

Review

Not peer-reviewed version

---

# Deep Learning Based Question Answering System

---

[Nayyab Saeed](#) , Humaira Ashraf , NZ Jhanjhi \*

Posted Date: 25 November 2025

doi: 10.20944/preprints202312.1739.v2

Keywords: deep learning; question answering system; natural language processing; BERT; recurrent neural networks



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# Deep Learning Based Question Answering System

Nayyab Saeed <sup>1</sup>, Humaira Ashraf <sup>2</sup> and NZ Jhanjhi <sup>3,\*</sup>

<sup>1</sup> 1135-FOC/MSCS/F22 Faculty of Computing-Department of CS and IT International Islamic University Islamabad, Pakistan

<sup>2</sup> Faculty of Computing-Department of CS and IT International Islamic University Islamabad, Pakistan

<sup>3</sup> School of Computer Science, SCS Taylor's University, Subang Jaya, Malaysia

\* Correspondence: noorzaman.jhanjhi@taylors.edu.my

## Abstract

Deep learning-based question answering systems have transformed the discipline of natural language processing (NLP) by automating the extraction of answers from textual data. This survey paper provides a captivating overview of these systems, exploring methodologies, techniques, and architectures such as recurrent neural networks (RNNs), BERT model, and transformer models. Extractive and generative approaches are examined, alongside the challenges of handling complex questions, managing noisy input, and addressing rare or unseen words. This survey serves as a stimulating reference, offering valuable insights to researchers and practitioners, fueling innovation and advancement in question answering systems within NLP.

**Keywords:** deep learning; question answering system; natural language processing; BERT; recurrent neural networks

---

## 1. Introduction

Deep learning-based question answering systems have gained popularity over the past few years mainly because of their capacity to achieve something in a rapid and precise way for questions posed in natural language. Deep learning models are used by these systems to comprehend the question's context and collect relevant data from a variety of resources such as text, pictures, as well as knowledge sources. Deep learning-based question answering systems have revolutionized the area of natural language processing (NLP) by enabling automated extraction of answers from textual data. Significant progress has been achieved in creating reliable and accurate question-answering systems as a result of the development of deep learning algorithms and the accessibility of enormous datasets. These systems have become vital in various applications, including information retrieval, virtual assistants, customer support, and educational platforms. Traditional approaches to question answering often struggled to handle the complexities and ambiguities of natural language queries. However, deep learning techniques have shown immense potential in overcoming these challenges by leveraging powerful neural network architectures capable of learning intricate patterns and representations.

In this survey paper, our objective is to provide an in-depth exploration of deep learning-based question answering systems. We will delve into the methodologies, techniques, and architectures that have been successfully employed in developing these systems. From recurrent neural networks (RNNs) to BERT model and transformer models, we will unravel the diverse range of deep learning approaches used to tackle question answering tasks. Furthermore, we will examine both extractive and generative approaches, where answers are either extracted directly from text or generated based on learned representations. Beyond methodologies, our survey will shed light on the key challenges faced by deep learning-based question answering systems. These challenges encompass handling complex questions requiring reasoning and inference, effectively dealing with noisy and ambiguous input, supporting multiple languages, processing non-textual data such as images or audio, and

addressing the difficulties posed by rare or unseen words. Addressing these challenges is vital to advancing the field and improving the performance of question answering systems in real-world scenarios.

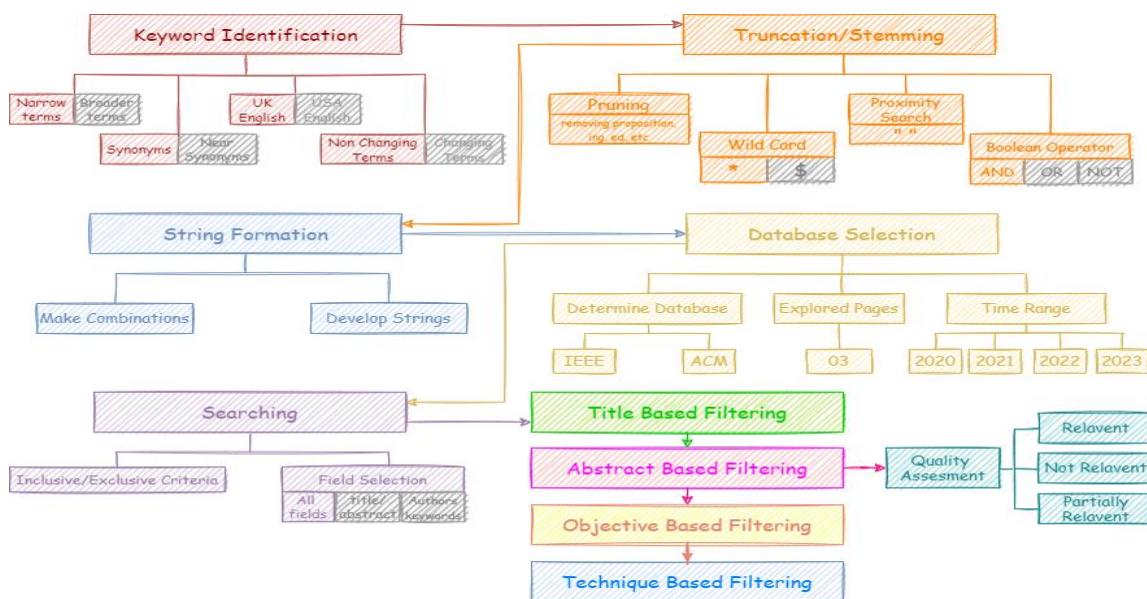
By presenting a comprehensive overview of deep learning-based question answering systems, their methodologies, and the challenges they face, this survey paper aims to provide valuable insights to researchers, practitioners, and enthusiasts in the field of NLP. Our goal is to offer a reference point for understanding the state-of-the-art techniques, sparking innovation, and fostering the development of more accurate and robust question answering systems.

## 2. Systematic Literature Review

A systematic literature review is a purposeful and systematic planning approach to discovering, assessing, and combining every piece of research material related to a specific research issue or subject of interest. It involves a rigorous and transparent process of searching for, selecting, appraising, and summarizing existing studies to give an in-detail and objective summary of the data that is accessible on the topic. A systematic literature review usually consists of multiple phases, such as identifying the research question, creating a search plan, identifying and choosing relevant research, evaluating the integrity of those that were included, obtaining and analyzing data, and combining the results.

### Activities of Systematic Literature Review

The following activities are performed to conduct SLR in this research as shown in Figure 1.



**Figure 1.** The Overview of SLR in Research.

### 2.1. Keywords Identification:

Keywords are words or phrases that represent the main ideas or concepts of a topic. To identify relevant keywords, it's important to brainstorm a list of terms, refine and prioritize the list based on relevance and specificity, and test and adjust the keywords through trial searches. This helps researchers to focus and refine their search for relevant information. In this research following criteria is defined for this activity.

### 2.1.1. Broader/Narrow Terms:

Broader terms refer to more general concepts or ideas, while narrower terms refer to more specific subsets within those concepts. Using both broader and narrower terms can help researchers refine their search for relevant information. For example, Digital media, Communication technology or Online platforms are broad terms and Facebook, Twitter, Instagram, TikTok, YouTube are its narrower terms. Starting with broad terms and gradually narrowing the search, or starting with narrow terms and expanding the search, can ensure a comprehensive and targeted search for information. In this research I have selected narrow keywords as shown in Table 1. For “Deep Learning”, Artificial Intelligence, Context Aware, Deep Architectures, and alike broad terms are used whereas for “Question” I have focused only on terms related to “Answering System”.

**Table 1.** Keyword Identification for the Problem.

<i>Deep Learning</i>	<i>based</i>	<i>Question Answering System</i>
<i>Artificial Intelligence</i>	-	<i>Query Answering System</i>
<i>Context Aware</i>	-	<i>Automated Answering System</i>
<i>Deep Architectures</i>	-	-
<i>Deep Models</i>	-	-
<i>Deep Neural Network</i>	-	-
<i>Knowledge</i>	-	-
<i>Machine Learning</i>	-	-

### 2.1.2. Synonyms/Near Synonyms:

There should be a relevancy between selected synonyms, such as workers, employees, personnel. In this survey three synonyms are identified for keyword “question” and seven are for “deep learning” which are selected on close relevance as shown in Table 1.

### 2.1.3. UK/US English:

UK English and US English are two variants of the English language that have differences in spelling, pronunciation, vocabulary, and grammar. Some examples include differences in the spelling of certain words, pronunciation of certain words, meaning of certain words, and usage of certain grammar constructs. It’s important to be aware of these differences and use the appropriate variant based on the context and audience. I have focused on UK English because it is a standard for technical write ups like research articles.

### 2.1.4. Terminology Change over Time:

Terminology can change over time due to social, cultural, technological, and scientific developments. Factors such as globalization and migration can also influence changes in terminology. It’s important to stay up-to-date with changing language and terminology to effectively communicate and collaborate with others. There is no such keyword in our research problem that might change.

### 2.2. Truncation/Stemming:

Truncation, also called stemming, is a technique used in database searching that involves adding a symbol ( ) at the end of a word to retrieve variations of that word. It can help broaden a search to find more results, but it may also bring up irrelevant results. Truncation should be used carefully and in combination with other search techniques to get the best results. Table 2 shows the results of truncation on identified keywords.

**Table 2.** Steps of Truncation on one of the string.

<i>Deep Learning</i>	<i>Based</i>	<i>Question</i>	<i>Answering System</i>
<i>Deep Learning</i>	<i>Base</i>	<i>Question</i>	<i>Answer System</i>
<i>Deep Learning</i>	<i>Base</i>	<i>Question</i>	<i>Answer System</i>

### 2.2.1. Pruning:

It is the process in which certain words like propositions, 'ing', 'ed' are trimmed from the keywords. For example, "Based" will become "Base" etc.

### 2.2.2. Wild Card:

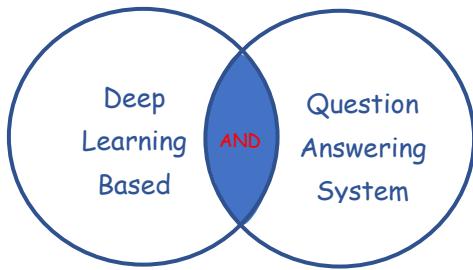
Wildcard is a character that signifies specific meaning to the search engine. For example: "safe" will retrieve "safety", "safely", "safest" etc. where \* is a wild card. In this research wildcards are used on selected for effective search of relevant research articles.

### 2.2.3. Proximity Search:

Double-quoting words allows you to conduct a search for a certain expression, for example, "Deep Learning" and "Question Answering System".

### 2.2.4. Boolean Operators:

When searching for information, words are combined using the Boolean operators AND, OR, and NOT. Look for specific phrases or topics initially, then use the searching history to organise collections of results applying Boolean operators. In this research we use the Boolean operator AND to fetch all relevant researches containing all keywords as shown in Figure 2.

**Figure 2.** Application of Boolean Operator in Search Clause.

### 2.3. String Development:

After defining keywords next step is the formation of substrings. String formation/development is the process of creating search terms or queries that accurately represent the concepts and ideas being searched for. It involves identifying relevant keywords and using them to construct a search string that will retrieve the desired information. Effective string development is important for ensuring accurate and comprehensive search results. The following Table 3 shows the strings which are formed by making combinations of the keywords.

**Table 3.** List of strings developed from selected keywords.

<i>Deep Learning</i>	<i>Based</i>	<i>Question Answering System</i>
<i>Artificial Intelligence</i>	<i>Based</i>	<i>Question Answering System</i>
<i>Artificial Intelligence</i>	<i>Based</i>	<i>Query Answering System</i>
<i>Artificial Intelligence</i>	<i>Based</i>	<i>Automate Answering System</i>
<i>Context Aware</i>	<i>Based</i>	<i>Question Answering System</i>

---

<i>Context Aware</i>	<i>Based</i>	<i>Query Answering System</i>
<i>Context Aware</i>	<i>Based</i>	<i>Automate Answering System</i>
<i>Deep Architectures</i>	<i>Based</i>	<i>Question Answering System</i>
<i>Deep Architectures</i>	<i>Based</i>	<i>Query Answering System</i>
<i>Deep Architectures</i>	<i>Based</i>	<i>Automate Answering System</i>
<i>Deep Learning</i>	<i>Based</i>	<i>Question Answering System</i>
<i>Deep Learning</i>	<i>Based</i>	<i>Query Answering System</i>
<i>Deep Models</i>	<i>Based</i>	<i>Question Answering System</i>
<i>Deep Models</i>	<i>Based</i>	<i>Query Answering System</i>
<i>Deep Models</i>	<i>Based</i>	<i>Automate Answering System</i>
<i>Deep Neural Network</i>	<i>Based</i>	<i>Question Answering System</i>
<i>Deep Neural Network</i>	<i>Based</i>	<i>Query Answering System</i>
<i>Deep Neural Network</i>	<i>Based</i>	<i>Automate Answering System</i>
<i>Knowledge</i>	<i>Based</i>	<i>Question Answering System</i>
<i>Knowledge</i>	<i>Based</i>	<i>Query Answering System</i>
<i>Knowledge</i>	<i>Based</i>	<i>Automate Answering System</i>
<i>Machine Learning</i>	<i>Based</i>	<i>Question Answering System</i>
<i>Machine Learning</i>	<i>Based</i>	<i>Query Answering System</i>
<i>Machine Learning</i>	<i>Based</i>	<i>Automate Answering System</i>

---

#### 2.4. Database Selection:

Database selection is the procedure of finding and picking out the best databases to use while doing a search for knowledge on a particular topic. It involves evaluating the scope, content, and relevance of different databases and selecting those that are most likely to contain the desired information. Effective database selection is important for ensuring that search results are comprehensive, relevant, and reliable. The selection of databases to be searched is an important step for collection of relevant research articles. This activity comprises of following steps.

##### 2.4.1. Determine Databases:

The following highly-ranked two databases are searched in this research.

