AI in Healthcare: A Survey on Medical Question Answering System

Hemant Yadav[1], Pooja Yadav[2*], Nagender Yadav [3], Prakhar Chaudhary[4]

[1]Future Institute of Engineering and Technology, Future University, Bareilly, Uttar Pradesh, India. Email: drhemantyadav22@gmail.com
[2] Deptt. of CSIT, FET, MJP Rohilkhand University, Bareilly, Uttar Pradesh, India. Email: pooja.yadav@mjpru.ac.in*
[3]Specialist Master, Deloitte Consulting LLP , Sales force Tower, 111 Monument Circle Suite 4200, Indianapolis, IN, 46204 , USA. Email: nagyadav@deloitte.com
[4] Senior Manager Consulting, Cognizant Technology Solutions, Pune, Maharashtra, India. Email: prakharpc@gmail.com
[*]Corresponding Author: Pooja Yadav, Email: erpoojafet15@gmail.com ; pooja.yadav@mjpru.ac.in

KEYWORDS
Information Retrieval, Deep Learning, Machine Learning, Question Answer System, Neural Network.

ABSTRACT:
The new era of question-answering systems in information retrieval has the potential to open up previously uncharted territory soon. Information retrieval (IR) is the area of competence for Question Answering Systems. It is useful in several fields, including personal help, healthcare, finance, and education. Instead of giving document lists as retrieval systems often do, the question-answering job aims to recover brief passages of text that answer a question. The fundamental concept of the question-answering system is to emphasize human-machine interaction. It is a method for retrieving accurate answers to questions posed by people. The three separate components that make up Question Answering are Question Classification, Question Pre-processing, and Answer Retrieval. This paper provides an overview of the Question Answering system along with the usage of three basic modules with their detailed description. This study aims to present a thorough review of the state of AI-driven medical question-answering technology.

1. Introduction

In the healthcare industry, artificial intelligence (AI) has become a disruptive force that holds the potential to improve patient care, diagnosis, treatment planning, and decision-making [1]. The medical question-answering (QA) system is one of the numerous AI applications in healthcare that are essential to providing patients and healthcare providers with fast, accurate, and trustworthy medical information. These technologies are intended to help clinical decision-making and enhance patient education by deciphering complex medical inquiries and offering evidence-based responses. It looks at the fundamental technologies that support these systems, assesses how effective they are, and pinpoints important issues that require attention [2][3][4].

Medical-based question-answering (MQA) systems have advanced to the point that they can now access enormous databases of medical data and comprehend complex medical terminology because of the quick development of Natural Language Processing (NLP) and machine learning techniques. However, the intricacy of the medical field creates special difficulties, such as the necessity to understand medical jargon precisely, deal with moral dilemmas, and protect patient privacy [5].

The aim of a QA system is to help the user with the most appropriate solutions or remedies to the posed inquiry along with its justification rather than displaying the entire long lists of matching documents stored in the databases which most information retrieval (IR) presently do. The expedited way of QA systems to solutions to user problems and the user-friendly features make it more convenient in the long run. It understands the natural language question, searches in the dataset finds the most précised answers based on various machine learning algorithms, and presents it to the user in a suitable manner.

This paper describes this whole process in a detailed manner for a medical domain-restricted question-answer system [6].
This research [7] tackles the problem of big language models and knowledge expansions using graphs to develop a retrieval-based medical question-answering system.
While research from a variety of disciplines is included in the topic of question answering, only a few related fields—information extraction and retrieval—have a scope. We are living in the Information Age. We may get any information with the help of web search engines. We can visit the page from anywhere in the world. The goal of the query answering system is to determine the precise answer to a query given a set of documents [8]. Two main objectives of the question-answering system are to develop different issues and to enhance the computer interface. A specialized type of information retrieval (IR) called a question-answering system seeks knowledge in which we are interested in receiving precise responses to questions. Therefore, instead of offering whole documents or the best-matching sections, as the majority of information retrieval systems now do, the primary goal of all question-answering systems is to obtain answers to queries [9][10].

2. Question Pre-processing

The most complicated step of the Question Answering System is question classification [11]. Question pre-processing is an important stage in the pipeline of medical question-answering systems since it helps to refine and structure the input query so that the underlying artificial intelligence models can better understand and interpret it. The major goal of this step is to convert the user's raw inquiry into a format that Natural Language Processing (NLP) systems can understand [12]
Processing consists of tokenizing the Question, adding POS tagging, and parsing the natural language questioning with the appropriate parser. The final stage in pre-processing is to apply Named Entity Recognition (NER) [13].

