Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

A Comprehensive Survey of Question Answering Systems: From Traditional Approaches to Large Language Models

Amit Virmani*, Alok Kumar, Vineeta Singh
Department of Computer Science and Engineering, School of Engineering and Technology (UIET), Chhatrapati Shahu Ji Maharaj University, Kalyanpur, Kanpur 208024, U.P. India

Abstract: Question Answering Systems (QAS) aim to answer users' natural language questions automatically, correctly, and meaningfully. There are still problems in question understanding and information retrieval, even with the advancements in Natural Language Processing (NLP). Problems connected with complicated, multi-hop reasoning questions and dirty data continue to be challenging. The challenge is to create an effective QAS that interprets diverse question forms, extracts salient information, and provides accurate and contextually appropriate answers efficiently. These problems entail the combination of knowledge representation, deep language comprehension, and retrieval methodologies for greater user satisfaction and reduced user effort in search. This survey offers a comprehensive survey of multiple methods for QAS, focusing on Deep Learning (DL)-based, Large Language Models (LLM)-based, and knowledge graph-based methods. A total of 25 research papers related to QAS were accumulated and analyzed. The survey also provides the general outline of QAS. Subsequently, the gathered research papers are classified on the basis of the methods exploited, and an elaborated literature survey is presented for each category. Following this, the cons encountered by the baseline models are discussed. Lastly, the analysis of the methods is conducted according to tools, publication years, datasets, metrics, and categories.

Keywords: Question Answering Systems, Machine Learning, Deep Learning, Knowledge Graph, Natural Language Processing

1. INTRODUCTION

In recent years, due to the quick development of big data, the amount of complex network information has grown rapidly. Thus, individuals mostly depend on search engines for effectively retrieving the most relevant information. Nevertheless, conventional search engine techniques failed to satisfy the requirements of users because of the large amount of data. Hence, the requirement for more effective information retrieval approaches is increasing globally. Moreover, QAS [1] is regarded as the most suitable technique for users' habits. This process is simple to use since only the natural language questions are fed as input for providing relevant answer information, which also effectually minimizes the cost and time for obtaining information [2]. In the initial stages, pattern matching was exploited by researchers for performing QA retrieval, and further, question answering is considered a database query task. Furthermore, QAS mostly concentrates on syntax and semantic analysis of user input, which is carried out for translating a natural language description into a logical expression before computing the semantic and statistical similarity amongst the question and query or changing into Structured Query Language (SQL) query [3]. In NLP, the QAS is a progressive Information Retrieval System (IRS), which is employed for answering the questions that are expressed by users in a natural language format [4]. Moreover, intelligent QASs are designed as an advanced information service system incorporating Artificial Intelligence (AI), semantic analysis, information retrieval, and NLP. This system is comprised of three major components, namely answer extraction, information retrieval, and question analysis. Based on these components, the system has the ability to offer users more convenient, fast, and precise answering services [5].

In NLP, answering questions is regarded as a challenging task, since machines are required to understand certain text passages for generating accurate answers [6] [7]. More recently, AI technologies have been developed as a robust technology, with the capacity to deal with machines for executing repeated tasks that usually need human intelligence, such as decision-making, perception, perception, and reasoning [8]. Large pre-trained language methods termed as LLMs include hundreds, billions, or even trillions of parameters, yet they exhibited emergent skills and produced more promising outcomes on multiple NLP tasks. In daily life, LLMs have become increasingly common because of their exceptional performance in production practices, personal assistants, and human-machine dialogue [9]. Recently, DL techniques have usually adopted advanced Knowledge Base Question Answering (KBQA) systems for addressing the complexity in the human language. Further, this system is used for managing ambiguities, identifying the entities and their relations, interpreting complex questions, and developing trust in the capacity of the system for handling the uncertainties [10]. Additionally, Graph Neural Networks (GNNs) are employed by a few KBQA systems for

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

analyzing and interpreting the knowledge graphs effectively, enabling more precise mining of important information on the basis of the knowledge base structure [11] [12]. In the QAS field, numerous DL models have emerged, which include Bidirectional Long Short-Term Memory (BiLSTM), Document Reader for Question Answering (DrQA), Long Short-Term Memory (LSTM), Bidirectional Encoder Representations from Transformers (BERT) [13], Generative Pre-trained Transformer (GPT), Question Answering Network (QANet), Document Reader for Question Answering (DrQA), CNN-attention [14] and A Lite BERT (ALBERT) [6] [7].

This survey is carried out to investigate several approaches applied to QAS. Further, the techniques are categorized into numerous classes, such as ML-based, knowledge graph-based, DL-based, and LLM-based models. 25 research papers considered in this survey were sourced from standard platforms, including IEEE, MDPI, Wiley Online Library, and so on. In addition, the assessment highlights main characteristics, like the most commonly employed tools, evaluation metrics, datasets, categories, and publication years. At last, the survey discusses the challenges and issues faced in the reviewed papers.

The residual sections of this survey are ordered as below: Section 2 presents the general outline with steps for QAS. Further, Section 3 gives categorization based on QAS. The literature survey of the gathered research papers is given in Section 4, the research gaps and drawbacks faced by QAS are discussed in Section 5, Section 6 presents the analysis of the works reviewed and lastly, the conclusion of the survey is presented in Section 7.