- IEEEXplore <https://ieeexplore.ieee.org/>
- ACM digital Library <https://dl.acm.org/>

##### 2.4.2. Explored Pages:

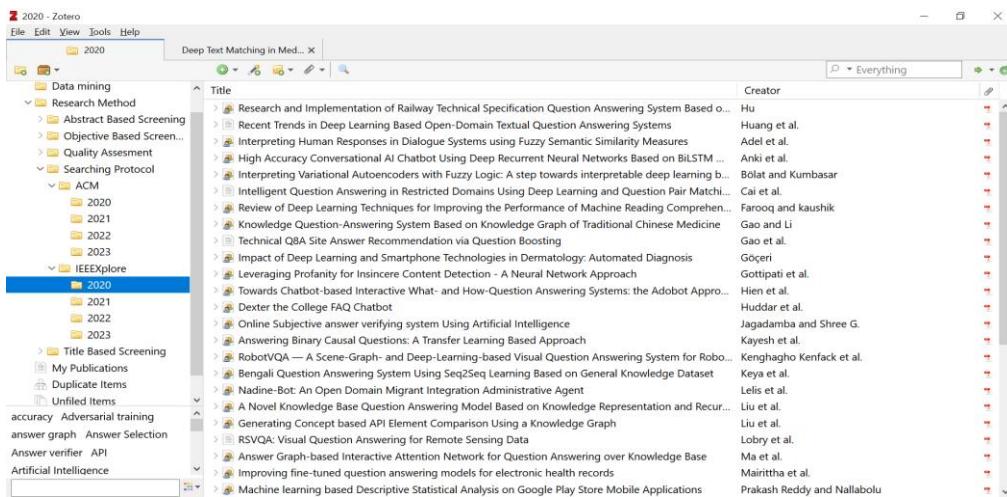
For each input string first three pages of each database is searched.

##### 2.4.3. Time Range

Four folders are created in Zotero for downloading researches of years 2020, 2021, 2022, 2023.

#### 2.5. Searching:

The next step is to execute searching, once strings are ready. All relevant papers are downloaded in Zotero in respective folders year-wise as shown in Figure 3.



**Figure 3.** Zotero Folder View.

During this activity the following steps are performed:

#### 2.5.1. Inclusion/Exclusion Criteria:

The following inclusion/exclusion criterion is defined for this survey.

- Only those articles are downloaded those must have keywords of a string
- Articles published before 2020 are not considered
- The articles for which only citation is available are not considered.

#### 2.5.2. Field Selection:

The number of articles retrieved by a search clause depends on fields selected in a database. Following options are available for field selection.

- All fields: You can use your search terms in any field. There will certainly be a large number of outcomes from this.
- Title/abstract: If the title and abstract contain search terms, the item is probably very relevant. It emphasizes the titles and abstracts with strong, descriptive writing.
- Keyword: Searches for what you entered in the list of terms that the author has provided.

In this research all field option is selected so that any relevant paper does not get missed. The following Table 4 shows the details of searching which was performed in this research. So far we have downloaded 798 articles in folder 2023 and 119 articles in folder 2022.

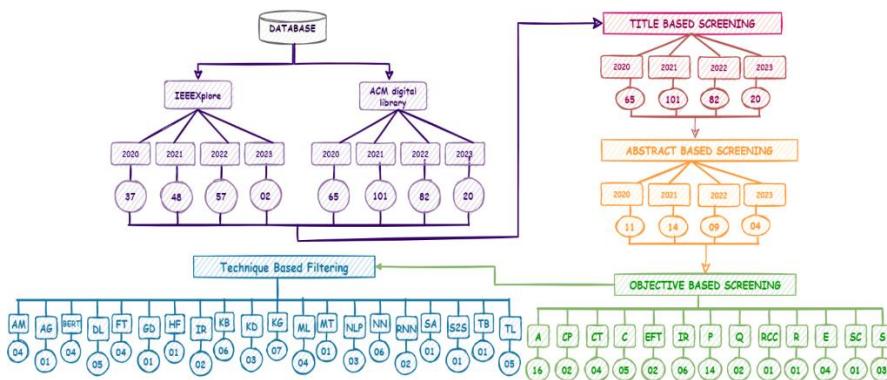
**Table 4.** Searching Details.

Year	Database	Pages Explored	Available Articles (per page)	Related Articles	Total
2022	IEEEExplore	3	20	37	102
	ACM	3	30	65	
2021	IEEEExplore	3	20	48	149
	ACM	3	30	101	
2022	IEEEExplore	3	20	57	142
	ACM	3	30	82	

2023	<i>IEEE Xplore</i>	3	20	02	22
	ACM	3	30	20	

## 2.6. Filtering:

Filtering in research involves carefully reviewing potential sources of information to identify and exclude irrelevant or low-quality studies or data. This helps researchers focus on high-quality and relevant information, which improves the accuracy and trustworthiness of their findings. The detailed filtering process is shown\* in figure+4.



**Figure 4.** Different steps of Filtration+ Process+.

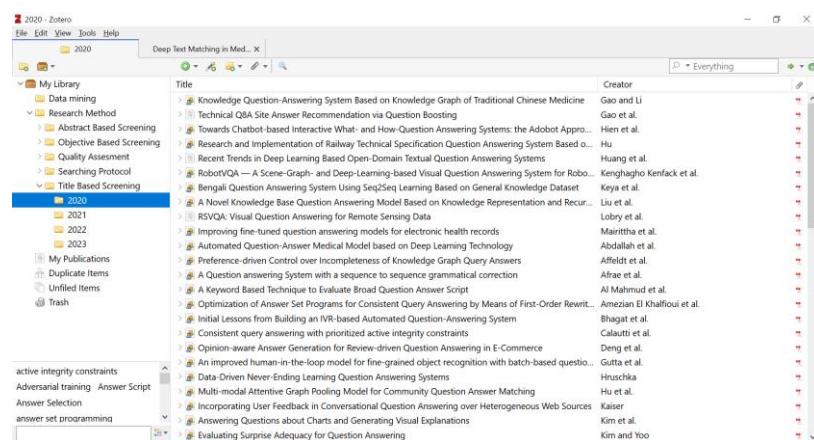
### 2.6.1. Title Based Filtering:

It is the first stage of filtering which is called *title-based* filtering as shown in Table 5. Title-based filtering involves selecting studies based on the information provided in their titles. It is a quick and efficient way to identify potentially relevant studies, but may result in the exclusion of relevant studies that have different titles or use different terminology.

**Table 5.** Title Based filtering.

Title Based Filtering					
Year	2020	2021	2022	2023	Total
Relevant Papers	66	102	94	16	278

Zotero files were made up of papers, but those that weren't useful to the topic at present were eliminated, which is diagrammatically represented in Figure 5.



**Figure 5.** Title Based Filtering folder in Zotero.

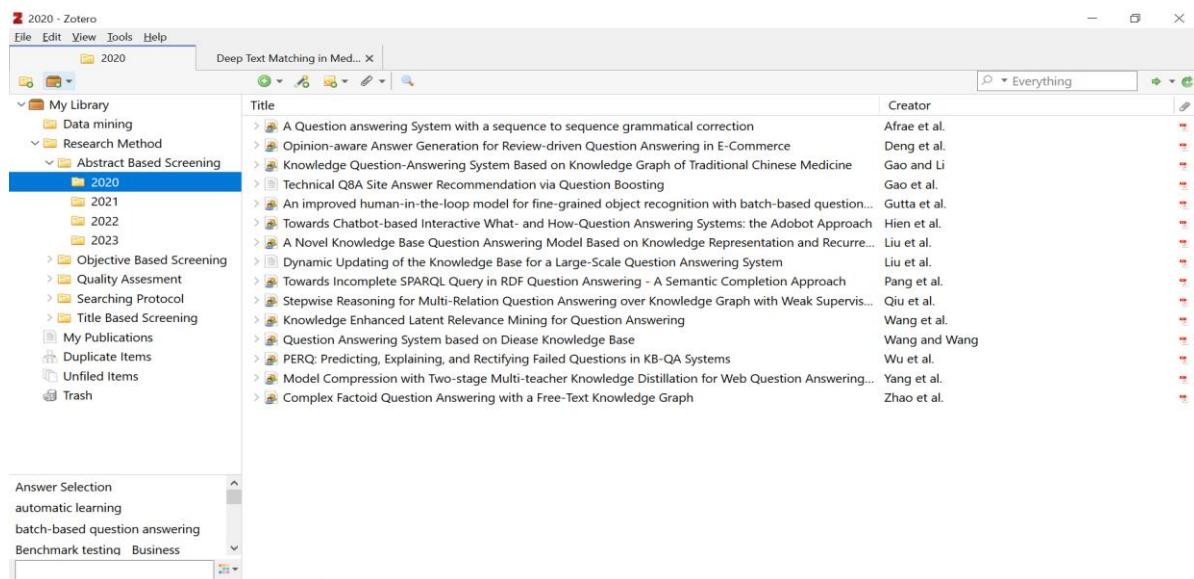
### 2.6.2. Abstract Based Filtering:

In the second level of filtering, abstract-based filtering is conducted. Abstract-based filtering is a screening process that involves selecting studies based on the information provided in their abstracts. It is a quick and efficient way to identify potentially relevant studies, but it may result in the exclusion of relevant studies that do not provide sufficient information in their abstracts. To make sure the studies fit the inclusion requirements and are related to the study issue, a full-text analysis is required. All publications whose abstracts are unrelated to the issue are eliminated from all the chosen sources as shown in Table 6.

**Table 6.** Abstract Based filtering.

<i>Abstract Based Filtering</i>					
<i>Year</i>	<i>2020</i>	<i>2021</i>	<i>2022</i>	<i>2023</i>	<i>Total</i>
<i>Relevant Papers</i>	11	14	09	04	38

All papers, based on their abstracts, which failed to address the topic were excluded from folders created in Zotero which is diagrammatically represented in Figure 6.



**Figure 6.** Abstract Based Filtering folder in Zotero.

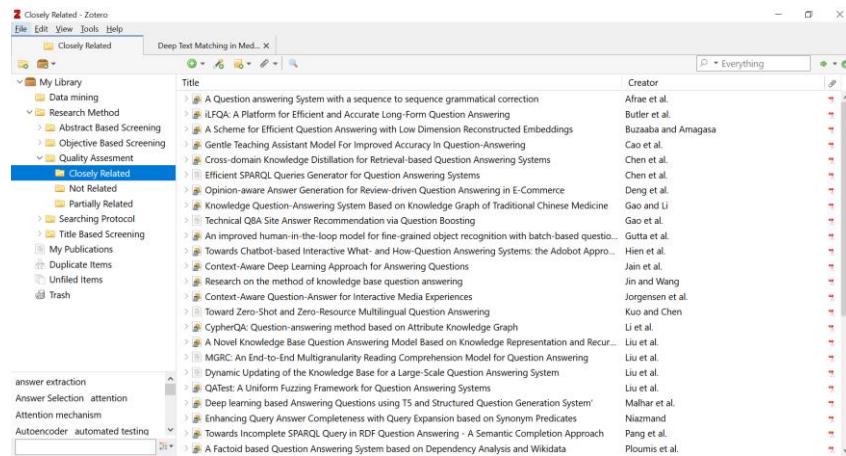
### 2.6.3. Quality Assessment:

In this filtering process I have perform one more filtering process which is quality assessment in which I am determining that how much research articles are related e.g. Closely Relevant, partially relevant or not relevant as shown in Table 7. This step is done while doing abstract based filtering.

**Table 7.** Quality Assessment.

<i>Quality Assessment</i>			
<i>Relevancy</i>	<i>Closely Relevant</i>	<i>Partially Relevant</i>	<i>Not Relevant</i>
<i>Papers</i>	38	77	146

The folder created in Zotero which is diagrammatically represented in Figure 7 contains all the papers that are closely related, partially related or not related.



**Figure 7.** Quality Assessment folder in Zotero.

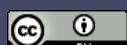
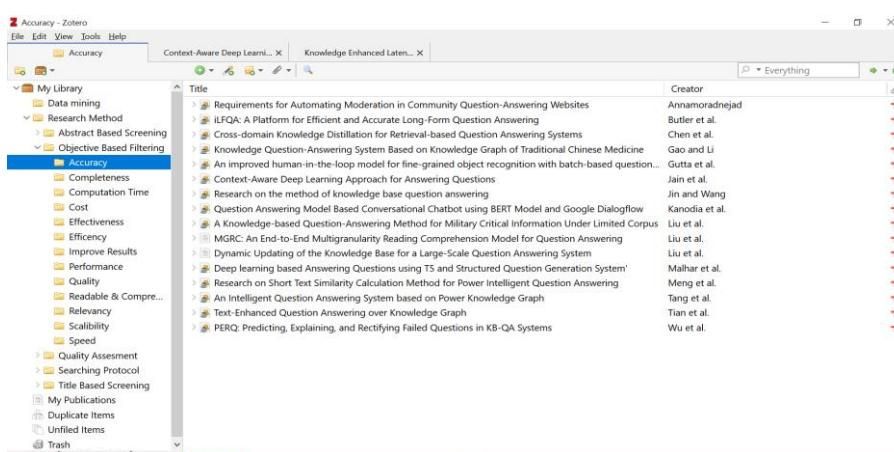
#### 2.6.4. Objective Based Filtering:

Objective-based filtering in a systematic literature review is a method of selecting studies based on predetermined criteria that are directly related to the research question or objective. This assists in ensuring that only high-quality and related papers are reviewed, lowering the possibility of bias and ensuring that the findings are trustworthy and related to the research issue. The categories of objectives were as shown in Table 8.

**Table 8.** Notation Table of terms.

Notation	Abbreviation	Notation	Abbreviation
Accuracy	A	Performance	P
Completeness	CP	Quality	Q
Computation Time	CT	Readability	&
Cost	C	Comprehend	RCC
		Context	
Effectiveness	EFT	Relevancy	R
Efficiency	E	Scalability	SC
Improved Results	IR	Speed	S

Objective-based filtering conducted in Zotero is diagrammatically represented in Figure 8.



**Figure 8.** Objective Based filtering folder in Zotero.

Table 9 illustrates this objective-based screening by listing and grouping the objectives of the research articles.