Steps of Question Answering System: Creating a question-and-answer system consists of the following fundamental steps:

1. Categorization of similar types of questions.
2. Tokenizing and pre-processing the question.
3. Parsing the tokens to understand the meaning of the NL question.
4. Finding the relevant answers from the data set, repository, and information storage.
5. Extracting the most exact or accurate answer based on Machine learning algorithms and information extraction methods.
6. Presenting the chosen answer to the user in an appropriate suitable form using NLP Techniques.

Key Terminology in Question Pre-Processing:
1. **Tokenization:** Medical questions are split down into smaller parts called tokens, which might be words, phrases, or symbols.
example: "What are the symptoms of diabetes?" → ["What", "are", "the", "symptoms", "of", "diabetes", "?"]

2. **Stop Word Removal:** Commonly used words like "the," "is," "and," and "of" are deleted since they do not significantly add to comprehending the medical inquiry.
example: "What are the symptoms of Diabetes?" ["symptoms" and "diabetes"]
3. **Stemming and Lemmatization*:*** *Stemming* removes suffixes from words to reduce them to their base or root form (e.g., "diagnosing" → "diagnos").
- *Lemmatization i*s the process of mapping words to their dictionary base form, or "lemma"
example: "symptoms" → "symptom".

This helps in the standardization of words, which is especially important in medical language, where forms such as "symptom" and "symptoms" should be handled equally.

4. **Named Entity Recognition (NER):** The system recognizes essential medical entities in the query, such as illnesses, symptoms, treatments, and anatomical components.
example: in the question "What is the treatment for hypertension?" the system will detect "hypertension" as a medical condition.

5**. Spelling Correction and Synonym Expansion:** Ensures appropriate query interpretation by correcting spelling problems. For example, "dibetes" is changed to "diabetes."
- Synonyms are also detected and extended to offer comprehensive search results.
example: "hypertension" may be viewed with its substitute, "high blood pressure."

6. **Identifying Query Intent:** It's important to understand the intent behind the inquiry. The system must detect whether the user is looking for information on diagnosis, therapy, symptoms, or side effects.
example: "What is the treatment for diabetes?" identifies the intended treatment.

7. **Question Type Classification:** Medical QA systems categorize questions as factoids (seeking a specific fact), lists, definitions, and diagnoses.
example: "What causes migraines?" is classed as a factoid.

8. **Extracting Context and Temporal Information:** Some queries may include time-sensitive information, such as symptom onset or duration.
example: "How long does it take for the flu to go away?" → Temporal data (duration) is extracted.

9. **Syntactic Parsing:** This analyses phrase structure to understand word connections. Also aids in finding significant subject-verb-object links.
example: in "Can high cholesterol cause heart disease?" the syntactic structure links "high cholesterol" as the cause with "heart disease" as the outcome.

10. **Disambiguation:** Medical queries can sometimes be unclear. The system attempts to resolve misunderstandings by referring to context or asking clarifying questions.
example: the term "cold" can apply to either a low temperature or the common cold.

Explanation: The flow of the concept of pre-processing is explained with the help of the following example:
Input Query: "What are the common symptoms of type 2 diabetes?"
Pre-Processing Steps are:

1. Tokenization: ["What", "are", "the", "common", "symptoms", "of", "type", "2", "diabetes", "?"] 2. Stop Word Removal: ["common", "symptoms", "type", "2", "diabetes"] 3. Lemmatization: ["common", "symptom", "type", "2", "diabetes"] 4. Named Entity Recognition: Identifies "type 2 diabetes" as a medical condition and "symptoms" as a symptom-related query. 5. Synonym Expansion: Expands "diabetes" to include synonyms like "T2D." 6. Intent Identification: Detects intent as "symptom inquiry."

Through these steps, question pre-processing ensures that the query is structured, standardized, and ready for further analysis by the medical QA system, leading to more accurate and relevant responses.