2. GENERAL OUTLINE OF QAS

QAS processes a user's question and a specific passage to generate an accurate answer. The major elements involved in this procedure are signified beneath,

i) Input: Question and Passage

The procedure starts with two prime inputs, such as a question and a passage. The question signifies what the user is seeking to know, whereas the passage involves the context from which the answer may be derived.

ii) Question Processing

Once the question is received, it is forwarded to the question processing phase. Here, the scheme evaluates the question's intent and structure. It also identifies the question type, extracts the relevant keywords, and comprehends the semantics for the purpose of directing answer extraction.

iii) Passage Processing

Concurrently, the designated passage was sent to passage processing. In this stage, the text is analyzed to identify important information, which is then broken down into smaller, more manageable pieces to facilitate correlation with the question.

If required, the scheme may also recover content to enhance the passage.

iv) Source of Answer

In order to optimize the information accessible for answering the question, the scheme accesses an additional source of answers. These sources can involve documents and the web. Further, the sources allow the system to acquire more comprehensive data, specifically if the passage does not provide enough information to answer the question precisely.

v) Answer Generation Module

In this module, the processed question and passage are fed together. It exploits modern algorithms based on NLP and ML, especially DL techniques, like Transformer-based architectures that employ attention mechanisms to model complex relationships between the question and passage. For answer retrieval, these techniques frequently perform span prediction, selecting the accurate start and end tokens in the passage that best answer the question. In addition, retrieval-based approaches can combine ranking algorithms to select the most relevant passages from a huge corpus before answer excerption, incorporating both retrieval and comprehension in a combined model.

vi) Output: Generated Answer

Lastly, the answer is generated by the system on the basis of earlier processing and matching phases. It is then presented to the user as the outcome of the scheme. The intention is to deliver the precise and relevant response that directly addresses the question posed by the user. In addition, the general outlines of the QAS process are signified in Fig. 1.

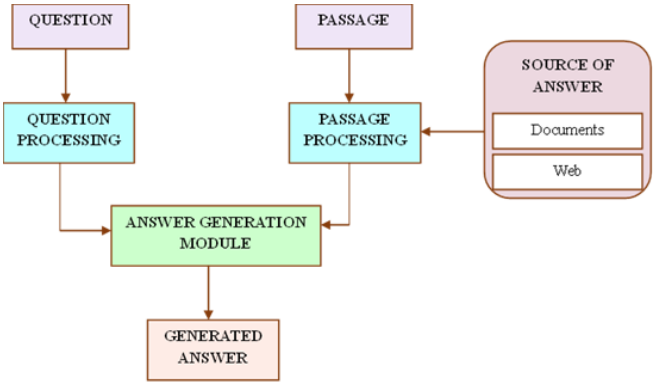

Fig. 1 - General Outline of QAS

3. CATEGORIZATION OF THE PAPERS

The research papers reviewed in this survey are classified based on their methodologies in QAS. The collected papers are divided into four major groups, such as LLM-based methods, DL-based approaches, and knowledge graph-based methods. This categorization assists in understanding the evolution of approaches and detecting drawbacks and research gaps in the research. The categorization of the QAS schemes is outlined in Fig. 2.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

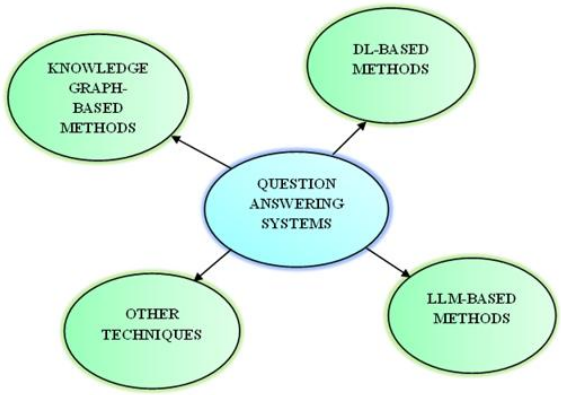

Fig. 2 - Categorization of QAS Techniques

4. DESCRIPTION AND LITERATURE SURVEY OF QAS

A comprehensive literature survey of the approaches, including LLM-based models, DL-based models, knowledge graph-based models, and ML-based models exploited for QAS, is specified in the subsequent section.

4.1 DL-based Techniques

In QAS, deep learning (DL)-based methods use neural networks to comprehend and produce answers to unstructured documents. With models like BERT and GPT, these methods acquire a contextual understanding of word relationships, which leads to more accurate and fluent responses. The use of DL methods alleviates the burden of manual feature engineering and handles the intricacies of language. They achieve remarkable performance across different disciplines.

i) Literature Survey of DL-based models

Wang, H. [15] described an Automatic Question Answering (AQA) model aimed at QAS. With the AQA strategy, he improved the accuracy and response time by a significant margin and optimized the TF-IDF weight computation to a great degree. Furthermore, it showed remarkable performance in fulfilling the daily demands of text-based question answering and online teaching and thus, evident practical value for students and teachers. Nonetheless, it was predominantly focused on text-based Q&A, thus, lacking support for any image or speech recognition, which limited its overall potential in addressing the student question-answering requirements. A BiLSTM model was developed by Katib, S.S.R. and Abdulameer, M.H., [7] for QAS. A BiLSTM is a kind of neural network termed for its efficacy in natural language understanding and text analysis. This approach enhanced precision by contemplating both preceding and following context in sentences. Moreover, preprocessing was applied to eradicate unwanted data, and the SQuAD 2.0 database was employed for training. Nonetheless, this technique was neither extended to numerous languages nor combined with attention mechanisms to better focus on relevant input features.