**Table 9.** Objective Based Filtering.

REF. PAPERS TITLE	OBJECTIVES											
	A	C P	C T	C	EF T	E R	I	P	Q	RC C	R	S C
"Question" answering"System" with "a" sequence" to "sequence" grammatical" correction" [1]	✓	-	-	-	-	-	✓	-	-	-	-	-
Opinion-aware"Answer"Generation"for"Review-driven Question"Answering"in E-Commerce" [2]	-	-	-	-	-	-	-	✓	-	-	-	-
Knowledge"Question-Answering"System"Based on"Knowledge"Graph"of"Traditional Chinese Medicine" [3]	✓	-	-	-	-	-	-	-	-	-	-	-
Technical'Q8A"Site"Answer"Recommendation via"Question Boosting" [4]	-	-	-	-	-	-	✓	-	-	-	-	-
An"improved"human-in-the-loop"model"for fine-grained"object"recognition"with"batch-based question"answering" [5]	✓	-	✓	-	-	-	-	✓	-	-	-	-
A Novel"Knowledge"Base Question"Answering Model"Based"on"Knowledge"Representation and"Recurrent"Convolutional"Neural Network" [6]	-	-	-	-	-	-	-	✓	-	-	-	-
Dynamic"Updating"of the Knowledge"Base for a Large-Scale"Question Answering"System" [7]	✓	-	-	✓	-	-	-	-	-	-	✓	-
Stepwise"Reasoning"for"Multi-Relation" Question"Answering over"Knowledge-Graph with"Weak"Supervision" [8]	-	-	✓	✓	-	-	-	-	-	-	-	-



<p><i>Reading a Comprehension Model for a Question Answering [20]</i></p> <p><i>Research and a Application of a Intelligent Question-Answer Algorithm a Based on a Multi-Channel a Dilated Convolution [21]</i></p> <p><i>Research on a Short Text a Similarity Calculation</i></p> <p><i>Method a for a Power a Intelligent a Question Answering [22]</i></p> <p><i>A Factoid a based Question a Answering a System based a on a Dependency a Analysis a and Wikidata [23]</i></p> <p><i>An Intelligent a Question a Answering a System based on Power a Knowledge a Graph a [24]</i></p> <p><i>A BERT- Based a Semantic a Matching a Ranker for Open-Domain Question a Answering [25]</i></p> <p><i>A a Multi-Level Semantic a Fusion Algorithm Based a on Historical a Data in a Question Answering [26]</i></p> <p><i>Requirements a for a Automating a Moderation in Community a Question- Answering a Websites [27]</i></p> <p><i>iLFQA: a Aa Platform a for Efficient a and Accurate Long- Form a Question a Answering [28]</i></p> <p><i>Toward a Zero-Shot a and a Zero- Resource Multilingual Question a Answering [29]</i></p> <p><i>CypherQA: a Question- answering a method based on Attribute a Knowledge a Graph [30]</i></p> <p><i>Deep learning a based Answering a Questions</i></p>	<p>- - - - - ✓ - - - - -</p> <p>✓ - - - - ✓ - - - - -</p> <p>- - - - - ✓ ✓ - - - -</p> <p>✓ - - - - - - - - - ✓</p> <p>- - - - - ✓ - - - - -</p> <p>✓ - - - - - - - - - -</p> <p>- - - - - ✓ - - - - -</p> <p>✓ - - - - - - - - - -</p> <p>- - - - - ✓ - - - - -</p> <p>✓ - - - - - - - - - -</p> <p>- - - - - ✓ - - - - -</p> <p>✓ - - - - - - - - - -</p> <p>- - - - - ✓ - - - - -</p> <p>✓ - - - - - - - - - -</p>
--	---

<i>using a T5a and a Structured a Question Generation a System [31]</i>	-	-	-	-	-	-	-
<i>Enhancing a Query Answer a Completeness with Query a Expansion a based on a Synonym Predicates [32]</i>	-	-	-	-	-	-	-
<i>Text-Enhanced a Question a Answering a over Knowledge a Graph [33]</i>	✓	✓	-	-	-	-	-
<i>A Legal a Question Answering a System a Based on BERT&amp;#xa0;[34]</i>	-	-	-	-	-	✓	-
<i>Gentle a Teaching a Assistant a Model a For</i>	-	-	-	-	-	✓	-
<i>Improved a Accuracy a In a Question-Answering a [35]</i>	-	-	-	-	-	-	-
<i>Efficient SPARQL a Queries Generator a for</i>	-	-	-	✓	-	-	✓
<i>Question a Answering a Systems [36]</i>	-	-	-	-	-	-	-
<i>QATest: A Uniform a Fuzzing Framework a for</i>	-	-	-	-	✓	-	-
<i>Question a Answering Systems [37]</i>	-	-	-	-	-	-	-
<i>Memory-Aware a Attentive a Control a for</i>	-	-	-	-	-	-	-
<i>Community a Question a Answering a With Knowledge-Based</i>	-	-	-	-	-	✓	-
<i>Dual a Refinement a [38]</i>	-	-	-	-	-	-	-

### 2.6.5. Technique-Based Filtering:

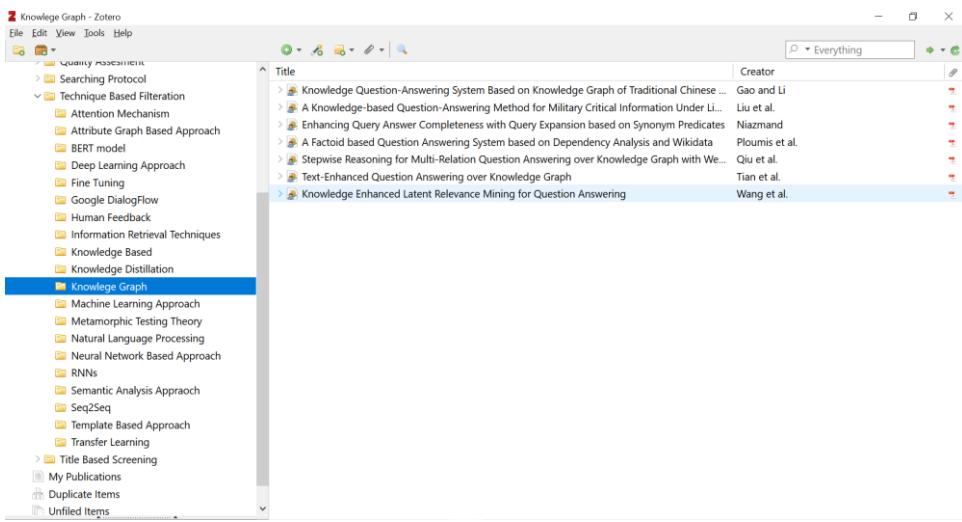
Technique-based filtering is a way to select research studies based on the methods or techniques used in the research. For example, if a research question requires a specific type of survey or statistical analysis, technique-based filtering can help to identify relevant studies. The categories of techniques were as shown in Table 10.

**Table 10.** Notation Table of Techniques terms.

Notation	Abbreviation	Notation	Abbreviation
<i>Attention Mechanism</i>	AM	<i>Knowledge Graph</i>	KG
<i>Attribute Graph Base Approach</i>	AG	<i>Machine Learning Approach</i>	ML
<i>BERT model</i>	BERT	<i>Metamorphic Testing Theory</i>	MT
<i>Deep Learning Approach</i>	DL	<i>Natural Processing Languages</i>	NLP
<i>Fine-Tuning</i>	FT	<i>Neural Network Approach</i>	NN
<i>Google Dialog-Flow</i>	GD	<i>Recurrent Neural Network</i>	RNN
<i>Human Feedback</i>	HF	<i>Semantic Analysis Approach</i>	SA

Information Retrieval Techniques	IR	Sequence-to-Sequence	Seq2Seq
Knowledge Base	KB	Template Based Approach	TB
Knowledge Distillation	KD	Transfer Learning	TL

Technique-based filtering conducted in Zotero is diagrammatically represented in Figure 9.



**Figure 9.** Technique Based filtering folder in Zotero.

**Table 10.** illustrates this technique-based screening by identifying and grouping the research publications' techniques into clusters.

**Table 10.** Technique Based Filtering.

Re f.	TECHNIQUES																			
	A	A	BER	D	F	G	H	I	K	K	K	M	M	NL	N	RN	S	S2	T	T
M	G	T	L	T	D	F	R	B	D	G	L	T	P	N	N	A	S	B	L	
[1]	✓	-	-	✓	-	-	-	-	-	-	-	-	-	-	✓	-	✓	-	-	
[2]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	
[3]	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	
[4]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	
[5]	-	-	-	-	-	-	✓	-	-	-	✓	-	-	-	-	-	-	-	-	
[6]	-	-	-	✓	-	-	✓	-	-	-	-	-	-	-	✓	-	-	-	-	
[7]	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	
[8]	✓	-	-	-	-	-	-	-	-	✓	-	-	-	-	✓	-	-	-	-	
[9]	✓	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	
[10]	-	-	-	-	-	-	✓	-	-	-	-	-	-	✓	-	-	-	-	-	
[11]	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	
[12]	-	-	-	-	-	✓	-	-	-	✓	-	-	-	-	-	-	-	-	✓	



### 3. Datasets

Datasets are repositories of organized information used for a variety of tasks, such as preparing question-and-answer training programs. Datasets like these are frequently used in question-answering algorithms:

- ***The Stanford Question Answering Dataset (SQuAD)*** is a database of questions and responses derived from Wikipedia articles.
- ***MS MARCO (Microsoft MAchine Reading COmprehension)***: Contains sections from web documents and genuine user inquiries from Bing search.
- ***CoQA (Conversational Question Answering)***: Focusing on conversational question-answering utilizing short tales.
- ***QuAC (Question Answering in Context)*** is a tool for thinking through several phrases while answering questions in conversation.
- ***TriviaQA***: Presents complex questions from various groups to test question-answering algorithms.
- ***Natural Questions***: Queries generated from actual user inquiries from the internet with responses given in sections from Wikipedia.
- ***WikiQA***: WikiQA is a dataset that emphasizes the choice of answer sentences. It includes questions derived from web searches and offers a list of potential answers for each query, leaving it up to models to choose the most pertinent one.

These datasets offer tools that allow scholars and researchers for developing and evaluating question-answering systems in certain situations since they span a variety of topics and question kinds.

#### 4. Detailed Literature:

A comprehensive and systematic strategy to finding and synthesizing all pertinent papers on a certain topic is part of a detailed literature review in a SLR. The major purpose is to present an in-depth overview of the available data on the research topic or target. It takes a high degree of technique as well as focus in order to conduct an extensive literature evaluation as part of a systematic review. The following is a detailed description of all the research that has been done on the topic.

“Deep learning” is the method used in [1], especially a collection of deep learning models for natural language retrieval and synthesis that include recurrent neural networks, sequence-to-sequence models, and attention mechanisms. For a website that sells goods electronically, the models are used to create an answering queries chat-bot that assists users with goods research as well as

inquiries. By feeding the outcomes of this initial model into the follow-up one, a linguistic correction mechanism is also put into place, which increases the accuracy of the entire system.

A “multi-view pointer-generator network” describes the method that is used [2] to generate answers that are knowledgeable to the opinions of users. It comprises utilizing a single model to concurrently train response production and opinion extraction tasks, distilling and collecting key opinion data via two different opinion merging algorithms, and producing opinion-aware replies through a multi-view pointer-generator network.

“Knowledge graph-based question-answering system” is the methodology employed in [3]. The knowledge graph is used by the system in order to provide responses to user inquiries. It pulls triplicates of TCM information from records, saves them in a Neo4j graph database, then utilizes the graph as a storage location for the triplets. In addition, the system uses natural language processing methods including object recognition, TF-IDF, and word vector matching to comprehend queries from users and correlate it to the most suited question-answering patterns. In accordance with the semantics of the patterns and the objects included in the queries, the answers are then looked for in the knowledge graph and given back to the user.

A “neural network-based approach” was utilized in [4], notably the DeepAns model, which has three stages: question boosting, labelling creation, and response suggestion. The most suitable response may be determined and suggested to the user by using the neural network to generate matching scores among the query and the response options.

Regarding fine-grained identification of objects, the [5] technique is described as a “batch-based local question answering method” that combines model responses from humans. It is not mentioned which specific method has been employed, but it incorporates a combination of human-computer recognizing approach via sole question response with a person in the looping process, as well as the plug-in of computer vision algorithms when photographs are provided. To increase the accuracy of recognition, it therefore probably uses a mix of machine learning and human input.

To create a KB-QA framework, the method described in [6] combines “knowledge representation and recurrent neural network (RNN)”. The model is composed of three sections: the production of potential replies, the mining of object connections, and the acquisition of knowledge’s representation from a knowledge base. Furthermore, a formula is created to calculate the rankings of applicant replies that relate to the knowledge base. The model uses deep learning approaches for extracting features and representational learning in order to increase the accuracy of the responses through the use of knowledge base structural data.

The approach implemented in [7] is referred to be a “vector modelling technique to facilitate automatic learning of KB-QA systems,” and it tackles the problems with scaling and knowledge base update. The suggested ALKB-QA system offers fundamental patterns and periodically updates the KBs, increasing accuracy. It was initially introduced as a practical commercial application and outperformed methods based on manual annotation.

A “neural method based on reinforcement learning called Stepwise Reasoning Network” is the mechanism employed in [8]. It establishes multi-relation query answering as an example of progressive choice issue, conducts efficient pathway exploration across the knowledge graph, and takes advantage of the attention mechanism and neural networks to improve the distinctive influence of various portions of a specific query over triple choice. To tackle the postponed and insufficient give issue brought on by inadequate supervision, the approach also suggests a potential-based reward shaping technique. The results of the trials demonstrate that the suggested approach outperformed the cutting-edge methods.

In order to incorporate ordinary knowledge from external knowledge graphs (KGs) into an appropriate selector for response decision-making, the study [9] presents an approach referred to as “knowledge-aware context-based attention mechanism.” For the purpose of capturing additional interacting aspects between queries and responses, the approach employs a “compare-aggregate” architecture. It is a neural technique that improves choice of responses by employing external knowledge graphs and attention processes.

A couple of methods are employed in the research study [10]: “Crawler technology” to crawl related disease queries and organized disease knowledge bases on disease query and response sites and disease websites; recommending a “joint model” for recognizing both entities and characteristics in the question being asked, accordingly, and utilizing a “joint model” to determine the groups and characteristics in the question itself.

To enhance the accuracy of a knowledge-based question answering system, the research [11] presents an approach called PERQ, which consists of three steps: prediction, explanation, and correction. The research presents a broad framework for enhancing KB-QA accuracy rather than a specific approach employed inside each phase. The paper’s proposed PERQ framework consists of three steps: The first prediction forecasts whether a particular question can be successfully answered by a KB-QA system. The second step is Explanation, which analyses the causes behind the KB-QA system’s failure to answer the question, and the last step is Rectification, which corrects the answer using the prediction and explanation findings. The paper suggests and evaluates tools for completing these three phases.

The “Two-stage Multi-teacher Knowledge Distillation method for web Question Answering” is presented in the paper [12] and includes preliminary training the learner’s model via a general Q&A distillation task and fine-tuning with multi-teacher knowledge distillation on downstream tasks. The technique intends to minimize over fitting and transfer more general information to the learner’s model.

Dimensionality reduction for embedding representations using an autoencoder is the method utilized in [13]. By using an autoencoder for acquiring low-dimensional features of input embeddings, the substantial computational expenses and resource demands associated with embedding-based approaches for question answering are intended to be reduced. The linguistic match that exists between the supplied response and the natural language inquiry is then determined using the reduced dimension embeddings. This technique can reach performance that is equivalent to that of standard baselines despite being longer-lasting and memory efficient, according to testing and analyses on the insuranceQA benchmark.