3. Question Classification

A question-answering system must handle many big, complicated queries. Therefore, it's important to arrange these many different queries sensibly. This is accomplished by classifying the questions according to several standards. The development of the Question Answering System, is a highly significant and critical phase. Question classification, often called question categorization, is the act of grouping questions into several pre-established classifications. The wide range of questions that may be asked of a question-answer system can be expressed in several ways, which makes it extremely challenging and time-consuming to manually create rules that, if followed, will enable the system to respond to these massive amounts of questions [14] [15].

Let's take the query, "What is Diabetes?" as an example. Additionally, the user can express it in a broad range of ways and ask the system for it, such as:

• What is diabetes defined as?
• What is meant by diabetes?
• Describe what diabetes is.
• Could you explain what diabetes means?

Furthermore, there are other ways to phrase the papers, sections, or paragraphs that include the answers to any of these questions:

• Diabetes is a condition caused by ………….
• Diabetes can be defined as follows…………
• Diabetes is a condition where.........................

Therefore, the purpose of question classification is to categorize several questions with the same meaning.

This is also done to guarantee that the inquiry is the kind that gives a hint as to the kind of response that is expected. It is easier to find the precise response in the pertinent sections and paragraphs when you are aware of the sort of query. For instance, a date or time will be provided as a response to "when" inquiries. Likewise, a place or location may be mentioned in the response to inquiries that contain the "where" keyword. As a result, a question is initially categorized according to its nature and the words it employs, such as what, when, who, where, how, why, why not, etc.

Questions are often categorized according to particular linguistic patterns.

Since questions often adhere to particular linguistic patterns, they are categorized using taxonomies. There are primarily two types of taxonomies: hierarchical, which contains sublevels, and flat, which do not. Moldovan et al. presented a hierarchical taxonomy (Table 1) in the TREC-8 proceedings[16] that categorized the question types into nine groups, each of which had several subclasses.

After identifying the kind, a question is categorized based on the information it requests, and the goal is to group questions with similar meanings into a single category. For example, take the two queries, “What is medicine?” and “What is the medicine for diabetes?”.

Following the establishment of the taxonomy, queries are further categorized using the two primary methods of rule-based classifiers and machine-learning classifiers.

Table 1: Hierarchical Taxonomy [16]

Question Class	Question Sub-Classes	Answer Type
WHAT what-who what-when what-where	basic-what	Money / Number / Definition / Title / NNP / Undefined
WHO		Person / Organization
HOW how-many how-long how-much how-much <modifier> how-far how-tall how-rich how-large	basic-how	Manner Number Time / Distance Money / Price Undefined Distance Number Undefined Number
WHERE		Location
WHEN		Date
WHICH which-where which-when which-what	which-who	Person Location Date NNP / Organization
NAME name-where name-what	name-who	Person / Organization Location Title / NNP
WHY		Reason
WHOM		Person /

While machine learning classification can be challenging, rule-based classifiers categorize questions based on taxonomy by simply constructing a set of predetermined heuristic rules. The machine learning approach uses an interpreted corpus of labelled questions to create and train a machine learning model. The process will automatically extract the patterns that are beneficial for creative categorization from the corpus [17]. Here, training data is used to teach the system the notion of classification, while test data is used to evaluate the system's performance.

Additionally, supervised and unsupervised machine learning algorithms use distinct learning mechanisms. Even yet, the classifiers that are used to train the system for the syntactic and semantic aspects of linguistic analysis are crucial. Support Vector Machines (SVM), Nearest Neighbours (NN), Naïve Bayes (NB), Decision Trees (DT), and Sparse Networks of Winnows (SNOW) are a few examples of machine learning classifiers that offer varying degrees of classification performance and efficiency.

3.1. Formulation of queries:

After determining the kind and topic of the inquiries, they are next developed into queries. In order to create queries that must be fed into information retrieval systems, the focus and key phrases from the inquiries are utilized. The relevant information and paragraphs about the question and its keywords are recovered, making them even more prepared for the response extraction procedure.

4. Information Retrieval

Finding brief text passages on the internet or in another collection of documents is the aim of information retrieval-based question answering [7]. Fig. 1 illustrates the information retrieval system's operation.