The PAL-BERT model was introduced by Zheng, W., *et al.* [2] for QAS. The PAL-BERT technique was devised by introducing a first-order network pruning approach on the basis of A Lite BERT (ALBERT) model. Here, a parameter optimization technique was created which replaced the ReLU activation function with the Mish function for better performance optimization. While the pruning-based compression technique improved training efficiency and substantially reduced training time, the approach still had high complexity which may lead to loss of accuracy due to pruning.

In [16], Shi, D., et. al. introduced the Fundus Fluorescein Angiography-Generative Pre-trained Transformer (FFA-GPT) for visual QAS. This technique consisted of two parts, the first being the interactionaling Q&A using the GPT-based module (Llama 2) and the second was an image-text alignment module for generating reports. In addition, the performance of this model was assessed using both automated and manual evaluation methods. The method exhibited outstanding performance with respect to generating reports that were grammatically correct, coherent, and semantically aligned. Nonetheless, this approach lacked validation, exploiting a fully external database, relying instead on time-split data, and further analysis and optimization of its generalizability were required for various spatial data and entirely new data.

COBERT technique was devised by Alzubi, J.A., *et al.* [17] for QAS. This scheme was designed to assist researchers and clinical workers in accessing authentic COVID-19 scientific data easily. Moreover, this model employed a three-component architecture, such as Retrieval, Reader, and Ranker, to capture query similarity with large text databases. An advantage of this technique was its portability, allowing it to be easily modified to various domains. However, this model was developed on the structured CORD-19 database, which limited its ability to manage unstructured data.

Zhang, J., *et al.* [3] introduced a Hierarchical Semantic Matching-Question Answering (HSM-QA) for QAS. The HSM-QA technique was devised with two major steps, including query-answer matching employing a single-stream structure to rank relevance of candidate answers, and query-question matching employing a Siamese network to identify similar questions and their answers. This method resulted in outstanding results in several metrics, demonstrating operational practicality. That said, the model did not improve generalization capabilities because it did not integrate other knowledge sources, particularly knowledge graphs, and other external elements.

Mallikarjuna, C. and Sivanesan, S., [18] developed TweetRoBERTa for Question Answer Systems. They worked on Tweet Question Classification to improve tweet-based Question Answering Systems by classifying tweets into various questions. This classification helped tackle the informal, elliptical, noisy tweets for response generation at a higher accuracy and system efficacy in a context-sensitive manner. A positive aspect of the model is the enhanced ability to preprocess structured queries for more precise user intent recognition. However, the integration of the Expected Answer Type was missing, which stunted the model's precision and efficiency in a greater manner.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

In [19], Qiu, Q., and others, worked on a BERT-based Question and Answer system in the context of Question Answer Systems. This system aimed to create automatically generated Q&A material based on ontology and structured definitions, culminating in a prototype of a chatbot system integrated with the WeChat ecosystem which was particularly high in accuracy relative to the computational demands and was thus effective. Nevertheless, this method needed a group of additional domain texts and reports to optimize BERT's training data and enhance the semantic understanding of domain-specific vocabulary.

The Interactive Attention-BiLSTM (IA-BiLSTM) model was introduced by Xingguang, L., *et al*. [20] for QAS. The IA-BiLSTM model, incorporating BiLSTM with interactive attention, was established to compute semantic similarity among sentences by capturing semantic associations via inter-sentence attention. This model enhanced the precision of similarity correlations and assisted pressure vessel designers in rapidly detecting answers to common design problems. This method was efficient in supporting intelligent manufacturing via precise sentence correlation. But this model relied solely on design standards and did not combine real-time data from the Internet of Manufacturing Things (IoMT), limiting its adaptability.

In [21], a Siamese LSTM-fusion model was presented by Liu, S., *et al*. for QAS. This approach was devised to learn different feature representations of questions by incorporating a Siamese LSTM network with the fusion MatchPyramid technique, combining both word significance and context for multi-semantic expression. Further, this approach revealed strong learning ability and handled information loss caused by MatchPyramid's single-channel matching, thus enhancing question answering performance. Though this model optimized semantic understanding, the model's increased complexity led to higher computational costs during tuning and inference.

A robust technique termed the BERT-Softmax model was devised by Duan, K., *et al*. [22] for QAS. This method was designed to optimize candidate entity features and bridge the semantic gap among questions and the knowledge base. This approach exploited problem encoding and candidate information excerpted by BERT, revealing strong generalization across multi-domain knowledge bases. The pros of this model were its efficacy in enhancing entity disambiguation. But further enhancements, such as tuning different subtasks in incorporating additional knowledge graph representation learning approaches, were required to acquire richer features.