The method in [14] is a “combination of transfer learning and knowledge distillation”, which utilizes a teacher-student paradigm for cross-domain transfer learning in Question Answering (QA) systems, with the goal of leveraging knowledge from different fields and the “dark knowledge” from the vast teacher model to enhance the student model’s performance within the domain of interest.

The “deep learning” method employed in [15] was created expressly for the purpose of creating a question-answering system which is capable of comprehending and reading documents written in natural language. With no previous knowledge of linguistic structure, the framework transforms words and paragraphs into an internal representation to produce precise responses to queries that are presented in natural language. The system’s capacity to respond to hypothetical queries is assessed using sizable datasets for training and testing.

The methods employed in the research [16] are “template-based approach, semantic analysis approach, and deep learning approach”. This research additionally offers an in-depth evaluation of these three algorithms and recommends further directions for deep learning models, built-in traits of language, and the fundamental idea of integrating questions and answers when creating query sentences.

A “light-weight context-aware QA framework called Context-QA” is the method utilized in the study [17]. It features a ground-breaking algorithm called the Staged QA Controller that aims to confine the hunt for solutions within the parameters that are most closely related to the inquiry. In comparison to traditional QA approaches, the system is made to offer QA experiences on multimedia material, enhance the level of quality of the responses by up to 49%, and consume up to 56% less time. Ninety percent of those who took part in the subjective assessments used to evaluate the strategy said they enjoyed using this new media format.

The method used in research [18] is a “Question Answering Model (QAM)” that was created using the BERT model and Google Dialogflow to imitate human-like relationships in a virtual assistant for comprehension activities..

A “knowledge-based question-answering (QA) technique for handling militarily critical information” is the approach employed in [19]. The procedure entails creating a collection of military crucial data questions, classifying the queries in accordance with the question categories using a standard matching approach centered on the naive Bayes classification algorithm, followed by accessing the crucial data knowledge graph to attain quality assurance. The approach employs methods including semantic information mining, match between templates, and knowledge graph queries to accomplish adaptive quality assurance for military-critical materials within the situation of restricted reservoir.

“An end-to-end multigranularity reading comprehension model for extractive question answering” is the phrase used to characterize the method stated in [20]. In a single scheme, this approach explicitly demonstrates three comparing granularities: paragraph identification, sentence preference, and response generation. By combining characteristics into a single framework, this method enhances the acquisition of depictions of various matching granularities. This model benefits from the fact that it reduces the issue of propagation of mistakes in the initial training and interpretation processes and beats both vanilla BERT models and current multistage matching techniques in a thorough comparison on four massive datasets. To confirm the efficiency of the suggested elements in the constructed structure, a procedure called ablation research is additionally carried out.

“A text matching model based on multi-channel dilated inversion” is the method employed in [21]. The approach is employed by electricity service providers to compare fresh user inquiries to previous conversation data. The system can instantly resolve user issues because to the model’s ability to capture text data at various scales to assess sentence similarity based on deeper semantic elements. Testing of the suggested model revealed good performance.

The method utilized in [22] is “a question-answering system based on a knowledge base” in the discipline of electricity, which employs a model for algorithms of cosine similarity mixed with TF-IDF to successfully match phrase similarity and segment vocabulary. It has been evaluated and shown to reach an accuracy of 75.8%, which is much higher than previous models. This approach can help save labor expenses and help energy staff members to resolve issues they meet while at work.

“An efficient methodology for automated question answering using knowledge bases” is the term utilized in [23]. In order to get the response provided by the Wikidata Query Service API, the platform analyses the inquiry, identifies connections, builds major triplets, and then prepares a SPARQL query. In addressing queries using open-domain factoids, the technique performs admirably.

“A domain knowledge graph with graph database and computing technologies for an IQA system” is the method utilized in [24] in the electric power industry. It creates graph data question phrases, conducts accurate information search and evaluation for clear visualization, and utilizes NLP to find purpose and restrictions. The technique provides multi-hop knowledge correlation reasoning analysis at rapid rate.

“A BERT-based semantic similarity ranker” is the method implemented in the study [25] to determine the semantic similarity within a sentence and a query for open-domain query answering. Examining three distinct representation aggregated methods, research on two open datasets reveal considerable gains in rankings and QA performance.

The method utilized in the research [26] is “a multi-level semantic fusion neural network model” that integrates deep learning and natural language comprehension to enhance conceptual comprehension in question answering. On the SQuAD dataset, the framework employs attentions and cross-layer fusions to enhance results. With regard to precise match rate and F1 value, the suggested approach operates better than conventional models.

The research [27] suggests “a machine-learning approach to automate content quality detection in community Q&A websites”, tackling the difficulties of sustaining standards in a setting of high user and contribution quantities. With a focus on context-related aspects and transfer learning for adaptability to changing material, the suggested approach contains modules for extraction of features, enhancement of features, and classification model training. The approach places emphasis on the requirement for concise and convincing justifications for regulation choices.

The method described in the study [28] uses a framework called “a machine learning-based long-form question-answering (iLFQA)” which allows the utilization of components for textual retrieval, textual synthesis, and zero-shot classification to provide replies to unstructured queries depending on an open-domain knowledge repository.

Leveraging fine-tuned multilingual bidirectional encoder representations via transformers (mBERT), the research [29] provides “a framework for zero-shot multilingual question answering in low-resource domains”. The model employs information obtained from high-resource languages to enhance low-resource languages’ functionality with zero-shot and zero-resource overhead. The suggested method executes better than a number of standard methods that call for millions of unprocessed data points for low-resource languages.

According to the study [30], “a knowledge-based query answering method that utilizes an attribute graph and a two-layer network with structured attention mechanism to optimize object boundaries recognition” is proposed. In order to identify associations and objects, it pulls references from the text, converts them into a slot-filling Cypher phrase, and uses this phrase for querying the response. The suggested method performs admirably when resolving complicated queries with murky object boundaries and challenging path prediction.

In accordance with the study [31], “a customized query generator methodology that involves fine-tuning a transformer and concentrating on both multiple-choice and long answer varieties” should be used. The evaluating model locates pertinent query-answer configurations, while the generator version creates queries from chosen words and develops semantically driven interruptions for MCQ answers.

“A method that improves question answering over inadequate community-based knowledge graphs” is what the study [32] suggests. It proposes a method to handle queries effectively in order to assess how comprehensive the answers are and improve the output. With embeddings created on a knowledge graph to identify synonym predicates, the methodology suggests a unique question extension technique. The strategy enhances question response comprehensiveness, according to the initial research, although additional work is needed for query expansion to maximize response completion.

In order to improve the visualization of knowledge, the study [33] suggests “a textual-enhanced query answering strategy over knowledge graph” that makes use of detailed contextual data in a text corpora. The framework employs a Transformer Encoder network to collect input interpretations and an attention mechanism to interactively merge internal as well as external data. This strategy exceeds similar cutting-edge QA techniques when it comes to of accuracy, according to findings from experiments on the WebQuestions dataset.

The study [34] suggests utilizing the Milvus vector query engine and BERT-based problematic vector format “a strategy to enhance cognitive query-answering systems” in certain applications. When the threshold level is set at 0.2, the recommended approach has a rate of retrieval of 86% and a discrepancy rate of 84%.

In the article [35], a method known as “an efficient knowledge distillation approach” was applied. To make it more precise, an individual’s model is taught more effectively and with fewer constraints by using a more advanced version as “an instructor assistant”.

In order to lower the request cost of creating SPARQL for query-answering systems, the study [36] presents “a low-cost SPARQL generator named Light-QAWizard” that incorporates multi-label categorization into a network of recurrent neural networks. All of these alternatives are outperformed by Light-QAWizard, which also has a request cost that is almost half that of QAWizard.

According to the study [37], “a fuzzing framework called QATest based on metamorphic testing theory which produces tests using arbitrary data autonomously for various QA systems” is proposed. The generating procedure is directed by N-Gram coverage and perplexity priority. The model is tested against four QA systems, and the findings demonstrate that it effectively identifies numerous incorrect QA system behaviors.

According to the study [38], “internally knowledge-based end-to-end model for open-domain query answering” that takes interpretation and context into account using responsive storage networks and persistent memory. Comparing how it performs to cutting-edge methods, it is fiercely competitive.

## 5. Critical Analysis:

In research, critical analysis entails analyzing a study’s methods, results, and prospects for advancement. It is beneficial to note advantages and disadvantages, rate the quality of the testimony, and take moral considerations into account. To improve the breadth and fineness of the conversation, I will conduct a critical review of the foundational methods described in Table 10 in this part.

A significant feature of query-answering systems nowadays is attention mechanism. These strategies enable the technology to concentrate on particular portions of the incoming data deemed crucial to answering the query. The attention mechanism is applicable to several fields, notably computer vision and natural language processing. The attention mechanism often pays attention to terms that are close to the right response, exist in the inquiry, and have the greatest pertinent to the situation [39]. The resemblance score function, which determines the attention assessment of similarity among two textual parts, is a crucial component of the attention mechanism [40]. Systems for answering questions have found application for the attention mechanism. There are several restrictions on this approach, too. The typical drawbacks of attention mechanisms in query-answering systems encompass computational aspects and memory demands that are exponential to the total number of words of the input sequence [41], the attention mechanism simply examining the encoded query matrix along with one candidate answer word embedded therein at a time, which might not be adequate for capturing the entirety of the context of the query and answer [42], and the attention mechanism might not operate effectively whenever the input data is noisy or consists of irrelevancies [39]. A further limitation is the reality that the attention mechanism only takes into account how the question will affect the response and overlooks how the answer will affect the query [43]. Moreover, for complicated inquiries, shallow attention processes may not be able to choose the pertinent information from the simultaneous representation of the question and image [42]. Considering these drawbacks, attention mechanisms are nevertheless an effective tool for question-answering systems and have been effectively used in a number of disciplines [44].

A graph-based technique utilized in question-answering systems is called the Attribute Graph Base technique. With this method, a graph of objects and their characteristics is created, with the points denoting the entities in question and the edges denoting their features. The method then employs this graph to provide answers to queries by navigating it and looking for the required response depending on the query entities [45]. This method has the drawback of requiring a lot of data to build the graph, which can be time-consuming and costly computationally [46]. The method may not perform well when the input data is noisy or contains unnecessary information [47], which is still another restriction. Furthermore, because the network might grow too big to be processed effectively, the method might not scale to enormous datasets [48]. Although it has these drawbacks, the Attribute Graph Base Approach is nonetheless a valuable tool for question-answering systems and has been effectively used in a number of disciplines. [49].

In 2018 [50], Google AI Language created the BERT (Bidirectional Encoder Representations from Transformers) machine learning model for understanding natural language. It is a very sophisticated language framework that may be applied to a variety of NLP tasks, include answering questions [50]. A number of applications, including COVID-19-related inquiries [51], online commerce [52], and healthcare [53], have deployed BERT-based question-answering models. In a number of NLP tasks,

BERT has demonstrated state-of-the-art performance [50]. Nevertheless, the quality and amount of data used for training together with the particular domain of the questions being asked determine how effectively BERT performs in question answering systems [52,53].

The capacity of deep learning algorithms to represent complicated issues has led to their widespread utilization in query answering systems. In two distinct streams of data extraction technique and end-to-end neural network-based approaches, deep learning techniques have demonstrated state-of-the-art performance in answering straightforward problems [54]. Modern designs of MQA have been studied, and deep learning techniques were additionally utilized in these systems [55]. Nevertheless, the quality and amount of the training data, the predictive architecture used, and the optimization of the model parameters all affect how well deep learning algorithms perform in question-answering systems [56,57]. Although deep learning techniques have demonstrated significant promise in question-answering algorithms, they additionally come with major drawbacks. The need for a lot of labelled data to train the algorithms is one of the primary restrictions. This can be difficult in fields where getting annotated data is difficult or expensive [58]. A further drawback is the inability of deep learning models to be interpreted, resulting in it challenging to comprehend how a model generates its results. This might be a concern in operations where transparency and responsibility are crucial, such as the legal or medical fields [57]. Models based on deep learning are also typically computationally costly and demand powerful hardware for deployment, which can be a hurdle to adoption in situations with limited resources [59]. Ultimately, overfitting is a risk factor for deep learning models, that can result in subpar generalization performance on new data [57]. While using deep learning models in applications that are practical, it is crucial to thoroughly assess the effectiveness of these models in question-answering systems while taking into account their limits.

Deep learning often uses fine-tuning in QA systems to achieve great generalization outcomes on subsequent tasks with comparatively minimal training data [60]. It is common practice in NLP to fine-tune a transformer model for query answering, which may end up in a high-performing QA model [61]. It enables the knowledge transfer from already-trained models to particular tasks, which can enhance the functionality of the QA system [61]. By changing the model's weights to more accurately suit the particular job, fine-tuning can also aid in enhancing the system's accuracy [62]. By changing the model's weights to better suit the particular job, fine-tuning can also aid in enhancing the system's accuracy [63]. A further drawback is that only small modifications to the language model, primarily to its top layers, may be made during fine-tuning [63]. Furthermore, if the inquiry contains grammar errors, the machine might not locate a match [63]. Nevertheless, optimizing a transformer model for QA is a well-liked NLP technique and can result in a high-performing QA model [61,64]. As a whole, fine-tuning in QA systems offers benefits and drawbacks that must be taken into account when designing a QA system.

Natural language processing (NLP) is a cognitive AI technology used by Google Dialogflow to build chatbots and voice assistants [65,66]. The developers are now able to build robots capable of decipher requests from users, correlate them to intentions, and retrieve objects from them [67]. The query outcomes do not specifically refer to Dialogflow's shortcomings in a QA system. It is crucial to keep in mind, nevertheless, that Dialogflow is a conversational artificial intelligence (AI) system built on natural language processing (NLP) that can be employed for creating user interfaces that are conversational for websites, mobile applications, and other business-related topics [68]. Although Dialogflow is a strong tool for building conversations with agents, it's crucial to keep in mind the drawbacks of NLP-based systems, such as the possibility for user questions to be misunderstood and the requirement for continual upkeep and upgrades to assure correctness.

The use of human feedback in quality assurance systems is crucial. It aids in enhancing the system's dependability and fostering user confidence [69,70]. In QA mechanisms, human feedback has certain constraints. It might be challenging to get adequate input that is of high quality to enhance the system, which is one of the major challenges [71,72]. Furthermore, human feedback may be biased, which might reduce the system's accuracy [71]. A further constraint on the deployment of

conversational platforms is the expense involved in gathering and analyzing human feedback [72]. Regarding these drawbacks, some research has looked at integrating collaborative input to raise the precision and comprehensibility of QA systems [73]. Discrepancies in human feedback have already been used in other experiments to increase the accuracy of the system [74]. In summary, even if receiving human feedback might be useful, it's critical to solve the issues that arise in order to make sure that the QA processes are reliable and successful.