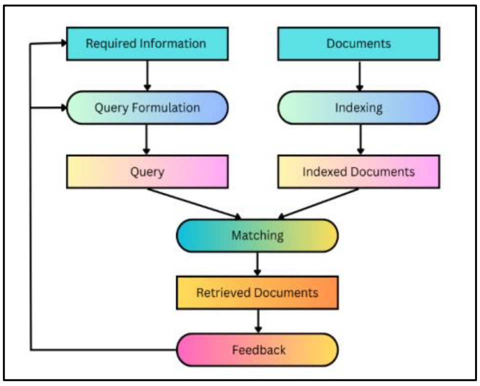

Fig 1: Information Retrieval Process

A corpus and dataset of the medical field must be gathered to build an automatic question-answer system. We gave the name MEDTALK to our medical-based question-answer system.

4.1. Corpus

Information about the medical question-answering system is gathered from the Clinical Question Answer System [CLINIQA] and websites. It provides Unique Medical Phrases taken from the knowledge base of CLINIQA [18]. CLINIQA is simple and in the English language. Its architecture is robust and applicable to any disease and knowledge source. Its response time is very fast.

4.2. Dataset

Using this gathered corpus, we create our database related to diabetes. However, this corpus contains noise, which we wish to eliminate to create a dataset. By converting the entire content to lowercase, eliminating punctuation, stop words, numerals, and excess whitespace, and filtering out undesired phrases, data mining performs data cleaning. Tokenization is another name for the data cleansing process. After that, the remaining words must be converted to floating-point or integer values. This is known as vectorization and is used as input for machine learning algorithms. Our main goals are counting words, determining their frequency, and converting data into distinct numbers that may be utilized for testing or training.

4.3. Training

We now have a dataset in the form of grids of integers called matrices. We need to train the dataset in order to fix the issue. We constructed a neural network, with a higher-level rule, for training. Every input is assigned a weight, which may be either positive or negative. We assign a random number to each weight before we begin.

Based on the Probability Ranking Principle, the probabilistic model is defined. The task of an information retrieval system is to assign a ranking to the document according to the likelihood that it is relevant to the query. The necessary data is given ambiguously. Numerous sources are used in probabilistic retrieval techniques, with statistical (graphical) distribution being the most often used. The link that connects the pieces guarantees the consistency of the system and offers methods for connecting models to data provided by probability theory. Both a data structure that naturally lends itself to the building of effective general-purpose algorithms and an intuitively appealing interface for modelling highly interacting sets of variables are provided by the graph theoretic side of graphical models [19].

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \ldots\ldots\ldots\ldots\ldots\ldots eqs(1)$$

The majority of directed graphs are Bayesian networks, where each node represents a hypothesis or a random process. The conditional probabilities are represented by the arrows connecting nodes, which show how information about one node's state affects the probability distribution of another related node. Assuming that the hypothesis is H and the evidence is E, the Bayes theorem states that the connection between the hypothesis's probability before receiving the evidence, P(H), and its probability after receiving the evidence, P(H|E), represents the hypothesis or a random process. The conditional probabilities are represented by the arrows connecting nodes, which show how knowledge of one node's state affects the probability distribution of another node that is related to it [19].

5. Work (Survey): In this, we divide our work into two parts: First to evaluate the performance of Model MedTalk in terms of F1 score, which measures the balance between precision (correctly predicted relevant items) and recall (all relevant items retrieved). that represents the percentage of user queries for which the MQAS provides the F1- score is 85 % and the second work is to collect user feedback on satisfaction.

The survey is conducted on 188 people to collect their views regarding the working of the medical chat box. The survey's main objective is to assess user satisfaction, usability, and the effectiveness of a medical chatbot in providing healthcare information or assistance. The participants are above 18 age general users. The findings revealed that a significant percentage of participants appreciated the chatbot's 24/7 availability and quick response times, making it a convenient tool for addressing basic medical queries. Many users found the interface intuitive and easy to navigate.

