

AI-DRIVEN PUBLIC HEALTH CHATBOT FOR DISEASE AWARENESS

Mrs. Ch. Pravalika¹, Abhijeet Varma², Harsh Gaur³, Kammari Rishik⁴, Jalapally Varshith⁵,
Mohammad Eshal⁶

¹ Assistant Professor, Department of Computer Science and Engineering, Hyderabad Institute of Technology and Management, Telangana, India

²³⁴⁵⁶ Department of Computer Science and Engineering, Hyderabad Institute of Technology and Management, Telangana, India

1. Abstract-Public health systems across the world continue to face challenges in delivering timely, accurate, and accessible disease-related information to the general population. Misinformation, delayed awareness, and limited access to healthcare professionals often worsen the impact of outbreaks and common illnesses, particularly in underserved and rural communities. The AI-Driven Public Health Chatbot for Disease Awareness addresses this gap by offering an intelligent, conversational system that provides users with instant, reliable guidance on disease symptoms, prevention measures, and treatment options. The proposed system leverages Natural Language Processing (NLP) and Machine Learning techniques, grounded through Retrieval-Augmented Generation (RAG) on verified sources such as WHO, ICMR and MoHFW guidelines, to understand user queries expressed in everyday language and respond with medically verified, safety-constrained information.

Keywords-Development of an AI-Driven Public Health Chatbot to provide timely, accurate, and accessible disease-related information and combat misinformation, especially in underserved/rural areas, using NLP + RAG grounded on WHO/ICMR/MoHFW verified guidelines.

2. INTRODUCTION

Access to reliable health information remains uneven across the world, particularly in rural and semi-urban areas where trained medical professionals and verified information sources may be limited. In the absence of easily accessible and credible guidance, people often depend on unverified sources such as social media, online forums, and general search results, which can contribute to the spread of health misinformation. The rapid development of Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP) provides an opportunity to improve access to healthcare information through intelligent conversational systems. An AI-based healthcare information assistant can understand questions expressed in natural language and provide users with relevant information about common symptoms, diseases, preventive measures, general health practices, and healthcare resources in a simple and understandable manner.

2.1 Background of Public Health Awareness

Access to reliable health information is an important part of public health. People regularly search for information about diseases, symptoms, prevention methods, and general healthcare practices. However, the availability of information does not always mean that the information is reliable or easy to understand. Users may come across content from social media, online forums, blogs, or general search results without knowing whether the information is medically trustworthy. This can become particularly problematic when people are trying to understand a disease or decide whether a health concern requires professional attention. The problem can be more noticeable in rural and semi-urban communities where access to trained healthcare professionals and verified information may be limited. In such situations, people may depend heavily on easily available online information. A system that can provide basic disease-awareness information in a simple conversational format can therefore be useful as an initial source of health education. The purpose of such a system is not to replace healthcare professionals, but to make general and reliable health information easier to access.

2.2 Challenges in Accessing Reliable Health Information

One of the major challenges in public health communication is the difference between available information and reliable information. A person can find thousands of search results for a disease within seconds, but determining which information is accurate can be difficult for someone without a medical background. Health information can also be difficult to understand because many medical resources use technical terminology. A user looking for a simple explanation may therefore struggle to interpret information that was originally written for healthcare professionals or academic purposes. Another issue is the spread of misinformation. Social media platforms and online communities allow health-related claims to spread rapidly, sometimes without appropriate verification. When users rely on such information, incorrect beliefs about symptoms, prevention, or treatment may develop.

2.3 Role of Artificial Intelligence in Healthcare

Artificial Intelligence has increasingly been applied to different areas of healthcare. Machine Learning can be used to identify patterns in data, while NLP enables systems to process and interpret human language. Together, these technologies can support applications such as information retrieval, clinical decision support, patient communication, and health education. For public health awareness, AI can be particularly useful when users need information in a conversational format. A conversational system can process the question and present the relevant information in a more direct format. However, healthcare applications require greater care than ordinary question-answering systems. An incorrect response can potentially lead to misunderstanding or inappropriate decisions. Therefore, an AI-based health-information system should place importance on reliable information sources, appropriate response generation, and clear limitations.