The use of information retrieval methods is crucial in question-and-answer (QA) systems. Information retrieval methods used by contemporary quality assurance systems range from conventional sparse vector keyword matching to deep learning models and neural networks [75]. Among the key approaches for quality assurance is based on IR methods, where the content of a user inquiry is assessed to derive a pertinent response [76]. Information retrieval methods have some drawbacks even though they are essential for QA systems. The fact that the performance of QA models is constrained by the performance of the IR system [77] is one significant restriction. Another drawback is that particular QA datasets, like SQuAD, might hinder the model's performance by giving it access to a small amount of data [77]. Furthermore, for more complicated queries that call for a better comprehension of its proper context, conventional sparse vector word matching methods, like Elasticsearch, might not be adequate [75]. Additionally, even though deep learning models and neural networks are used by current QA systems for information retrieval, these methods can be computationally costly and demand a lot of resources [75]. Information retrieval methods are crucial for QA systems, but they also have inherent limits that must be taken into account while creating and assessing these systems.

The implementation of knowledge-based strategies is crucial in question-answering (QA) systems. These methods rely on a machine for inference along with a knowledge base to draw conclusions from the data stored in the knowledge base [78]. The way that systems based on knowledge handle difficulties may differ since some systems encode professional expertise as rules, while others employ a case-based reasoning [78]. To accurately respond to customer inquiries, QA systems combine natural language processing, information retrieval, logical reasoning, knowledge representation, and machine learning [79]. Although knowledge-based approaches are a crucial part of QA systems, they do have significant drawbacks. The absence of techniques for verifying and validating knowledge-based systems is one of their most important limitations [80]. Protection, quality assurance, and dependability difficulties may result from this [81]. A further drawback is the amount of effort and money required for the creation and upkeep of knowledge-based systems [82]. Additionally, despite the fact that knowledge-based systems might be efficient for some sorts of issues, they might not be appropriate for all topics or domains [83]. Though knowledge-based procedures are a crucial component of QA systems, it is necessary to keep in mind their limits when designing and implementing these systems.

The condensation of knowledge from a complicated model into a more straightforward one is known as knowledge distillation [84]. The less complex model can attain equivalent accuracy to the bigger model thanks to a training strategy that distributes information from a larger, more complicated model to a lighter, simpler one. For the purpose of enhancing the performance of QA systems, and knowledge distillation has been deployed. For instance, retrieval-based QA systems' accuracy has been increased by the distillation of cross-domain knowledge [86]. In a different research, knowledge was transferred from an ensemble of models to a single model using knowledge distillation and active learning, which enhanced question-answering performance [87]. Although knowledge distillation is a useful method for enhancing the functionality of QA systems, there are several restrictions that should be taken into account when applying this method. One drawback is that not every models or activities can benefit from knowledge distillation [88]. Furthermore, zero-shot transfer learning, which aims to conveying information from a particular field to a different one without any previous training information in the desired field, could prove to be useful for knowledge distillation [89]. Knowledge distillation could not be successful for all types of data, especially when the data is noisy or contains errors [84]. This is another drawback. Additionally, not

all designs can benefit from knowledge distillation, especially those that have become extremely complicated or specialized [87]. Last but not least, not every applications, especially those requiring high degrees of precision or accuracy, may benefit from knowledge distillation [86]. As a whole, knowledge distillation, which involves moving information from more complicated, larger models to less complicated, smaller ones, is a useful strategy for QA system performance improvement.

The adoption of knowledge graph approaches in question-answering (QA) systems is crucial. Knowledge graphs offer well-organized relational data between things, that can be useful for responding to inquiries in plain language [90]. By combining knowledge graphs into variational reasoning, QA systems' performance has significantly increased [91]. Systems for autonomously responding to inquiries about knowledge graphs in natural language (KGQA) are being developed [92]. Although they do have certain limitations, knowledge graph approaches are a vital part of QA systems and are crucial for giving correct and trustworthy replies to user inquiries. A potential drawback is that knowledge graphs might be faulty or insufficient, which could result in mistakes in the solutions offered by the QA system [92]. It can also be difficult to translate natural language inquiries into formal query representations that can be utilized to retrieve responses from a knowledge graph [93]. The simple fact that many queries need multi-hop logic reasoning across the knowledge network in order to get the solutions is another limitation [91]. Additionally, a lot of user inquiries might include errors or different pronunciations, which could make it challenging for the QA system to correlate the given items to the knowledge graph [91].

Systems for answering questions (QA) have been created using machine learning approaches. The aforementioned systems may be trained to respond to queries with or without context depending on the situation. Systems for quality assurance (QA) that use deep learning have demonstrated promising results and can lessen the requirement for manually created heuristic rules [94]. Machine learning possesses the ability to enhance QA systems' precision and effectiveness, but it is crucial to carefully assess their capabilities and limits in each unique situation. Being utilized in question-answering (QA) systems, machine learning approaches have several drawbacks. The requirement for substantial training data and processing means is one of the major issues [95]. The caliber of the training data and the difficulty of the queries can also have an impact on how accurate these systems are [96]. It's likewise crucial to keep in mind that machine learning-based QA systems could not be appropriate for all sorts of questions and might have trouble with inquiries that call for contextual knowledge or common sense reasoning [97].

A potential method for assessing question-and-answer (QA) systems from the users' viewpoints is metamorphic testing. Finding metamorphic linkages and deciding which of them to test are its two key shortcomings. Finding metamorphic relationships may be difficult, particularly for intricate systems like QA systems. Furthermore, choosing which metamorphic connections to test might be challenging because there may be a large number of potential links [98]. A further drawback of metamorphic testing involves the fact that not all QA systems could possibly be compatible with it, as certain systems might lack obvious metamorphic relations that can be tested [99]. In conclusion, although metamorphic testing has the potential to enhance QA systems, it is crucial to carefully assess its use and limits in each unique situation.

Many question-answering (QA) systems incorporate utilization of Natural Language Processing (NLP) methods. When answering queries, these systems often take an invasive strategy that consists of an interpreter and a retriever [100]. QA systems may employ a variety of architectural designs, although they often use predictive indexing techniques. Analyzing natural language subtleties, particularly necessitates advanced NLP tools, is one of the key issues in QA [102]. In QA systems, NLP approaches are additionally utilized for expressing knowledge, response display, and social networking analysis [103]. Natural Language Processing (NLP) approaches have several limitations in Question Answering (QA) systems. The intricacy of feature design is one of the key drawbacks, however deep learning techniques can get around this [104]. The accurateness of the system may also be hampered by the level of detail of the supplied data [105]. For NLP-based QA systems, retaining contextual throughout duration and comprehending human feelings as well as phrases with diverse

meanings are equally difficult tasks [105]. The usefulness of NLP approaches, query mapping, and response inferencing has to be examined [106] since semantic-based QA systems may also encounter issues. Some applications may be constrained by the query restrictions imposed by closed domain systems [107]. Overall, while NLP techniques are critical for QA systems, they have some limitations that need to be addressed for optimal performance.

Although neural network-based approaches have been extensively used in Question Answering (QA) systems, they do have significant drawbacks. The intricacy of obtaining features is one of the key drawbacks, however deep learning techniques can get over this [108]. Massive amounts of labelled data are a further constraint that certain applications may find difficult to meet [109]. Another obstacle is the challenge of comprehending phrases with numerous interpretations, human emotions, and context across time [110]. It can be difficult for neural network-based QA systems to manage unorganized data and keep context over time [110]. Likewise these systems' productivity is still beneath what people would anticipate, and unresolved QA systems' responses are frequently general and call for additional investigation [110]. Deep learning techniques, such as determining feature intricacy, are capable of getting around some of these constraints [108]. Although neural network-based approaches have helped QA systems operate more effectively in general, there are still several issues that must be resolved for them to operate at their best.

Recurrent Neural Networks (RNNs) have recently been demonstrated to be successful in Question Answering (QA) systems. Input patterns of varying length may be handled by RNNs, resulting in such an excellent choice for applications like statistical analysis of time series, speech recognition, and natural language processing [111]. A quality assurance system answers queries given by human users in natural language [112]. Nevertheless, long-term dependencies in the order of inputs might be difficult for RNNs to record, which can lead to the network forgetting critical information [111]. In QA systems, RNNs have several drawbacks. The fact that RNN functions cannot be parallelized because of their linear structure, making both training as well as inference sluggish, is one of RNNs' constraints in Question Answering (QA) systems [113]. a consequence of the recurrent hidden vector's dependency on the prior concealed vector, RNNs experience memory loss [114]. RNNs face a few drawbacks in QA systems, including delayed inference and training, a lack of powerful representations, a challenge in preserving dependencies over time, and memory loss.

Semantic analysis is a technique utilized in inquiry Answering (QA) systems to accurately analyze the user's inquiry and offer a response. Semantic analysis is used to determine a question's interpretation and correlate it with the pertinent facts in the information provided source [115]. The particular focus of existing Question Answering (QA) systems, flaws in the frameworks that have been utilized, and the difficult task of accurately labelling the content with the right meaning behind it represent a few of the limits of semantic analysis in QA systems [116].

A approach called Seq2Seq has been adopted for use in Question Answering (QA) systems in particular to create effective conversational modelling for sequences [117]. In order to construct queries based on the context and likely responses discovered, Seq2Seq models were used [118]. Since 2014[119], Seq2Seq models have dominated Natural Language Processing (NLP) activities like QA systems. The question answering assignment specified in the Stanford Question Answering Dataset (SQuAD) was completed using a sequence-to-sequence attention comprehension of reading model [120]. Still, a number of the drawbacks of Seq2Seq models in QA systems include the narrow emphasis of existing QA systems, flaws in the models that are employed, and the challenge of correctly semantically labelling the text [121].

A template-based method is a mechanism used in Question Answering (QA) systems to create all feasible queries from a particular set of data utilizing a variety of patterns [122]. Template-based QA systems retrieve the question's query format using custom rules [123]. For Conceptual Knowledge Base Question Answering (C-KBQA) systems, an inherited responsive template-based strategy has been suggested for responding to complicated inquiries using a semantic template-matching approach [124]. The benefit of template-based QA solutions is that they are simple to utilize and comprehend, and they may be employed to respond to factual queries [125]. Yet, a couple of the

drawbacks of template-based QA systems pertain to the narrow emphasis of present QA systems, flaws in the models that are employed, and the challenge of resolving complicated queries that call for inference and reasoning [126].

Transfer learning is a technique used in Question Answering (QA) systems to pass on data obtained from a single role to another, with the goal of improving the system's performance [127]. Transfer learning's application to QA was recently investigated [128]. It has been demonstrated to be effective for activities like object and voice recognition. Numerous previously trained models were utilized as the initial teaching emulate in one research, and their experience was then transmitted to a number of subordinate QA models, improving the QA performance [129]. A further investigation employed supervised transfer learning to improve the accuracy of the item details, and it saw an improvement of roughly 10%[130]. By transferring the information acquired from a single endeavor to a different one, transfer learning may be utilized to enhance the performance of QA systems. Nonetheless, the necessity for substantial volumes of information as well as the challenge in choosing the best already trained algorithm for the intended task are some of the drawbacks of transfer learning in QA systems [128].

**Table 10.** Summary of Critical Analysis.

REF. PAPERS	TECHNIQUES	SHORTCOMINGS
[1,8,9,30]	<i>Attention Mechanism (AM)</i>	<i>Quadratic Computational And Memory Requirements, Limited Context Capture, Sensitivity To Noise And Irrelevance, Neglecting Answer Impact On Question, Inadequate Selection From Complex Inputs.</i>
[30]	<i>Attribute Graph Base Approach (AG)</i>	<i>Data-intensive graph construction, sensitivity to noise and irrelevance, potential scalability issues with large datasets.</i>
[18,20,25,34]	<i>BERT model (BERT)</i>	<i>Dependency on high-quality and domain-specific training data for optimal performance.</i>
[1,6,15,16,30]	<i>Deep Learning Approach (DL)</i>	<i>The requirement for a lot of labelled data, the difficulty of interpretation, the cost of computing, and the danger of over-fitting, which require careful evaluation and consideration before deployment in real-world applications.</i>
[12,29,31]	<i>Fine-Tuning (FT)</i>	<i>Shallow changes to the language model, sensitivity to grammar mistakes, and the need for careful consideration during system development.</i>
[18]	<i>Google Dialog-Flow (GD)</i>	<i>The potential for misinterpretation of user queries and the need for ongoing maintenance and updates for accuracy and relevance.</i>
[5]	<i>Human Feedback (HF)</i>	<i>Challenges in obtaining high-quality feedback, potential biases in the feedback, and the associated cost of obtaining and processing it.</i>
[6,10,17]	<i>Information Retrieval Techniques (IR)</i>	<i>Performance dependency on the IR system, dataset limitations, insufficiency for complex questions, and computational expense of advanced techniques.</i>
[7,11,21–23]	<i>Knowledge Base (KB)</i>	<i>Lack of methods for verifying and validating the systems, time-consuming and expensive development and maintenance, and limited suitability for all types of questions and domains.</i>
[12,14,35]	<i>Knowledge Distillation (KD)</i>	<i>Potential ineffectiveness for all types of models or tasks, challenges in zero-shot transfer learning, limited effectiveness for noisy or error-prone data, potential</i>

			<i>limitations for specialized or complex architectures, and limitations for applications requiring high levels of accuracy or precision.</i>
[3,8,9,19,24,32,33]	Knowledge Graph (KG)		<i>Potential incompleteness or inaccuracy of knowledge graphs, challenges in mapping natural language questions to formal query representations, computational expense of multi-hop reasoning, and difficulties in matching entities in noisy or variable user queries to the knowledge graph.</i>
[5,27–29,35]	Machine Learning Approach (ML)		<i>The need for a lot of processing power and training data, as well as the importance of trained quality of information and question difficulty, and potential challenges in handling questions that require common sense reasoning or contextual understanding.</i>
[37]	Metamorphic Testing Theory (MT)		<i>Identifying and selecting metamorphic relations, especially for complex systems, and may not be suitable for all types of QA systems that lack clear metamorphic relations for testing.</i>
[10,17]	Natural Processing Languages (NLP)		<i>Feature engineering complexity, data quality, and context maintenance, understanding human nuances, semantic-based QA, and closed domain limitations.</i>
[2,4,8,13,20,26]	Neural Network Approach (NN)		<i>Limitations in feature extraction, data requirements, context understanding, and generating specific answers, but deep learning approaches help mitigate some of these challenges.</i>
[1,6]	Recurrent Neural Network (RNN)		<i>Slow training and inference, limited representational power, difficulty in capturing long-term dependencies, and memory loss.</i>
[16]	Semantic Analysis Approach (SA)		<i>The limited application of existing systems, flaws in the models employed, and challenges in precise semantic tagging.</i>
[1]	Sequence-to-Sequence (Seq2Seq)		<i>Focused nature, weaknesses in the underlying models, and challenges in accurate semantic tagging.</i>
[16]	Template Based Approach (TB)		<i>Limitations in handling complex queries that require reasoning and inference.</i>
[12,14,29,31]	Transfer Learning (TL)		<i>Requires sufficient data and careful selection of pre-trained models for the target task.</i>

## 6. Performance Analysis:

Performance analysis is an essential tool that helps researchers and organizations to achieve their goals and objectives in technology-based research papers. It helps to evaluate and compare the performances of different technologies according to specific use cases. Additionally, it may be used to analyze how technology performance has changed and find inconsistencies between existing and future performance. The Table 11 shows performance analysis of different research articles in order to achieve goals and objectives.