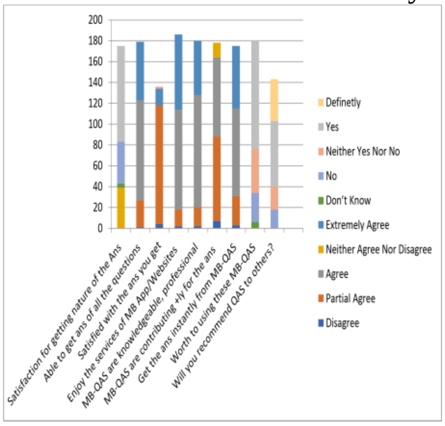

Fig 2: Graph of Survey

Column1		*Column2*		*Column3*		*Column4*		*Column5*		*Column6*		*Column7*		*Column8*		*Column9*	
Mean	3.133333	Mean	2.98333333	Mean	2.91666667	Mean	2.91667	Mean	3.1	Mean	3	Mean	2.966667	Mean	3	Mean	2.383333
Std Err	0.087225	Std Err	0.09355402	Std Err	0.12864278	Std Err	0.10707	Std Err	0.097134	Std Err	0.09205746	Std Err	0.123432	Std Err	0.1344585	Std Err	0.140804
Median	3	Median	3	Median	3	Median	3	Median	3	Median	3	Median	3	Median	3	Median	3
Mode	3	Mode	3	Mode	4	Mode	3	Mode	3	Mode	3	Mode	3	Mode	4	Mode	3
Std Dev	0.675646	Std Dev	0.7246663	Std Dev	0.99646267	Std Dev	0.82937	Std Dev	0.752397	Std Dev	0.71307403	Std Dev	0.956098	Std Dev	1.0415113	Std Dev	1.090664
Sample Var	0.456497	Sample Var	0.52514124	Sample Var	0.99293785	Sample Var	0.68785	Sample Var	0.566102	Sample Var	0.50847458	Sample Var	0.914124	Sample Var	1.0847458	Sample Var	1.189548
Kurtosis	1.979332	Kurtosis	-0.2612552	Kurtosis	-1.2597426	Kurtosis	-0.3418	Kurtosis	0.496094	Kurtosis	0.76025408	Kurtosis	- 0.183651	Kurtosis	- 0.9900536	Kurtosis	-1.335495
Skewness	-0.84792	Skewness	-0.2511296	Skewness	-0.2537548	Skewness	-0.394	Skewness	-0.66194	Skewness	-0.5802947	Skewness	- 0.774463	Skewness	-0.558704	Skewness	-0.014738
Range	3	Range	3	Range	3	Range	3	Range	3	Range	3	Range	3	Range	3	Range	3
Minimum	1	Minimum	1	Minimum	1	Minimum	1	Minimum	1	Minimum	1	Minimum	1	Minimum	1	Minimum	1
Maximum	4	Maximum	4	Maximum	4	Maximum	4	Maximum	4	Maximum	4	Maximum	4	Maximum	4	Maximum	4
Sum	188	Sum	179	Sum	175	Sum	175	Sum	186	Sum	180	Sum	178	Sum	180	Sum	143
Count	60	Count	60	Count	60	Count	60	Count	60	Count	60	Count	60	Count	60	Count	60

Table 2 a: Summary

However, some challenges were identified, such as occasional inaccuracies in responses, limited handling of complex medical terminology, and image-related data. Fig 2, represents the result of the survey while Table 2a -2b represents the statistical ANOVA test on survey data.

There is a statistically significant difference between groups if the P-value is smaller than the alpha; in this instance, the p-value of .000288 is less than .05. As a result, the test shows that the satisfaction ratings between the columns are statistically significant.

Table 2 b: Anova: Single Factor

Columns	Count	Sum	Average	Variance		
1	60	188	3.1333333	0.4565		
2	60	179	2.9833333	0.52514		
3	60	175	2.9166667	0.99294		
4	60	175	2.9166667	0.68785		
5	60	186	3.1	0.5661		
6	60	180	3	0.50847		
7	60	178	2.9666667	0.91412		
8	60	180	3	1.08475		
9	60	143	2.3833333	1.18955		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*Fcrit*
Between Groups	23	8	2.875	3.73623	0.000288	1.95583
Within Groups	408.6	531	0.7694915			
Total	431.6	539				

Comparative Study: Based on the existing literature [20] small comparative study is summarized in Table 3.

Table 3: Comparative Study of Medical-Based Question Answering System.