Lyu, P., *et al.* [23] introduced a GlobalPointer and BiLSTM (GPB) for QAS. The approach combined BiLSTM into GlobalPointer via concatenation to assure precise detection of entities and their relationships. To optimize user experience in an intelligent motor fault maintenance QAS, the BiLSTM+Attention+Conditional Random Field (CRF) (BAC) was designed for named entity recognition, whereas BERT-Whole Word Masking (BERT-wwm) was exploited for user intent classification, enhancing answer quality. Additionally, this approach contributed to reduced downtime and less loss in production costs. However, this approach did not incorporate graph convolutional networks and CRF regarding dynamic updates on knowledge graphs, which affected the level of flexibility.

Behmanesh et al. [12] introduced ColBERTv2 for QAS, which was a token-level interaction scalable retrieval enhanced approach focused on managing complicated queries, which included multiple relations and entities. This model showed remarkable improvement when dealing with large scale hybrid databases such as Simple Questions and Reverb Simple Questions where it reached new levels of accuracy and improvement and large scale hybrid databases. An additional value was the strong flexibility and precision in dealing with queries of varying complexity. However, this approach did not integrate multiple knowledge sources at different levels, improve precision and recall on entities and relations, address dataset imbalance, and errors which has an impact on the overall robustness.

Gao et al. [5] improved BERTserini for QAS. The first stage focused on text segmentation using a preprocessing strategy designed for multi-document legal texts followed by Anserini to excerpt highly relevant paragraphs.

The second phase applied a two-step fine-training of the Chinese BERT technique for accurate answer generation, also recovering chapter, document, and page details. This technique enhanced answer accuracy and traceability. Nonetheless, this approach needed significant computational resources for fine-tuning.

4.2. Knowledge Graph-Based Methods

Knowledge graph-based techniques employ structured data in the form of entities and relationships to answer questions by detecting relevant connections. They offer precise, interpretable outcomes, particularly for fact-based queries. These approaches are highly efficient in domains with well-defined knowledge graphs.

i) Literature Review of Knowledge Graph-Based Methods

Multi-hop English Language Teaching (ELT) knowledge test method was developed by Wang, L., [24] for QAS. The enhanced multi-hop ELT knowledge test technique was designed on the basis of knowledge graph embedding, establishing a relational link scoring module to output all candidate entities on the similar relational link, addressing answer omission, and optimizing robustness. Further, the question embedding approach was optimized to better capture English semantics for the ELT domain. This technique attained link prediction ability and strong inference on incomplete knowledge graphs. Still, this model did not combine representation learning to filter ideal solutions from candidate answers, limiting further enhancement in test quality.

Isah, M.A. and Kim, B.S., [8] established QAS for Tunnel project Risks (QASTRisk). The QASTRisk aimed to minimize time and effort in risk detection during the preconstruction phase of tunnel projects and optimize risk management efficacy. This technique offered project managers with an AI-based technique to automatically retrieve tunnel risk information. But the knowledge graph lacked robustness for other tunnel kinds and construction

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

projects, limiting its broader applicability. Moreover, additional development was required to extend its use to various infrastructure projects.

In [25], a template-based approach was devised by Dhandapani, A. and Vadivel, V., for QAS. This method was devised for classifying question types and mapping them to suitable SPARQL query templates, such as superlatives and comparatives, for implementation on the DBpedia endpoint. This technique demonstrated potential in managing a broad range of factoid questions more efficiently. A significant benefit of this model was its flexibility in dealing with different question types. Nonetheless, this method did not support total question types and needed manual addition of novel templates for broader coverage.

4.3. LLM-based Methods

LLM-based techniques inspire LLM to understand questions and generate human-like answers. Further, they excel at managing complicated reasoning, queries, and generating fluent responses from unstructured text. These approaches require minimal task-specific training and adapt well across domains.

i) Literature Review Of LLM-based Methods

Lai, H., *et al*. [26] introduced the Website General Language Model (WebGLM) for QAS. An efficient LLM-based retrieval QA scheme termed WebGLM was designed by combining enhancements from practical deployment experience. This technique employed in-context learning and a rigorous filtering technique to generate a large, high-quality database with referenced long answers for tuning, alongside a self-check mechanism to minimize hallucination from low-quality references. A significant advantage of this approach was its cost-effectiveness and dependability in real-world scenarios. Still, the self-check module did not fully remove hallucinations in complex queries.

An efficient model termed Automatic Identification System Stream-Model Context Protocol (AISStream-MCP) technique was introduced by Chen, S., *et al*. [27] for QAS. This technique assimilated an LLM with four MCP-enabled components, like knowledge graph lookup, live AIS data query, persistent dialogue memory, and result assessment to optimize maritime situational awareness and operational efficacy. Using this method in maritime operations showed promising improvement in decision making. One of this method's strength is the integration of different data sources to perform in-depth analysis. That said, to evaluate the system's generalizability, scaling up with a larger, more varied set of queries, is further required.

For Questions Answer Systems, in [9], Wang, F., et al. described the LLM-Knowledge Graph Multi-hop Question-Answering (LLM-KGMQA). Knowledge graph-based multi-hop question answering involved the formulation of a three-step multi-hop knowledge path reasoning method and an entity fast-linking algorithm. To ensure accurate linking, the technique ranked entities via multi-attribute classification, feature intersections, similarity calculations, and a pre-trained model. This method provided an efficient tool for precise knowledge retrieval in the hands of healthcare professionals and patients. This model's lack of the more advanced dynamic contextual embeddings and augments for fine-tuning that deal with the medical field for more intricate, contextually rich queries is a setback, however.