**Table 11.** Performance Analysis.

REF. PAPER S	OBJECTIVES												
	A	CP	CT	C	EFT	E	IR	P	Q	$\frac{RC}{C}$	R	SC	S
[1]	Hig h	-	-	-	-	-	Hig h	-	-	-	-	-	-
[2]	-	-	-	-	-	-	-	Hig h	-	-	-	-	-
[3]	Hig h	-	-	-	-	-	-	-	-	-	-	-	-
[4]	-	-	-	-	-	-	Hig h	-	-	-	-	-	-
[5]	Hig h	-	Lo w	-	-	-	-	Hig h	-	-	-	-	-
[6]	-	-	-	-	-	-	-	Hig h	-	-	-	-	-
[7]	Hig h	-	-	Lo w	-	-	-	-	-	-	Hig h	-	-
[8]	-	-	Lo w	Lo w	-	-	-	-	-	-	-	-	-
[9]	-	-	-	-	-	-	-	-	-	-	Hig h	-	-
[10]	-	-	-	-	Hig h	-	-	-	-	-	-	-	-
[11]	Hig h	-	-	-	-	-	-	-	-	-	-	-	-
[12]	-	-	-	-	-	-	Hig h	-	-	-	-	Hig h	-
[13]	-	Lo w	Lo w	-	Hig h	-	-	Hig h	-	-	-	-	-
[14]	Hig h	-	-	-	-	-	Hig h	Hig h	-	-	-	Hig h	-
[15]	Hig h	-	-	-	-	-	-	-	-	-	Hig h	-	-
[16]	Hig h	-	-	-	-	-	-	-	-	-	-	-	-
[17]	-	Lo w	-	-	-	-	Hig h	-	Hig h	-	-	-	-
[18]	Hig h	-	-	-	-	-	-	-	-	-	-	-	-
[19]	Hig h	-	-	-	-	Hig h	-	-	-	-	-	-	-

[20]	<i>Hig</i> <i>h</i>	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-	-	-	-	-
[21]	-	-	-	-	-	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-
[22]	<i>Hig</i> <i>h</i>	-	-	<i>Lo</i> <i>w</i>	-	-	-	-	-	-	-	-	-	-
[23]	-	-	-	-	-	-	-	-	<i>Hig</i> <i>h</i>	<i>Hig</i> <i>h</i>	-	-	-	-
[24]	<i>Hig</i> <i>h</i>	-	-	-	-	-	-	-	-	-	-	-	<i>Hig</i> <i>h</i>	-
[25]	-	-	-	-	-	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-
[26]	-	-	-	-	-	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-
[27]	<i>Hig</i> <i>h</i>	-	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-	-	-	-
[28]	<i>Hig</i> <i>h</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
[29]	-	-	-	-	-	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-
[30]	-	-	-	-	-	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-
[31]	<i>Hig</i> <i>h</i>	-	-	-	-	-	-	-	-	-	-	-	-	-
[32]	-	<i>Lo</i> <i>w</i>	-	-	-	-	-	-	-	-	-	-	-	-
[33]	<i>Hig</i> <i>h</i>	<i>Lo</i> <i>w</i>	-	-	-	-	-	-	-	-	-	-	-	-
[34]	-	-	-	-	-	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-
[35]	-	-	-	-	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-	-
[36]	-	-	-	<i>Lo</i> <i>w</i>	-	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-
[37]	-	-	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-	-	-	-
[38]	-	-	-	-	-	-	-	<i>Hig</i> <i>h</i>	-	-	-	-	-	-

## 7. Research Gaps & Solutions:

Research gap is referred to as an unknown or underexplored region having room for more investigation. Any research report must identify research gaps in order to improve knowledge and create long-term studies that have a positive influence on society. Finding gaps in the literature aids

in describing the goals of the study, the issue under investigation, and the significance of the issue. Below is a discussion of research gaps in deep learning-based question answering systems and potential remedies:

- **Data scarcity:** Deep learning-based systems for answering questions need an enormous amount of top-of-the training data that might be challenging to get in some fields or languages. One solution is to develop methods for generating synthetic training data, such as those based on data augmentation or generative adversarial networks [131].
- **Lack of annotated data:** Annotated data is essential for training deep learning models, but it can be difficult and expensive to obtain. One solution is to use transfer learning, where a pre-trained model is improved on a more restricted marked dataset [131].
- **Limited ability to handle out-of-domain questions:** Deep learning models are often trained on specific domains or datasets, which can limit their ability to answer questions outside of those domains. One solution is to use transfer learning or domain adaptation techniques, which can help the model generalize to new domains [131].
- **Limited interpretability:** It might be tricky to comprehend how deep learning models generate their results since they can be complicated to comprehend. Utilisation of attention mechanisms, which enable the model to concentrate on particular elements of the input when generating predictions, is one remedy.[132].
- **Limited ability to handle low-resource languages:** It might be difficult for low-resource languages to use deep learning models since they need a lot of metadata to function successfully. One solution is to use unsupervised or semi-supervised learning techniques, which can leverage unannotated data to improve performance [132].
- **Limited generalization:** The ability of deep learning models to generalize to new contexts or tasks to perform that are dissimilar from those they were trained on might be a challenge. One way is to employ multi-task learning that involves training a single model on a variety of related tasks. This can increase the generalizability of the mode [133].
- **Lack of Explainability:** Understanding the manner in which deep learning models generate their responses might be tricky since they can be tricky to evaluate. One solution is to use methods such as attention mechanisms or explainable AI, which can provide insights into how the model makes its predictions [133].
- **Limited ability to handle noisy or ambiguous input:** Natural language input can be noisy or ambiguous, which can make it challenging for deep learning models to accurately answer questions. One solution is to use models that can handle uncertainty, such as probabilistic models or fuzzy logic-based models [133].
- **Lack of comprehensive evaluation:** Many research papers evaluate their models on a limited set of metrics or datasets, which can make it difficult to compare different approaches. One solution is to use standardized benchmarks and evaluation metrics, which can facilitate fair comparisons between different models [134].
- **Limited ability to handle multi-hop reasoning:** Multi-hop reasoning involves answering questions that require multiple pieces of information to be combined. Deep learning models can struggle with this task, as they often rely on local context rather than global context. One solution is to use graph-based models, which can represent the relationships between different pieces of information [135].
- **Limited ability to handle complex questions:** Deep learning models can struggle with answering complex questions that require reasoning and inference. One solution is to use

models that can perform logical reasoning, such as knowledge graph-based models or symbolic reasoning models [135].

- **Limited ability to handle long documents:** Deep learning models can struggle with long documents, as they may have difficulty retaining relevant information over long periods of time. One solution is to use memory-augmented models, which can store relevant information in an external memory and retrieve it when needed [136].
- **Limited ability to handle multi-lingual questions:** Deep learning models trained on one language may not perform well on questions in other languages. Cross-lingual transfer learning, which trains a model on many languages and is capable of transferring knowledge across them, is one remedy [137].

## 8. Conclusion:

Deep learning-based question answering systems have emerged as a vibrant area of research within natural language processing. Although they have shown encouraging results, there are still a number of major issues that must be resolved if they are going to work better and be more useful in everyday situations. These challenges include handling complex questions requiring reasoning, dealing with noisy or ambiguous input, supporting multiple languages, processing non-textual data, and handling rare or unseen words. Proposed solutions involve text matching models, semantic matching rankers, fine-tuning of multilingual models, and alike. Further research is necessary to develop robust and accurate deep learning-based question answering systems that can effectively handle diverse real-world scenarios.

**Acknowledgments:** By the Grace of Almighty Allah, I Miss. Nayyab Saeed, the author of this paper expresses heartfelt gratitude to my parents for their unwavering support and encouragement throughout this journey. Special thanks Dr. Humaira Ashraf and the all-other contributors and collaborators for guiding me with valuable resources and contributions which helped me in writing this paper.

**Conflicts of Interest:** There is no conflict of interest for this study.

## References

1. B. Afrae, B. A. Mohamed, and A. A. Boudhir, "A Question Answering System with a sequence-to-sequence grammatical correction," in *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, Marrakech Morocco: ACM, Mar. 2020, pp. 1–6. doi: 10.1145/3386723.3387894.
2. Y. Deng, W. Zhang, and W. Lam, "Opinion-aware Answer Generation for Review-driven Question Answering in E-Commerce," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, in CIKM '20. New York, NY, USA: Association for Computing Machinery, Oct. 2020, pp. 255–264. doi: 10.1145/3340531.3411904.
3. R. Gao and C. Li, "Knowledge Question-Answering System Based on Knowledge Graph of Traditional Chinese Medicine," in *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Dec. 2020, pp. 27–31. doi: 10.1109/ITAIC49862.2020.9339040.
4. Z. Gao, X. Xia, D. Lo, and J. Grundy, "Technical Q&A Site Answer Recommendation via Question Boosting," *ACM Trans. Softw. Eng. Methodol.*, vol. 30, no. 1, p. 11:1-11:34, Dec. 2021, doi: 10.1145/3412845.
5. V. Gutta, N. B. Unnam, and P. K. Reddy, "An improved human-in-the-loop model for fine-grained object recognition with batch-based question answering," in *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, in CoDS COMAD 2020. New York, NY, USA: Association for Computing Machinery, Jan. 2020, pp. 134–142. doi: 10.1145/3371158.3371174.
6. C. Liu, T. He, Y. Xiong, H. Wang, and J. Chen, "A Novel Knowledge Base Question Answering Model Based on Knowledge Representation and Recurrent Convolutional Neural Network," in *2020 International Conference on Service Science (ICSS)*, Aug. 2020, pp. 149–154. doi: 10.1109/ICSS50103.2020.00031.

7. X.-Y. Liu, Y. Zhang, Y. Liao, and L. Jiang, "Dynamic Updating of the Knowledge Base for a Large-Scale Question Answering System," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 19, no. 3, p. 45:1-45:13, Feb. 2020, doi: 10.1145/3377708.
8. Y. Qiu, Y. Wang, X. Jin, and K. Zhang, "Stepwise Reasoning for Multi-Relation Question Answering over Knowledge Graph with Weak Supervision," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, in WSDM '20. New York, NY, USA: Association for Computing Machinery, Jan. 2020, pp. 474–482. doi: 10.1145/3336191.3371812.
9. D. Wang, Y. Shen, and H.-T. Zheng, "Knowledge Enhanced Latent Relevance Mining for Question Answering," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pp. 4282–4286. doi: 10.1109/ICASSP40776.2020.9053432.
10. X. Wang and Z. Wang, "Question Answering System based on Disease Knowledge Base," in *2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS)*, Oct. 2020, pp. 351–354. doi: 10.1109/ICSESS49938.2020.9237712.
11. Z. Wu, B. Kao, T.-H. Wu, P. Yin, and Q. Liu, "PERQ: Predicting, Explaining, and Rectifying Failed Questions in KB-QA Systems," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, in WSDM '20. New York, NY, USA: Association for Computing Machinery, Jan. 2020, pp. 663–671. doi: 10.1145/3336191.3371782.
12. Z. Yang, L. Shou, M. Gong, W. Lin, and D. Jiang, "Model Compression with Two-stage Multi-teacher Knowledge Distillation for Web Question Answering System," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, in WSDM '20. New York, NY, USA: Association for Computing Machinery, Jan. 2020, pp. 690–698. doi: 10.1145/3336191.3371792.
13. H. Buzaaba and T. Amagasa, "A Scheme for Efficient Question Answering with Low Dimension Reconstructed Embeddings," in *The 23rd International Conference on Information Integration and Web Intelligence*, in iiWAS2021. New York, NY, USA: Association for Computing Machinery, Dec. 2022, pp. 303–310. doi: 10.1145/3487664.3487705.
14. C. Chen, C. Wang, M. Qiu, D. Gao, L. Jin, and W. Li, "Cross-domain Knowledge Distillation for Retrieval-based Question Answering Systems," in *Proceedings of the Web Conference 2021*, in WWW '21. New York, NY, USA: Association for Computing Machinery, Jun. 2021, pp. 2613–2623. doi: 10.1145/3442381.3449814.
15. S. Jain, M. Khanna, and Ankita, "Context-Aware Deep Learning Approach for Answering Questions," in *2021 IEEE 6th International Conference on Computing, Communication and Automation (ICCCA)*, Dec. 2021, pp. 26–31. doi: 10.1109/ICCCA52192.2021.9666296.
16. T. Jin and H.-J. Wang, "Research on the method of knowledge base question answering," in *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLDBI)*, Dec. 2021, pp. 527–530. doi: 10.1109/MLDBI54094.2021.00106.
17. K. Jorgensen, Z. Zhao, H. Wang, M. Wang, and Z. He, "Context-Aware Question-Answer for Interactive Media Experiences," in *ACM International Conference on Interactive Media Experiences*, in IMX '21. New York, NY, USA: Association for Computing Machinery, Jun. 2021, pp. 156–166. doi: 10.1145/3452918.3458795.
18. N. Kanodia, K. Ahmed, and Y. Miao, "Question Answering Model Based Conversational Chatbot using BERT Model and Google Dialogflow," in *2021 31st International Telecommunication Networks and Applications Conference (ITNAC)*, Nov. 2021, pp. 19–22. doi: 10.1109/ITNAC53136.2021.9652153.
19. B. Liu, R. Yan, Y. Zuo, and Y. Tao, "A Knowledge-based Question-Answering Method for Military Critical Information Under Limited Corpus," in *2021 2nd International Conference on Computer Engineering and Intelligent Control (ICCEIC)*, Nov. 2021, pp. 86–91. doi: 10.1109/ICCEIC54227.2021.00025.
20. Q. Liu, X. Geng, H. Huang, T. Qin, J. Lu, and D. Jiang, "MGRC: An End-to-End Multigranularity Reading Comprehension Model for Question Answering," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2021, doi: 10.1109/TNNLS.2021.3107029.
21. X. Liu *et al.*, "Research and Application of Intelligent Question-Answer Algorithm Based on Multi-Channel Dilated Convolution," in *2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT)*, Nov. 2021, pp. 776–780. doi: 10.1109/ICESIT53460.2021.9696734.