Year	Reference	Dataset	Tools	Techniques	Key Findings
2023	[21]	MIMIC-III, PubMed	GPT variants, T5-base	Fine-tuning, in-context learning	Fine-tuning improves task-specific performance; dynamic prompting enhances relevance.
2023	[22]	MIMIC-IV, UMLS	Retrieval-Augmented Generation	Weighted graph construction, tagging	Graph-based methods enhance domain alignment; scalable to large datasets.
2023	[23]	Visual datasets (MIMIC-CXR, CheXpert)	Vision-Language Models	Multimodal fusion, SWOT analysis	Vision and text integration improve QA; dataset quality is a key limitation.
2023	[24]	MIMIC-IV	BERT, Neo4j Knowledge Graph	Knowledge graph integration, semantic search	Effective for structured data; unstructured inputs pose scalability challenges.
2022	[25]	PubMed	BioBERT, PubMedBERT	Domain-specific pre-training	Outperforms generic models in biomedical NLP tasks; requires extensive computational resources.
2022	[26]	BioASQ, PubMed	Transformers, Word2Vec	Hybrid deep learning with	Superior for evidence-based QA; struggles with ambiguous queries.

				hierarchical ranking	
2021	[27]	MedlinePlus, UMLS	Ontology-based pipelines	Semantic search, NER	Ontology improves precision but lacks robustness for unstructured queries.
2020	[28]	SQuAD, BioBERT benchmark datasets	ClinicalBERT, PyTorch	Fine-tuning on medical contexts	Domain-specific tuning enhances accuracy for clinical queries.
2019	[29]	PubMed	BioBERT	Transfer learning	Effective for evidence-based clinical answers but limited coverage of recent literature.

Application of Medical-Based Question-Answering System: Table 4 represents the domain where the Medical-Based Question Answering System (MQAS) is used.

Table 4: Application of Medical-Based Question Answering System.

Application Domain	Description	Example Queries
Patient Support and Education	Provides health education, symptom checks, and medication guidance to patients.	"What is hypertension?"
Clinical Decision Support	Assists healthcare professionals in the diagnosis, treatment recommendations, and access guidelines.	"What are the guidelines for treating diabetes?"
Telemedicine and Virtual Care	Supports virtual consultations by resolving real-time queries and reducing doctor workload.	"How do I manage high cholesterol virtually?"
Integration with IoT Devices	Analyzes data from wearables and IoT devices, providing insights and preventive care advice.	"What does a high heart rate indicate?"
Medical Training and Education	Helps medical students learn, provides case-based guidance, and supports exam preparation.	"What is the pathophysiology of asthma?"
Pharmaceutical Applications	Answers drug-related queries, alerts about adverse reactions, and suggests safe drug combinations.	"Can ibuprofen be taken with aspirin?"
Clinical Research Support	Aids researchers by retrieving literature, analyzing data trends, and supporting grant applications.	"What is the prevalence of Type 1 diabetes?"
Emergency Response	Provides real-time instructions for first aid and guides triage in emergencies.	"How to perform CPR?"

Mental Health Support	Offers self-help resources and therapeutic guidance for mental well-being.	"What are some strategies for managing anxiety?"
Personalized Medicine	Delivers tailored advice based on individual health data and genetic profiles.	"What diet is suitable for high cholesterol?"
Health Insurance Queries	Answers questions related to policy coverage, claims, and pre-authorization for procedures.	"Does my insurance cover physiotherapy?"
Public Health Campaigns	Disseminates health alerts, outbreak information, and preventive measures to the public.	"What are the symptoms of the latest flu?"
Accessibility in Remote Areas	Bridges healthcare gaps by providing basic medical advice and multilingual support in underserved regions.	"What to do for a fever in a rural area?"

Challenges: However, everything is dependent on AI nowadays but still there are several challenges in MQAS. Table 5 represents the challenges of MQAS.

Table 5: Challenges of Medical based Question Answering System [4][5][8][30][31]

Category	**Challenges**
Data-Related	Incomplete datasets for training and testing.
	Ambiguity in medical language, including overlapping terms and abbreviations. Ex: "MI" could mean "Myocardial Infarction" or "Mitral Insufficiency."
	Poor performance on rare disease queries.
	Difficult to standardize and extract meaningful information. like clinical notes, research papers, and reports.
Model Development	Risk of inaccurate or incomplete answers, leading to potentially harmful medical guidance.
	Difficulty in interpreting complex multi-faceted queries. Ex: "What are the side effects of statins in diabetic patients?"
User Interaction	Handling vague or poorly structured queries. Ex: "Symptoms of heart?" may lack clarity or context.
	Handling multilingual and region-specific language variations.
Ethical and Legal	Incorrect answers may result in malpractice or harm.
	The system might perform poorly for underrepresented demographics.
	Compliance with data privacy regulations. Ex: Potential legal issues with handling sensitive patient data.
	Lack of explainability and transparency in model outputs.
Deployment	Ensuring real-time performance in resource-constrained environments.
	Integration with existing healthcare systems. Ex: EHR, IoT.
	Scalability to handle large volumes of queries simultaneously.
Technical	Handling multimodal data such as text, images (Ex: X-rays), and real-time IoT data.