4.4. Other Techniques

Do, T.P.P., et al. [28] developed a Retriever-Reader-Generator Question Answering (R2GQA) system for QAS. The R2GQA system incorporates a Machine Reader, a Document Retriever, and an Answer Generator. With this system, an even more compact design which was efficient in resource and operational cost savings was achieved. However, this model also did not prioritize improving and tuning the language models to some extent, considering the contextual and linguistic specifics of the Vietnamese language.

Soni, S., et al. [29] for QAS developed querHy. querHy was developed to answer natural language questions drawn from structured Electronic Health Records (EHR) and retrieve the presented answer in an understandable manner. This method involved some degree of semantic parsing, time frame classification and concept normalization, a query module for FHIR mapping and processing, and visualization. With this method, enhanced accuracy, reliability, and scalability were gained. However, this technique still required manual intervention to some extent to broaden the mappings from phrases to logical predicates and FHIR resources.

Aithal, S.G., et al. [1] proposed a Question Similarity mechanism for QAS. This was designed to identify unanswerable, irrelevant questions, and filter them with human-like reasoning. This model allowed QAS to narrow down to only answerable questions. This helped improve overall performance with lower computational costs. This technique minimized capturing irrelevant input. However, this model had a problem with high time complexity and over errors in some cases. Wu, S., et al. [30] developed DIETNERD for QAS. This was proposed in order to strengthen dietary recommendations and foster a collaborative relationship between patients and healthcare providers. The model's improvement in personalized dietary support was one of its strengths. However, the knowledge base was shallow and therefore more scientific literature needs to be integrated to strengthen this model. In [4], a Concept Question Answering system applied to the Computer Domain (CQACD) model was devised by Wen, Y., et al. This model leveraged a domain ontology with rich semantic relationships to model basic computer knowledge and built a concept-centric knowledge approach. Additionally, this method utilized 80 input templates built with description logics that focused on capturing students' question objectives, and used a textual entailment algorithm to pair inputs and assess contributions to enhance the technique's flexibility. Also, an ontology-driven dialogue management module was implemented to automate the generation of the conversational content and order efficiently. Yet, the absence of voice input/output

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

capabilities in this configuration constrained user interaction and access.

5. RESEARCH GAPS AND ISSUES

The research challenges faced while employing DL-based approaches are detailed in this section. In [15], the AQA technique did not combine suitable algorithms for image and voice recognition, which was crucial for attaining more effectual and precise Q&A capabilities. The BiLSTM approach [7] did not enhance the interpretability of the model and efficacy in complicated scenarios. Furthermore, this model failed to consider numerous languages and combine an attention mechanism for enhancing the ability to dynamically focus on suitable features of input data. In [16], the FFA-GPT model required reevaluation and optimization for improving its generalizability across different spatial databases and completely novel data. COBERT in [17] did not contemplate the multilingual capabilities or combine real-time content from trusted online sources. HSM-QA devised in [3] for QAS failed to optimize the generalization, and it did not combine an external knowledge graph for enhancing the overall performance and usefulness. In [18], the TweetRoBERTa model failed to combine the expected answer type for enhancing the accuracy and efficacy of the tweet QA scheme. The BERT-based Q&A [19] technique was required to gather domain texts and reports to the BERT training data sources for enhancing the semantic properties between domain vocabularies. In [5], improved BERTserini required numerous domains for enhancing the generalizability of the algorithmic procedure, and did not consider algorithmic iterations for optimizing the efficacy of the QAS.

The challenges of prior knowledge graph-based techniques can be summarized as follows: The QASTRisk technique [8] failed to consider a more complicated ontology for enhancing the quality of decision-making. Also, this model did not minimize the time taken for processing. The Template-based approach in [25] needed the development and combination of several templates to handle real-time questions, which added complexity and limited scalability.

Conventional LLM-based techniques faced the following challenges: In [27], the AISStream-MCP model failed to test more advanced LLMs, like Claude-3 and GPT-5 for assessing performance enhancements. The LLM-KGMQA model in [9] did not combine dynamic contextual embeddings or fine-tuning techniques specifically devised for medical terminologies, limiting the model's ability to efficiently handle complex, context-rich queries.

The major limitations faced by other techniques are as follows: The R2GQA technique in [28] failed to explore and combine modern approaches, like reinforcement learning and DL, to further enhance the accuracy and performance of the scheme. In [29], the quEHRy model needed manual development of mappings among logical predicates, phrases, and FHIR resources, which was time-consuming and prone to errors. In [4], the CQACD model did not refine the domain ontology or optimize the input template library in CQACD, which could have further enhanced its total performance.

6. RESULTS AND ANALYSIS

This section deliberates the evaluation and examination of the numerous QAS reviewed in this survey. The research papers are assessed according to several criteria, including techniques, tools utilized, evaluation measures, year of publication, and dataset exploited.

6.1 Analysis based on Methods

This section evaluates the various research works based on the techniques used for QAS. The techniques are classified into different groups, namely DL, knowledge graphs, LLM, and others. As portrayed in the figure below, DL-based methods are the most common, with 14 papers utilizing them. In addition, 3 papers on knowledge graph and LLM-based approaches each, and 5 on other categories. This analysis demonstrates that DL-based techniques are the most frequently employed in QAS. In Fig. 3, the categorization of QAS models is illustrated.