22. F. Meng, W. Wang, and J. Wang, "Research on Short Text Similarity Calculation Method for Power Intelligent Question Answering," in *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)*, Sep. 2021, pp. 91–95. doi: 10.1109/CICN51697.2021.9574692.
23. T. Ploumis, I. Perikos, F. Grivokostopoulou, and I. Hatzilygeroudis, "A Factoid based Question Answering System based on Dependency Analysis and Wikidata," in *2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA)*, Jul. 2021, pp. 1–7. doi: 10.1109/IISA52424.2021.9555551.
24. Y. Tang, H. Han, X. Yu, J. Zhao, G. Liu, and L. Wei, "An Intelligent Question Answering System based on Power Knowledge Graph," in *2021 IEEE Power & Energy Society General Meeting (PESGM)*, Jul. 2021, pp. 01–05. doi: 10.1109/PESGM46819.2021.9638018.
25. S. Xu, F. Liu, Z. Huang, Y. Peng, and D. Li, "A BERT-Based Semantic Matching Ranker for Open-Domain Question Answering," in *Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval*, in *NLPiR 2020*. New York, NY, USA: Association for Computing Machinery, Feb. 2021, pp. 31–36. doi: 10.1145/3443279.3443301.
26. R. Zhao, S. Wu, Y. Mao, and N. Li, "A Multi-Level Semantic Fusion Algorithm Based on Historical Data in Question Answering," in *2021 7th Annual International Conference on Network and Information Systems for Computers (ICNISC)*, Jul. 2021, pp. 950–956. doi: 10.1109/ICNISC54316.2021.00178.
27. I. Annamoradnejad, "Requirements for Automating Moderation in Community Question-Answering Websites," in *15th Innovations in Software Engineering Conference*, in *ISEC 2022*. New York, NY, USA: Association for Computing Machinery, Feb. 2022, pp. 1–4. doi: 10.1145/3511430.3511458.
28. R. Butler, V. D. Duggirala, and F. Banaei-Kashani, "iLFQA: A Platform for Efficient and Accurate Long-Form Question Answering," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, in *WSDM '22*. New York, NY, USA: Association for Computing Machinery, Feb. 2022, pp. 1565–1568. doi: 10.1145/3488560.3502190.
29. C.-C. Kuo and K.-Y. Chen, "Toward Zero-Shot and Zero-Resource Multilingual Question Answering," *IEEE Access*, vol. 10, pp. 99754–99761, 2022, doi: 10.1109/ACCESS.2022.3207569.
30. C. Li, X. Lu, and K. Shuang, "CypherQA: Question-answering method based on Attribute Knowledge Graph," in *2021 The 9th International Conference on Information Technology: IoT and Smart City*, in *ICIT 2021*. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 243–247. doi: 10.1145/3512576.3512620.
31. A. Malhar, P. Sawant, Y. Chhadva, and S. Kurhade, "Deep learning based Answering Questions using T5 and Structured Question Generation System," in *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, May 2022, pp. 1544–1549. doi: 10.1109/ICICCS53718.2022.9788264.
32. E. Niazmand, "Enhancing Query Answer Completeness with Query Expansion based on Synonym Predicates," in *Companion Proceedings of the Web Conference 2022*, in *WWW '22*. New York, NY, USA: Association for Computing Machinery, Aug. 2022, pp. 354–358. doi: 10.1145/3487553.3524198.
33. J. Tian, B. Li, Y. Ji, and J. Wu, "Text-Enhanced Question Answering over Knowledge Graph," in *Proceedings of the 10th International Joint Conference on Knowledge Graphs*, in *IJCKG '21*. New York, NY, USA: Association for Computing Machinery, Jan. 2022, pp. 135–139. doi: 10.1145/3502223.3502241.
34. C. Wang and X. Luo, "A Legal Question Answering System Based on BERT&#xa0;," in *2021 5th International Conference on Computer Science and Artificial Intelligence*, in *CSAI 2021*. New York, NY, USA: Association for Computing Machinery, Mar. 2022, pp. 278–283. doi: 10.1145/3507548.3507591.
35. J. Cao, J. Cao, and Y. Wang, "Gentle Teaching Assistant Model For Improved Accuracy In Question-Answering," in *2022 5th International Conference on Machine Learning and Natural Language Processing*, in *MLNLP2022*. New York, NY, USA: Association for Computing Machinery, Mar. 2023, pp. 353–359. doi: 10.1145/3578741.3578816.
36. Y.-H. Chen, E. J.-L. Lu, and Y.-Y. Lin, "Efficient SPARQL Queries Generator for Question Answering Systems," *IEEE Access*, vol. 10, pp. 99850–99860, 2022, doi: 10.1109/ACCESS.2022.3206794.
37. Z. Liu, Y. Feng, Y. Yin, J. Sun, Z. Chen, and B. Xu, "QATest: A Uniform Fuzzing Framework for Question Answering Systems," in *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, in *ASE '22*. New York, NY, USA: Association for Computing Machinery, Jan. 2023, pp. 1–12. doi: 10.1145/3551349.3556929.

38. J. Wu, T. Mu, J. Thiyyagalingam, and J. Y. Goulermas, "Memory-Aware Attentive Control for Community Question Answering With Knowledge-Based Dual Refinement," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–14, 2023, doi: 10.1109/TSMC.2023.3234297.

39. "6879936.pdf." Accessed: Apr. 25, 2023. [Online]. Available: <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6879936.pdf>

40. Y. Shen, E. M.-K. Lai, and M. Mohaghegh, "Effects of Similarity Score Functions in Attention Mechanisms on the Performance of Neural Question Answering Systems," *Neural Process Lett*, vol. 54, no. 3, pp. 2283–2302, Jun. 2022, doi: 10.1007/s11063-021-10730-4.

41. Y. Bachrach *et al.*, "An Attention Mechanism for Answer Selection Using a Combined Global and Local View." *arXiv*, Sep. 20, 2017. Accessed: Apr. 25, 2023. [Online]. Available: <http://arxiv.org/abs/1707.01378>

42. A. Osman and W. Samek, "DRAU: Dual Recurrent Attention Units for Visual Question Answering," *Computer Vision and Image Understanding*, vol. 185, pp. 24–30, Aug. 2019, doi: 10.1016/j.cviu.2019.05.001.

43. S. Zhao and Z. Jin, "Research Progress of Automatic Question Answering System Based on Deep Learning," *Open Access Library Journal*, vol. 7, no. 6, Art. no. 6, Jun. 2020, doi: 10.4236/oalib.1106046.

44. G. Brauwers and F. Frasincar, "A General Survey on Attention Mechanisms in Deep Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3279–3298, Apr. 2023, doi: 10.1109/TKDE.2021.3126456.

45. S. Varma, S. Shivam, S. Biswas, P. Saha, and K. Jalan, "Graph NLU enabled question answering system," *Heliyon*, vol. 7, no. 9, p. e08035, Sep. 2021, doi: 10.1016/j.heliyon.2021.e08035.

46. "A knowledge graph based question answering method for medical domain - PMC." <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8444078/> (accessed Apr. 25, 2023).

47. P. Mital *et al.*, Graph based Question Answering System. 2018.

48. S. Vakulenko, J. D. F. Garcia, A. Polleres, M. de Rijke, and M. Cochez, "Message Passing for Complex Question Answering over Knowledge Graphs." *arXiv*, Aug. 19, 2019. Accessed: Apr. 25, 2023. [Online]. Available: <http://arxiv.org/abs/1908.06917>

49. M. Shi, "Knowledge Graph Question and Answer System for Mechanical Intelligent Manufacturing Based on Deep Learning," *Mathematical Problems in Engineering*, vol. 2021, p. e6627114, Feb. 2021, doi: 10.1155/2021/6627114.

50. "BERT 101 - State Of The Art NLP Model Explained." <https://huggingface.co/blog/bert-101> (accessed Apr. 25, 2023).

51. "How to Train A Question-Answering Machine Learning Model | Paperspace Blog." <https://blog.paperspace.com/how-to-train-question-answering-machine-learning-models/> (accessed Apr. 25, 2023).

52. "How to build Question Answering system for online store with BERT," *Grid Dynamics Blog*, Oct. 15, 2020. <https://blog.griddynamics.com/question-answering-system-using-bert/> (accessed Apr. 25, 2023).

53. "COBERT: COVID-19 Question Answering System Using BERT - PMC." <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8220121/> (accessed Apr. 25, 2023).

54. C. Zhang, Y. Lai, Y. Feng, and D. Zhao, "A review of deep learning in question answering over knowledge bases," *AI Open*, vol. 2, pp. 205–215, Jan. 2021, doi: 10.1016/j.aiopen.2021.12.001.

55. E. Mutabazi, J. Ni, G. Tang, and W. Cao, "A Review on Medical Textual Question Answering Systems Based on Deep Learning Approaches," *Applied Sciences*, vol. 11, no. 12, Art. no. 12, Jan. 2021, doi: 10.3390/app11125456.

56. "How to Train A Question-Answering Machine Learning Model," *Paperspace Blog*, Dec. 14, 2020. <https://blog.paperspace.com/how-to-train-question-answering-machine-learning-models/> (accessed Apr. 26, 2023).

57. Y. Sharma and S. Gupta, "Deep Learning Approaches for Question Answering System," *Procedia Computer Science*, vol. 132, pp. 785–794, Jan. 2018, doi: 10.1016/j.procs.2018.05.090.

58. R. Alqifari, "Question Answering Systems Approaches and Challenges," in *Proceedings of the Student Research Workshop Associated with RANLP 2019*, Varna, Bulgaria: INCOMA Ltd., Sep. 2019, pp. 69–75. doi: 10.26615/issn.2603-2821.2019\_011.

59. E. Stroh and P. Mathur, "Question Answering Using Deep Learning".

60. G. Shachaf, A. Brutzkus, and A. Globerson, "A Theoretical Analysis of Fine-tuning with Linear Teachers." arXiv, Nov. 07, 2021. Accessed: Apr. 26, 2023. [Online]. Available: <http://arxiv.org/abs/2107.01641>

61. "Building a QA System with BERT on Wikipedia," *NLP for Question Answering*, May 19, 2020. [https://qa.fastforwardlabs.com/pytorch/hugging%20face/wikipedia/bert/transformers/2020/05/19/Getting\\_Started\\_with\\_QA.html](https://qa.fastforwardlabs.com/pytorch/hugging%20face/wikipedia/bert/transformers/2020/05/19/Getting_Started_with_QA.html) (accessed Apr. 26, 2023).

62. "pdf.pdf." Accessed: Apr. 26, 2023. [Online]. Available: <https://openreview.net/pdf?id=ks4BvF7kpiP>

63. D. Dimitriadis and G. Tsoumacas, "Artificial fine-tuning tasks for yes/no question answering," *Natural Language Engineering*, pp. 1–23, Jun. 2022, doi: 10.1017/S1351324922000286.

64. J. Briggs, "How-to Fine-Tune a Q&A Transformer," *Medium*, Sep. 02, 2021. <https://towardsdatascience.com/how-to-fine-tune-a-q-a-transformer-86f91ec92997> (accessed Apr. 26, 2023).

65. "Solving internal search problems with Dialogflow," *Google Cloud Blog*. <https://cloud.google.com/blog/topics/developers-practitioners/solving-internal-search-problems-dialogflow> (accessed Apr. 26, 2023).

66. "What is Google Dialogflow?" <https://start-up.house/en/blog/articles/what-is-google-dialogflow> (accessed Apr. 26, 2023).

67. "Testing best practices (Dialogflow) | Dialogflow and legacy Actions SDK," *Google Developers*. <https://developers.google.com/assistant/df-asdk/dialogflow/testing-best-practices> (accessed Apr. 26, 2023).

68. "Dialogflow Review: Pricing, Pros, Cons & Features," *CompareCamp.com*, Feb. 01, 2019. <https://comparecamp.com/dialogflow-review-pricing-pros-cons-features/> (accessed Apr. 26, 2023).

69. Z. Li, P. Sharma, X. H. Lu, J. Cheung, and S. Reddy, "Using Interactive Feedback to Improve the Accuracy and Explainability of Question Answering Systems Post-Deployment," in *Findings of the Association for Computational Linguistics: ACL 2022*, Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 926–937. doi: 10.18653/v1/2022.findings-acl.75.

70. "send.pdf." Accessed: Apr. 26, 2023. [Online]. Available: [https://etd.ohiolink.edu/apexprod/rws\\_etd/send\\_file/send?accession=osu1609944750779603&disposition=inline](https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1609944750779603&disposition=inline)

71. S. K. Dwivedi and V. Singh, "Research and Reviews in Question Answering System," *Procedia Technology*, vol. 10, pp. 417–424, 2013, doi: 10.1016/j.protcy.2013.12.378.

72. J. A. Campos, K. Cho, A. Otegi, A. Soroa, E. Agirre, and G. Azkune, "Improving Conversational Question Answering Systems after Deployment using Feedback-Weighted Learning," in *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 2561–2571. doi: 10.18653/v1/2020.coling-main.230.

73. Z. Li, P. Sharma, X. H. Lu, J. Cheung, and S. Reddy, "Using Interactive Feedback to Improve the Accuracy and Explainability of Question Answering Systems Post-Deployment," in *Findings of the Association for Computational Linguistics: ACL 2022*, Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 926–937. doi: 10.18653/v1/2022.findings-acl.75.

74. "da8a13559cc19da8ab4eb8eb5f065e70019fd74c.pdf." Accessed: Apr. 26, 2023. [Online]. Available: <https://openreview.net/pdf/da8a13559cc19da8ab4eb8eb5f065e70019fd74c.pdf>

75. "Evaluating QA: the Retriever & the Full QA System," *NLP for Question Answering*, Jun. 30, 2020. [https://qa.fastforwardlabs.com/elasticsearch/mean%20average%20precision/recall%20for%20irqa/qa%20system%20design/2020/06/30/Evaluating\\_the\\_Retriever\\_&\\_End\\_to\\_End\\_System.html](https://qa.fastforwardlabs.com/elasticsearch/mean%20average%20precision/recall%20for%20irqa/qa%20system%20design/2020/06/30/Evaluating_the_Retriever_&_End_to_End_System.html) (accessed Apr. 27, 2023).