	Dealing with out-of-domain queries for which the system lacks knowledge. Ex: "What is the latest treatment for a rare disease?"
Trust and Adoption	Gaining trust of healthcare professionals and patients.
	Resistance to adoption in clinical practice due to workflow disruptions.

6. **Conclusion:**

This survey report addresses the healthcare system's question-answering mechanism. It displays the findings of the brief survey together with the statistical test which is the ANOVA test. It covers how it works and its benefits. It also covers the creation of questions, pre-processing and categorization, query creation, information retrieval, and processing, and the need to increase the system's accuracy and efficiency in the future.

5. Reference:

[1] Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. Future Healthc J. 2021 Jul;8(2):e188-e194. doi: 10.7861/fhj.2021-0095. PMID: 34286183; PMCID: PMC8285156.
[2] Adam Bohr, Kaveh Memarzadeh, The rise of artificial intelligence in healthcare applications, Adam Bohr, Kaveh Memarzadeh, Artificial Intelligence in Healthcare, Academic Press, 2020, Pages 25-60, ISBN 9780128184387, https://doi.org/10.1016/B978-0-12-818438-7.00002-2.
[3] Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, Aldairem A, Alrashed M, Bin Saleh K, Badreldin HA, Al Yami MS, Al Harbi S, Albekairy AM. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ. 2023 Sep 22;23(1):689. doi: 10.1186/s12909-023-04698-z. PMID: 37740191; PMCID: PMC10517477.
[4] Pawar, D., Phansalkar, S., Sharma, A., Sahu, G. K., Ang, C. K., & Lim, W. H. (2023). Survey on the Biomedical Text Summarization Techniques with an Emphasis on Databases, Techniques, Semantic Approaches, Classification Techniques, and Similarity Measures. *Sustainability*, *15*(5), 4216. https://doi.org/10.3390/su15054216.
[4] Mutabazi, E., Ni, J., Tang, G., & Cao, W. (2021). A Review on Medical Textual Question Answering Systems Based on Deep Learning Approaches. *Applied Sciences*, *11*(12), 5456. https://doi.org/10.3390/app11125456
[6] Ojokoh, Bolanle & Adebisi, Emmanuel. (2019). A Review of Question Answering Systems. Journal of Web Engineering. 17. 717-758. 10.13052/jwe1540-9589.1785.
[7] Guo, Q., Cao, S., & Yi, Z. (2022). A medical question answering system using large language models and knowledge graphs. *International Journal of Intelligent Systems*, *37*(11), 8548-8564.
[8] Jerfy A, Selden O, Balkrishnan R. The Growing Impact of Natural Language Processing in Healthcare and Public Health. Inquiry. 2024 Jan-Dec;61:469580241290095. doi: 10.1177/00469580241290095. PMID: 39396164; PMCID: PMC11475376.
[9] Prashant Upadhyay, Rishabh Agarwal, Sumeet Dhiman, Abhinav Sarkar, Saumya Chaturvedi, A comprehensive survey on answer generation methods using NLP,
Natural Language Processing Journal, Volume 8, 2024, 100088, ISSN 2949-7191,
https://doi.org/10.1016/j.nlp.2024.100088.
[10] Chowdhary, K.R. (2020). Information Retrieval. In: Fundamentals of Artificial Intelligence. Springer, New Delhi. https://doi.org/10.1007/978-81-322-3972-7_18
[11] Daniel Jurafsky & James H. Martin. (2019). Question Answering, Speech and Language Processing. October 2, 2019.
[12] Kell, G., Roberts, A., Umansky, S., Qian, L., Ferrari, D., Soboczenski, F., ... & Marshall, I. J. (2024). Question answering systems for health professionals at the point of care—a systematic review. *Journal of the American Medical Informatics Association*, *31*(4), 1009-1024.
[13] Archana Goyal, Vishal Gupta, Manish Kumar, Recent Named Entity Recognition and Classification techniques: A systematic review, Computer Science Review, Volume 29, 2018, Pages 21-43, ISSN 1574-0137, https://doi.org/10.1016/j.cosrev.2018.06.001.