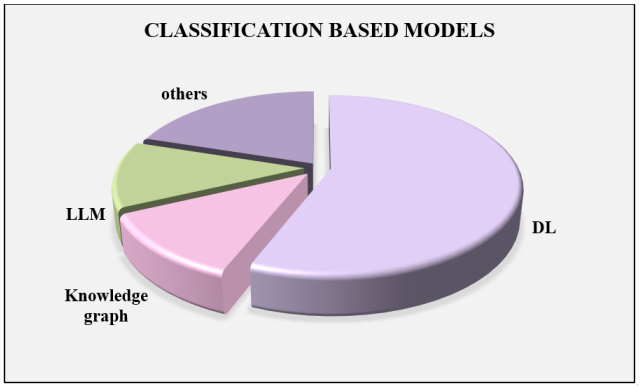

Fig. 3 - Analysis on the basis of Approaches

6.2 Analysis based on Tools

Fig. 4 illustrates the assessment of techniques on the basis of the tools utilized. The tools used for implementation include Python and Java, with Python being the most commonly employed, appearing in 6 research papers, whereas the Java tool is employed in only 1 research paper.

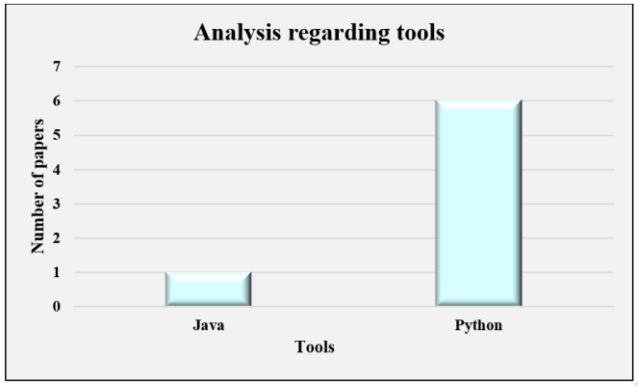

Fig. 4 - Analysis Concerning Tools

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

6.3 Analysis based on Published Year

Fig. 5 specifies an evaluation of the models concerning the count of research papers published year-wise. Specifically, 7 papers were published in 2023, 6 in 2025, 5 in 2024, 4 in 2022, and 3 in 2021. Further, this Figure reveals that the maximal number of papers was published in the year of 2023.

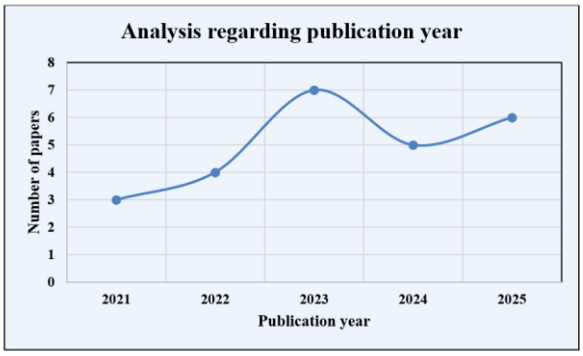

Fig. 5 - Assessment based on the Publication Year of Research Papers

6.4 Analysis based on the Dataset

Fig. 6 designates the examination of the works considering the datasets employed in the reviewed research. The datasets, namely Reverb Simple Questions, ViRHE4QA, MySQL, SQuAD 2.0, CMRC 2018, FFA, CORD-19, Quora Duplicate Question, QNLI, Liao Dynasty Question-and-Answer, CQA, baike2018qa, LCQMC, NLPCC 2016 DBQA, AIS, Chinese QA Pair, DBpedia Neural Question Answering, WebQA, TRisKG, CN-DBPedia, WebGLM, TweetQA-EAT, QASystemOnMedicalKG, QALD-8, QQP, CCKS, Liao Dynasty History QAB, Response Dataset, NLPCCICCPOL 2016 KBQA, motor fault maintenance, and Delta Reading Comprehension, are utilized. Among these, the SQuAD 2.0 dataset and CMRC 2018 dataset are the most commonly employed, each appearing in 3 research papers, whereas all other databases are exploited in only 1 research paper each.

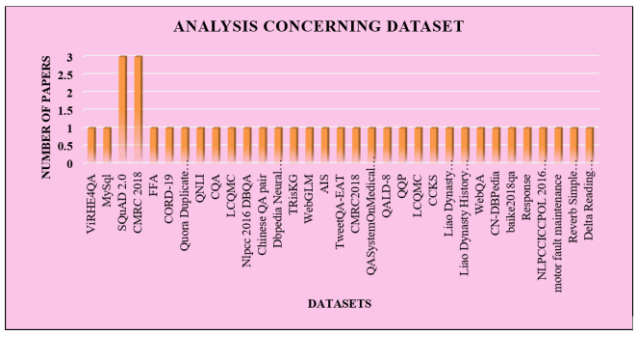

Fig. 6 - Analysis on the basis of the Database

6.5 Analysis based on Evaluation Measures

Fig. 7 signifies the valuation on the basis of the performance measures exploited in the reviewed papers. The measures used in the various works include, such as Recall, F1-score, Mean Average Precision (MAP), Recall-Oriented Understudy for Gisting Evaluation (ROUGE), Accuracy, Bilingual Evaluation Understudy (BLEU), Mean Reciprocal Rank (MRR), Precision, and Exact Match (EM). Among these, the F1-score metric is the most widely used, appearing in 15 research papers. Moreover, precision and recall were employed in 10 and 11 papers, correspondingly, while accuracy was exploited in 8 research papers. Further, EM was utilized in 3 papers, and BLEU, ROUGE, and MAP metrics were each employed in only 1 paper. This assessment reveals that the F1-score is the most frequently exploited evaluation metric in QAS.