76. Á. Rodrigo, J. Pérez-Iglesias, A. Peñas, G. Garrido, and L. Araujo, *A Question Answering System based on Information Retrieval and Validation*. 2010.

77. Z. Wang, V. X. Lin, and L. Mehr, "Deep Retriever: Information Retrieval for Multi-hop Question Answering".

78. "What is a Knowledge-based System? | Definition from TechTarget," *CIO*. <https://www.techtarget.com/searchcio/definition/knowledge-based-systems-KBS> (accessed May 08, 2023).

79. B. Zope, S. Mishra, K. Shaw, D. R. Vora, K. Kotecha, and R. V. Bidwe, "Question Answer System: A State-of-Art Representation of Quantitative and Qualitative Analysis," *Big Data and Cognitive Computing*, vol. 6, no. 4, Art. no. 4, Dec. 2022, doi: 10.3390/bdcc6040109.

80. "(PDF) A Structured Testing Methodology for Knowledge-Based Systems." [https://www.researchgate.net/publication/221464216\\_A\\_Structured\\_Testing\\_Methodology\\_for\\_Knowledge-Based\\_Systems](https://www.researchgate.net/publication/221464216_A_Structured_Testing_Methodology_for_Knowledge-Based_Systems) (accessed May 08, 2023).

81. "Testing Knowledge Based Systems: A Case Study and Implications - ScienceDirect." <https://www.sciencedirect.com/science/article/pii/S147466701749447X> (accessed May 08, 2023).

82. D. Diefenbach, V. Lopez, K. Singh, and P. Maret, "Core techniques of question answering systems over knowledge bases: a survey," *Knowl Inf Syst*, vol. 55, no. 3, pp. 529–569, Jun. 2018, doi: 10.1007/s10115-017-1100-y.

83. K. Singh, I. Lytra, A. S. Radhakrishna, S. Shekarpour, M.-E. Vidal, and J. Lehmann, "No one is perfect: Analysing the performance of question answering components over the DBpedia knowledge graph," *Journal of Web Semantics*, vol. 65, p. 100594, Dec. 2020, doi: 10.1016/j.websem.2020.100594.

84. "Knowledge Distillation: Principles & Algorithms [+Applications]." <https://www.v7labs.com/blog/knowledge-distillation-guide>, <https://www.v7labs.com/blog/knowledge-distillation-guide> (accessed May 08, 2023).

85. D. User, "Introduction to Knowledge Distillation," *Deci*, Jun. 29, 2022. <https://deci.ai/blog/knowledge-distillation-introduction/> (accessed May 08, 2023).

86. C. Chen, C. Wang, M. Qiu, D. Gao, L. Jin, and W. Li, "Cross-domain Knowledge Distillation for Retrieval-based Question Answering Systems," in *Proceedings of the Web Conference 2021*, in *WWW '21*. New York, NY, USA: Association for Computing Machinery, Jun. 2021, pp. 2613–2623. doi: 10.1145/3442381.3449814.

87. "Improving Question Answering Performance Using Knowledge Distillation and Active Learning," *ar5iv*. <https://ar5iv.labs.arxiv.org/html/2109.12662> (accessed May 08, 2023).

88. S. Soltan, H. Khan, and W. Hamza, "Limitations of Knowledge Distillation for Zero-shot Transfer Learning," in *Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing*, Virtual: Association for Computational Linguistics, Nov. 2021, pp. 22–31. doi: 10.18653/v1/2021.sustainlp-1.3.

89. "Limitations of knowledge distillation for zero-shot transfer learning," *Amazon Science*. <https://www.amazon.science/publications/limitations-of-knowledge-distillation-for-zero-shot-transfer-learning> (accessed May 08, 2023).

90. "Question Answering based on Knowledge Graphs." <https://www.linkedin.com/pulse/question-answering-based-knowledge-graphs-andreas-blumauer> (accessed May 08, 2023).

91. Y. Zhang, H. Dai, Z. Kozareva, A. Smola, and L. Song, "Variational Reasoning for Question Answering With Knowledge Graph," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, Art. no. 1, Apr. 2018, doi: 10.1609/aaai.v32i1.12057.

92. M. Yani and A. A. Krisnadhi, "Challenges, Techniques, and Trends of Simple Knowledge Graph Question Answering: A Survey," *Information*, vol. 12, no. 7, Art. no. 7, Jul. 2021, doi: 10.3390/info12070271.

93. S. Liang, K. Stockinger, T. M. de Farias, M. Anisimova, and M. Gil, "Querying knowledge graphs in natural language," *Journal of Big Data*, vol. 8, no. 1, p. 3, Jan. 2021, doi: 10.1186/s40537-020-00383-w.

94. "How to Train A Question-Answering Machine Learning Model," *Paperspace Blog*, Dec. 14, 2020. <https://blog.paperspace.com/how-to-train-question-answering-machine-learning-models/> (accessed May 08, 2023).

95. D. Marijan, A. Gotlieb, and M. Ahuja, *Challenges of Testing Machine Learning Based Systems*. 2019, p. 102. doi: 10.1109/AITest.2019.00010.

96. S. Omri and C. Sinz, "Machine Learning Techniques for Software Quality Assurance: A Survey." *arXiv*, Apr. 28, 2021. Accessed: May 08, 2023. [Online]. Available: <http://arxiv.org/abs/2104.14056>

97. "QA Systems and Deep Learning Technologies – Part 1 - Alibaba Cloud Community." [https://www.alibabacloud.com/blog/qa-systems-and-deep-learning-technologies-part-1\\_71899](https://www.alibabacloud.com/blog/qa-systems-and-deep-learning-technologies-part-1_71899) (accessed May 08, 2023).

98. E. Altamimi, A. Elkawakjy, and C. Catal, "Metamorphic relation automation: Rationale, challenges, and solution directions," *Journal of Software: Evolution and Process*, vol. 35, no. 1, p. e2509, 2023, doi: 10.1002/sm.2509.

99. K. Tu, M. Jiang, and Z. Ding, "A Metamorphic Testing Approach for Assessing Question Answering Systems," *Mathematics*, vol. 9, no. 7, Art. no. 7, Jan. 2021, doi: 10.3390/math9070726.

100. F. Höhn, "Natural Language Question Answering Systems," *SEEBURGER Blog*, Jan. 18, 2022. <https://blog.seeburger.com/natural-language-question-answering-systems-get-quick-answers-to-concrete-questions/> (accessed May 08, 2023).

101. "advanced-qa.pdf." Accessed: May 08, 2023. [Online]. Available: <https://www.cs.cornell.edu/courses/cs474/2005fa/Handouts/advanced-qa.pdf>

102. "Weights & Biases," *W&B*. <https://wandb.ai/mostafaibrahim17/ml-articles/reports/The-Answer-Key-Unlocking-the-Potential-of-Question-Answering-With-NLP--VmlldzozNTcxMDE3> (accessed May 08, 2023).

103. "Question answering," *Wikipedia*. May 06, 2023. Accessed: May 08, 2023. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Question\\_answering&oldid=1153385768](https://en.wikipedia.org/w/index.php?title=Question_answering&oldid=1153385768)

104. R. Alqifari, "Question Answering Systems Approaches and Challenges," in *Proceedings of the Student Research Workshop Associated with RANLP 2019*, Varna, Bulgaria: INCOMA Ltd., Sep. 2019, pp. 69–75. doi: 10.26615/issn.2603-2821.2019\_011.

105. "Limitations of NLP in building Question/Answering systems - Proxzar." <https://www.proxzar.ai/blog/limitations-of-nlp-in-question-answering/> (accessed May 08, 2023).

106. "TMLde1.pdf." Accessed: May 08, 2023. [Online]. Available: <https://upcommons.upc.edu/bitstream/handle/2117/118781/TMLde1.pdf>

107. A. Team, "Intelligent QA Systems With NLP for Optimizing Customer Interactions," *AI Oodles*, Jun. 05, 2020. <https://artificialintelligence.oodles.io/blogs/qa-system-with-nlp/> (accessed May 08, 2023).

108. R. Alqifari, "Question Answering Systems Approaches and Challenges," in *Proceedings of the Student Research Workshop Associated with RANLP 2019*, Varna, Bulgaria: INCOMA Ltd., Sep. 2019, pp. 69–75. doi: 10.26615/issn.2603-2821.2019\_011.

109. L. Kodra, "A Review on Neural Network Question Answering Systems," *International Journal of Artificial Intelligence & Applications*, vol. 8, pp. 59–74, Mar. 2017, doi: 10.5121/ijai.2017.8206.

110. "Limitations of NLP in building Question/Answering systems," *Proxzar*, Jul. 22, 2020. <https://www.proxzar.ai/blog/limitations-of-nlp-in-question-answering/> (accessed May 08, 2023).

111. "Recurrent Neural Network (RNN) Tutorial: Types and Examples [Updated] | Simplilearn," *Simplilearn.com*. <https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn> (accessed May 16, 2023).

112. A. Bansal, Z. Eberhart, L. Wu, and C. McMillan, "A Neural Question Answering System for Basic Questions about Subroutines." *arXiv*, Jan. 11, 2021. Accessed: May 16, 2023. [Online]. Available: <http://arxiv.org/abs/2101.03999>

113. [https://web.stanford.edu/class/cs224n/reports/final\\_summaries/summary215.html](https://web.stanford.edu/class/cs224n/reports/final_summaries/summary215.html) (Accessed May 16, 2023).

114. M.-H. Wang, S. S. Yalaburg, and L. Shankarrao, "Question Answering System".

115. S. I. A. Saany, A. Mamat, A. Mustapha, L. S. Affendey, and M. N. A. Rahman, "Semantics question analysis model for question answering system," *ams*, vol. 9, pp. 6491–6505, 2015, doi: 10.12988/ams.2015.54361.

116. S. S. Alanazi, N. Elfadil, M. Jarajreh, and S. Algarni, "Question Answering Systems: A Systematic Literature Review," *IJACSA*, vol. 12, no. 3, 2021, doi: 10.14569/IJACSA.2021.0120359.

117. N. N. Khin and K. M. Soe, "Question Answering based University Chatbot using Sequence to Sequence Model," in 2020 23rd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA), Nov. 2020, pp. 55–59. doi: 10.1109/O-COCOSDA50338.2020.9295021.

118. R. Agarwal, T. Jain, and K. Mehta, "Question Generation System using Seq2Seq".

119. "Question Answering (Part 4): Transformers | by Atul Singh, PhD | Medium." <https://atulsinghphd.medium.com/question-answering-part-4-transformers-6160a30fda62> (accessed May 16, 2023).

120. W. Hou, "Seq2seq-Attention Question Answering Model".

121. S. S. Alanazi, N. Elfadil, M. Jarajreh, and S. Algarni, "Question Answering Systems: A Systematic Literature Review," *IJACSA*, vol. 12, no. 3, 2021, doi: 10.14569/IJACSA.2021.0120359.

122. L. Biermann, S. Walter, and P. Cimiano, "A Guided Template-Based Question Answering System over Knowledge Graphs".

123. "(PDF) SPARQL Template-Based Question Answering." [https://www.researchgate.net/publication/240615221\\_SPARQL\\_Template-Based\\_Question\\_Answering](https://www.researchgate.net/publication/240615221_SPARQL_Template-Based_Question_Answering) (accessed May 16, 2023).

124. J. Gomes, R. C. de Mello, V. Ströele, and J. F. de Souza, "A Hereditary Attentive Template-based Approach for Complex Knowledge Base Question Answering Systems," *Expert Systems with Applications*, vol. 205, p. 117725, Nov. 2022, doi: 10.1016/j.eswa.2022.117725.

125. K. Höffner *et al.*, "User Interface for a Template Based Question Answering System," in *Knowledge Engineering and the Semantic Web*, P. Klinov and D. Mouromtsev, Eds., in *Communications in Computer and Information Science*. Berlin, Heidelberg: Springer, 2013, pp. 258–264. doi: 10.1007/978-3-642-41360-5\_21.

126. S. S. Alanazi, N. Elfadil, M. Jarajreh, and S. Algarni, "Question Answering Systems: A Systematic Literature Review," *IJACSA*, vol. 12, no. 3, 2021, doi: 10.14569/IJACSA.2021.0120359.

127. "Question answering as a 'lingua franca' for transfer learning - Amazon Science." <https://www.amazon.science/blog/question-answering-as-a-lingua-franca-for-transfer-learning> (accessed May 16, 2023).

128. Y.-A. Chung, H.-Y. Lee, and J. Glass, "Supervised and Unsupervised Transfer Learning for Question Answering," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 1585–1594. doi: 10.18653/v1/N18-1143.

129. Y. Boreshban, S. M. Mirbostani, G. Ghassem-Sani, S. A. Mirroshandel, and S. Amiriparian, "Improving question answering performance using knowledge distillation and active learning," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106137, Aug. 2023, doi: 10.1016/j.engappai.2023.106137.

130. T. M. Lai, T. Bui, N. Lipka, and S. Li, "Supervised Transfer Learning for Product Information Question Answering," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2018, pp. 1109–1114. doi: 10.1109/ICMLA.2018.00180.

131. E. Stroh and P. Mathur, "Question Answering Using Deep Learning".

132. C. Zhang, Y. Lai, Y. Feng, and D. Zhao, "A review of deep learning in question answering over knowledge bases," *AI Open*, vol. 2, pp. 205–215, Jan. 2021, doi: 10.1016/j.aiopen.2021.12.001.

133. Y. Sharma and S. Gupta, "Deep Learning Approaches for Question Answering System," *Procedia Computer Science*, vol. 132, pp. 785–794, 2018, doi: 10.1016/j.procs.2018.05.090.

134. S. Alanazi, N. Mohamed, M. Jarajreh, and S. Algarni, "Question Answering Systems: A Systematic Literature Review," *International Journal of Advanced Computer Science and Applications*, vol. 12, Jan. 2021, doi: 10.14569/IJACSA.2021.0120359.

135. M. Haroon, Deep Learning Based Question Answering System: A Review. 2019. doi: 10.13140/RG.2.2.18591.59044.

136. L. Nie and J. Ye, "Question answering algorithm based on deep learning," in *Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering*, in EITCE '21. New York, NY, USA: Association for Computing Machinery, Dec. 2022, pp. 731–734. doi: 10.1145/3501409.3501541.

137. H. Abdel-Nabi, A. Awajan, and M. Ali, "Deep learning-based question answering: a survey," *Knowledge and Information Systems*, vol. 65, pp. 1–87, Dec. 2022, doi: 10.1007/s10115-022-01783-5.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.