[14] Sarrouti, M., Lachkar, A., & Ouatik, S. E. A. (2015, November). Biomedical question types classification using syntactic and rule based approach. In *2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)* (Vol. 1, pp. 265-272). IEEE.
[15] Aarthi, D., Viswanathan, V., Nandhini, B., & Ilakiyaselvan, N. (2019). Question classification using a rule based model. *Int J Innov Technol Explor Eng*, *9*(1), 4172-4176.
[16] Dan Moldovan, Sanda Harabagiu, Marius Pasca, Rada Mihalcea, Richard Goodrum, Roxana Girju and Vasile Rus, "Lasso: A tool for Surfing the Answer Net" , Proceedings of TREC-8, Aug 1, 2000
[17] Ali Mohamed Nabil Allam and Mohamed Hassan Haggag, "Question Answering Systems: A Survey", International Journal of Research and Reviews in Information Sciences (IJRRIS) Vol. 2, No. 3, September 2012
[18] M.A.H. Zahid, Ankush Mittal, R. C. Joshi, Gowtham Atluri "CLINIQA: A Machine intelligence based clinical question answering system," IEEE Transaction on Information Technology in Biomedicine, (Communicated) 2006.
[19] George, C.P. (2010). A Realm-Based Question Answering System Using Probabilistic Modeling.
[20] Yudhono, E., Shaufiah, S., & Sardi, I. L. (2024). Application of a Question Answering System in the Ontology-Based Health Domain and Question Templates to Improve the Quality of Answers. *Journal La Multiapp*, *5*(5), 553-567.
[21] Yagnik, N., Jhaveri, J., Sharma, V., Pila, G., Ben, A., & Shang, J. (2024). MedLM: Exploring Language Models for Medical Question Answering Systems. *arXiv preprint arXiv:2401.11389*.
[22] Wu, J., Zhu, J., Qi, Y., Chen, J., Xu, M., Menolascina, F., & Grau, V. (2024). Medical graph rag: Towards safe medical large language model via graph retrieval-augmented generation. *arXiv preprint arXiv:2408.04187*.
[23] Al-Hadhrami, S., Menai, M. E. B., Al-Ahmadi, S., & Alnafessah, A. (2023). A critical analysis of benchmarks, techniques, and models in medical visual question answering. *IEEE Access*, *11*, 136507-136540.
[24] Ozonze, O., Scott, P.J. & Hopgood, A.A. Automating Electronic Health Record Data Quality Assessment. *J Med Syst* **47**, 23 (2023). https://doi.org/10.1007/s10916-022-01892-2
[25] Tung, N. Y., Hu, H. W., Chang, T. W., Hu, Y. M., & Lin, H. M. (2022, May). Multi-model Comparison for Classification of Medical Records using the BioBERT Models. In *2022 IEEE 4th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)* (pp. 221-224). IEEE.
[26] Nentidis, A., Katsimpras, G., Vandorou, E., Krithara, A., Miranda-Escalada, A., Gasco, L., ... & Paliouras, G. (2022, August). Overview of BioASQ 2022: the tenth BioASQ challenge on large-scale biomedical semantic indexing and question answering. In *International conference of the cross-language evaluation forum for European languages* (pp. 337-361). Cham: Springer International Publishing.
[27] Jing X, The Unified Medical Language System at 30 Years and How It Is Used and Published: Systematic Review and Content Analysis, JMIR Med Inform 2021; 9(8): e 20675
URL: https://medinform.jmir.org/2021/8/e20675, DOI: 10.2196/20675
[28]Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, Jaewoo Kang, BioBERT: a pre-trained biomedical language representation model for biomedical text mining, *Bioinformatics*, Volume 36, Issue 4, February 2020, Pages 1234–1240, https://doi.org/10.1093/bioinformatics/btz682
[29] Jin, Q., Dhingra, B., Liu, Z., Cohen, W. W., & Lu, X. (2019). Pubmedqa: A dataset for biomedical research question answering. *arXiv preprint arXiv:1909.06146*.
[30] Adnan K, Akbar R. Limitations of information extraction methods and techniques for heterogeneous unstructured big data. International Journal of Engineering Business Management. 2019;11. doi:10.1177/1847979019890771
[31] Shelke, B. A., & Mahender, C. N. Challenges For Development Of Question Answering Systems In Indian Languages.