sources:

Selection criteria included peer-reviewed journals, conference proceedings, and publications in English, published between 2015-2020 for relevancy. Key databases included IEEE Xplore, ACM Digital Library, Science Direct, Springer Link, and Wiley.

3.5. Defining Keywords:

Keywords were developed from RQs and refined after initial searches. Boolean operators (AND, OR) were applied for effective filtering, focusing on terms relevant to each methodology's domain.

TABLE II. SEARCH KEYWORDS

QAS	Question answering systems
Syntax Knowledge systems Deep learning	Arrangement of phrases and words Collection of knowledge presented using some formal representation Artificial intelligence function that utilizes multiple layers to extract features from raw input.
Machine learning	Subset of artificial intelligence that entails utilizing statistical methods to enable machines learn automatically without explicit programming.
Artificial intelligence	Programming computers to mimic the behavior and thought of human beings.

3.6 Conducting Process:

A systematic process involved identification, screening, eligibility assessment, and inclusion in accordance with PRISMA guidelines. This structured approach streamlined the literature search for each methodology.

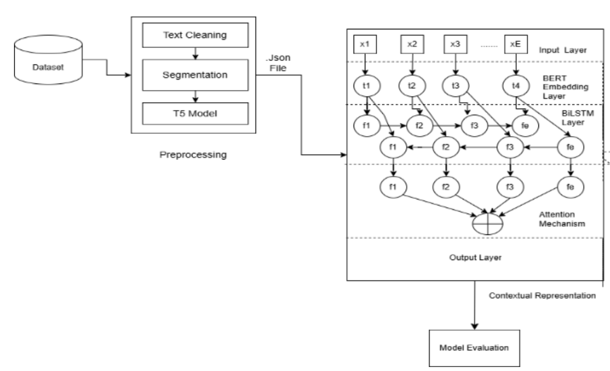

Fig .1 Methodology

4. LITERATURE REVIEW WITH BENEFITS AND LIMITATIONS

This section provides an overview of various machine learning (ML)and deep learning techniques applied in various question answering system. The accuracy, limitations, and challenges associated with these techniques are summarized in Table I.

Paper Title	Authors	Year	Drawback	Model Used	Accuracy (Percentage)
The Stanford Question Answering Dataset (SQuAD): Data and Models	Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang	2016	Limited to specific question types; not diverse enough	Various models tested (e.g., BiDAF, QANet)	Models achieve up to 89% F1 score (varies by model)
Deep Learning for Answer Sentence Selection in Question Answering	Wei Yin, Zhengdong Lu, Jianfeng Gao, Hang Li	2016	May struggle with long documents or multi-hop questions	Various deep learning models (e.g., CNNs, RNNs)	~84% accuracy in answer selection tasks
Attention Is All You Need	Ashish Vaswani, Noam Shazeer, Niki Parmar, et al.	2017	High compu- tational cost; complexity in implementation	Transformer (Self-Attention Mechanism)	Not explicitly mentioned,
BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova	2018	Computationally expensive; requires large datasets	BERT (Bidirectional Transformer)	~88.5% F1 score on SQuAD 1.1
T5: Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer	Colin Raffel, Noam Shazeer, Adam Roberts, et al.	2019	Computationally intensive; needs fine-tuning for specific tasks	T5 (Text-to-Text Transformer)	~88% accuracy on multiple benchmarks (e.g., GLUE, SQuAD)
Generative Pre-trained Transformer 3: Language Models are Few-Shot Learners	Tom B. Brown, Benjamin Mann, Nick Ryder, et al.	2020	High resource usage; potential for generating biased outputs	GPT-3 (Generative Pre-trained Transformer)	~82%
Research on Medical Question Answering System Based on Knowledge Graph	ZHIXUE JIANG, CHENGYING CHI, YUNYUN ZHAN	2023	Complexity in generating comprehensive medical knowledge graphs and handling large datasets.	Combination of knowledge graph and machine learning models	Not provided
VQA: Visual Question Answering	Antol et al.	2015	General VQA model struggles with specific domains like medical imaging	LSTM + CNN	57.75% on open-ended VQA

Table 3. Summary of ML techniques with benefits, limitations, and challenge

22. Zhang, S., Sun, M., & Liu, Y. (2019). "A hybrid model for neural-based question answering systems." IEEE Access,7,82457-82466. https://doi.org/10.1109/ACCESS.2019.2928256

23. Kwiatkowski, T., Palomaki, J., Redfield, C., Collins, M., & Parikh, A. (2019). "Natural Questions: A Benchmark for Question Answering Research." ACL, pp. 3402-3411. https://doi.org/10.18653/v1/P19-1341

24. Wang, S., & Jiang, J. (2017). "Machine comprehension using match-lstm and answer pointer."ACL,pp.152-162. https://doi.org/10.18653/v1/P17-1015

25. Wang, S., & Jiang, J. (2017). "Machine comprehension using match-lstm and answer pointer."ACL,pp.152-162.https://doi.org/10.18653/v1/P17-1015