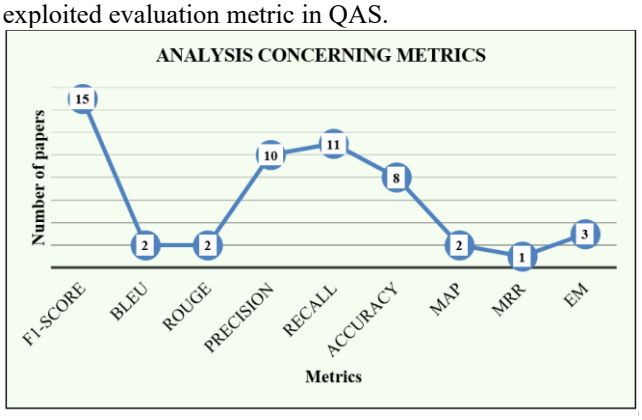

Fig. 7 - Analysis Concerning Evaluation Metrics

6.6 Values of Evaluation Metrics

Table 1 presents the evaluation of research papers on the basis of the F1-score range quantified. In this case, the paper [19] acquired the F1-score of above 95%, while papers [8], [18], and [4] recorded the F1-score between 91% to 95%. Moreover, the research papers [24] and [17] gained an F1-score ranging from 86% to 90%. Further, the maximal F1-score value of below 85% is measured in the research papers [15], [2], [3], and [5].

Table 1 - Analysis based on F1-score

Range of F1-score in %	Number of papers
Above 95%	[7]
91% to 95%	[8] [18] [4]
86% to 90%	[24] [17]
Below 85%	[15] [2] [3] [5]

7. CONCLUSION

A QAS is designed to automatically offer precise answers to user queries by understanding natural language, recovering relevant information, and delivering precise responses effectually across numerous domains. This work presents a comprehensive survey of numerous techniques for QAS. A total of 25 research papers were gathered from various

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

sources and reviewed, with a focus on approaches on the basis of DL-based methods, knowledge graph-based methods, and LLM-based approaches. Among these, Deep Learning-based strategies received more attention within this survey. The identification of key barriers, shortcomings, research gaps, and challenges within each of these categories. Afterward, the criteria utilized in the reviews included, but were not limited to, performance indicators, tools used, year of publication, and the associated dataset. The primary source material for the paper compilations was reputable sources such as Wiley Online Library, IEEE, and MDPI. The references noted a predominance of Python as a programming language, while the SQuAD 2.0 and CMRC 2018 datasets were the most popular for training and evaluation. The subsequent chapter will focus on the incorporation of more innovative tools and approaches to the challenges laid out within the research.

REFERENCES

[1] S. G. Aithal, A. . B. Rao and S. Singh, "Automatic question-answer pairs generation and question similarity mechanism in question answering system," Applied Intelligence, vol. 51, no. 11, pp. 8484-8497, 2021.

[2] W. Zheng, S. Lu, Z. Cai, R. Wang, L. Wang and L. Yin, "PAL-BERT: an improved question answering model," Computer Modeling in Engineering & Sciences, pp. 1-10, 2023.

[3] J. ZHANG, J. HE, Y. ZHOU, X. SUN and X. YU, "HSM-QA: Question Answering System Based on Hierarchical Semantic Matching," IEEE Access, 2023.

[4] Y. WEN, X. ZHU and L. ZHANG, "CQACD: A concept question-answering system for intelligent tutoring using a domain ontology with rich semantics," Ieee Access, vol. 10, pp. 67247-67261, 2022.

[5] M. Gao, M. Li, T. Ji, N. Wang, G. Lin and Q. Wu, "Key technologies of intelligent question-answering system for power system rules and regulations based on improved BERTserini algorithm," Processes, vol. 12, no. 1, p. 58, 2023.

[6] Z. HUANG, S. XU, M. HU, X. WANG, J. QIU, Y. FU, Y. ZHAO, Y. PENG and C. WANG, "Recent trends in deep learning based open-domain textual question answering systems," IEEE Access, vol. 8, pp. 94341-94356, 2020.

[7] S. S. R. Katib and M. H. Abdulameer, "Question Answering System Based on Bidirectional Long-Short-Term Memory (Bilstm)," Al-Furat Journal of Innovations in Electronics and Computer Engineering, pp. 105-120, 2024.

[8] M. A. Isah and B.-S. Kim, "Question-answering system powered by knowledge graph and generative pretrained transformer to support risk identification in tunnel projects," Journal of Construction Engineering and Management, vol. 151, no. 1, p. 04024193, 2025.

[9] F. Wang, D. Shi, J. Aguilar, X. Cui, J. Jiang, L. Shen and M. Li, "LLM-KGMQA: large language model-augmented multi-hop question-answering system based on knowledge graph in medical field," Knowledge and Information Systems, pp. 1-43, 2025.

[10] Y. Lan, G. He, J. Jiang, J. Jiang, W. X. Zhao and J.-R. Wen, "Complex knowledge base question answering: A survey," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 11, pp. 11196-11215, 2022.

[11] M. Blum, G. Nolano, B. Ell and P. Cimiano, "Investigating the Impact of Different Graph Representations for Relation Extraction with Graph Neural Networks," in In Proceedings of the Workshop on Deep Learning and Linked Data , 2024.

[12] S. BEHMANESH, A. TALEBPOUR, M. SHAMSFARD and M. M. JAFARI, "A Novel Open-Domain Question Answering System on Curated and Extracted Knowledge Bases with Consideration of Confidence Scores in Existing Triples," IEEE Access, 2024.

[13] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi and B. Alshemaimri, "BERT applications in natural language processing: a review," Artificial Intelligence Review, vol. 58, no. 6, pp. 1-49, 2025.

[14] M. A. KIA, A. GARIFULLINA, M. KERN, J. CHAMBERLAIN and S. JAMEEL, "Adaptable closed-domain question answering using contextualized CNN-attention models and question expansion," EEE Access, vol. 10, pp. 45080-45092, 2022.

[15] H. Wang, "Automatic question-answering modeling in English by integrating TF-IDF and segmentation algorithms," Systems and Soft Computing, vol. 6, p. 200087, 2024.

[16] X. Chen, W. Zhang, P. Xu, Z. Zhao, Y. Zheng and M. He, "FFA-GPT: an automated pipeline for fundus fluorescein angiography interpretation and question-answer," npj Digital Medicine, vol. 7, no. 1, p. 111, 2024.

[17] J. A. Alzubi, R. Jain, A. Singh, P. Parwekar and M. Gupta, "COBERT: COVID-19 question answering system using BERT," Arabian journal for science and engineering, vol. 48, no. 8, pp. 11003-11013, 2023.

[18] C. Mallikarjuna and S. Sivanesan, "Tweet question classification for enhancing Tweet Question Answering System," Natural Language Processing Journal, vol. 10, p. 100130, 2025.

[19] Q. Qiu, M. Tian, K. Ma, Y. J. Tan, L. Tao and Z. Xie, "A question answering system based on mineral exploration ontology generation: A deep learning methodology," Ore Geology Reviews, vol. 153, p. 105294, 2023.

[20] L. XINGGUANG, C. ZHENBO, S. ZHENGYUAN, Z. HAOXIN, M. HANGCHENG, X. XUESONG and X. GANG, "Building a question answering system for the manufacturing domain," Ieee Access, vol. 10, pp. 75816-75824 , 2022.

[21] S. Liu, N. Tan, H. Yang and N. Lukač, "An intelligent question answering system of the liao dynasty based on knowledge graph," International Journal of Computational Intelligence Systems, vol. 14, no. 1, p. 170, 2021.

[22] K. Duan , S. Du , Y. Zhang, Y. Lin, H. Wu and Q. Zhang, "Enhancement of question answering system accuracy via transfer learning and BERT," Applied Sciences, vol. 12, no. 22, p. 11522, 2022.

[23] P. Lyu, J. Fu, C. Liu, W. Yu and L. Xia, "GPB and BAC: two novel models towards building an intelligent motor fault maintenance question answering system," Journal of Engineering Design, pp. 1-21, 2024.

[24] L. Wang, "An Improved Knowledge Graph Question Answering System for English Teaching," Mobile Information Systems, no. 1, p. 3401074, 2022.

[25] A. DHANDAPANI and V. VADIVEL, "Question answering system over semantic web," IEEE Access, vol. 9, pp. 46900-46910 , 2021.

[26] H. LAI, X. LIU, H. YU, Y. XU, I. L. IONG, S. YAO, A. ZENG, Z. DU, Y. DONG and J. TANG, "WebGLM: Towards an Efficient and Reliable Web-Enhanced Question Answering System," ACM Transactions on Information Systems, 2025.

[27] S. Chen, R. Zhao, J.-B. Yang and Y. Huang, "AISStream-MCP: A Real-Time Memory-Augmented Question-Answering System for Maritime Operations," Journal of Marine Science and Engineering., vol. 13, no. 9, p. 1754, 2025.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

[28] P.-T. P. Do, D.-N. D. Cao, K. Q. Tran and K. V. Nguyen, "R2GQA: retriever-reader-generator question answering system to support students understanding legal regulations in higher education," Artificial Intelligence and Law, pp. 1-46, 2025.

[29] S. Soni, S. Datta and K. Roberts, "quEHRy: a question answering system to query electronic health records," Journal of the American Medical Informatics Association, vol. 30, no. 6, pp. 1091-1102, 2023.

[30] S. Wu, Z. Yacub and D. Shasha, "DietNerd: A Nutrition Question-Answering System That Summarizes and Evaluates Peer-Reviewed Scientific Articles," Applied Sciences , vol. 14, no. 19, pp. 2076-3417, 2024.