2.4 Role of Natural Language Processing and Conversational AI

Natural Language Processing is an important component of the proposed chatbot because users are not expected to communicate using predefined commands. They should be able to express their questions naturally. For example, the following questions have a similar intention:

- "What are the symptoms of dengue?"
- "How do I know if someone has dengue?"
- "Tell me the common signs of dengue."

A conversational AI system should be able to recognize that these questions are related to the same general topic. This type of interaction can make health-information systems easier to use, particularly for users who may not be familiar with technical terminology or complicated interfaces.

2.5 Retrieval-Augmented Generation for Reliable Responses

A significant concern with generative AI systems is that a language model may generate information that sounds convincing but is not necessarily supported by a reliable source. This is especially important in healthcare applications. The proposed system therefore incorporates Retrieval-Augmented Generation (RAG). Instead of relying only on the language model's previously learned information, the system can retrieve relevant information from its selected knowledge sources and use that information while generating a response. This approach provides a connection between conversational AI and a controlled collection of health information. In the proposed project, the retrieved information is intended to be grounded in verified sources such as WHO, ICMR, and

MoHFW guidelines. The basic process can be represented as:

User Query → Query Processing → Relevant Information Retrieval → AI Response Generation → User Response

2.6 Motivation of the Proposed Work

The motivation for this project comes from the need to make basic disease-awareness information easier to access and understand. Existing online information is extensive, but users may still have difficulty identifying reliable sources or understanding medical terminology. The proposed chatbot attempts to address this issue through a simple conversational interface. Rather than requiring users to search through multiple sources, the system provides an interface through which they can ask health-related questions directly. At the same time, the project recognizes that an AI chatbot should not be treated as a replacement for a medical professional. Its intended purpose is disease awareness and general health information, while situations requiring diagnosis or professional medical advice should be directed toward qualified healthcare professionals.

3. LITERATURE REVIEW

3.1 Artificial Intelligence in Healthcare

The application of Artificial Intelligence in healthcare has expanded from traditional machine-learning systems to systems capable of processing text, images, and natural-language queries. Machine Learning techniques can identify patterns from large datasets, while Deep Learning has enabled more complex forms of automated analysis. For health-information applications, NLP is particularly relevant because much of the information exchanged between patients and healthcare systems is expressed through natural language. AI-based systems can process questions and organize relevant information in a form that is easier for users to understand. However, healthcare AI also introduces challenges related to accuracy, reliability, privacy, and safety. A model that performs well in a general conversational setting may not necessarily be appropriate for providing health information without additional safeguards. This makes the selection of reliable information sources and appropriate system constraints important parts of healthcare chatbot development.

3.2 Healthcare Chatbots

Healthcare chatbots are conversational systems developed to interact with users through text or voice. Early chatbot systems generally relied on predefined rules and responses. Although rule-based systems can provide predictable outputs, they are limited when users phrase questions in unexpected ways. The development of NLP

and Machine Learning has allowed newer healthcare chatbots to handle more flexible user input. Such systems can support activities including health education, frequently asked questions, symptom-related information, appointment assistance, and general healthcare guidance. For disease-awareness applications, a chatbot can provide information in a conversational format rather than requiring users to navigate through a conventional website. This can reduce the effort required to locate basic information. At the same time, healthcare chatbots need to clearly communicate their limitations. Providing general information and diagnosing an individual are two different tasks. A disease-awareness chatbot should therefore be designed around information and education rather than presenting itself as a replacement for clinical assessment.

3.3 Natural Language Processing in Healthcare Applications

NLP provides the foundation for understanding user questions in conversational healthcare applications. Instead of treating every query as an exact keyword match, an NLP-based system can analyse the meaning and context of the input. For example, a user may ask:

"What should I look for if I think someone has dengue?"

The system needs to recognize that the user is asking about symptoms or warning signs, even though the question does not contain a formal phrase such as "symptoms of dengue." This ability to handle natural variations in language is important for public-facing applications because users have different levels of technical and medical knowledge.

3.4 AI-Based Disease Awareness Systems

Disease-awareness systems generally focus on making information about specific diseases more accessible to users. These systems may provide details such as symptoms, causes, risk factors, prevention, and general management. A limitation of many information systems is that users have to search for the required information themselves. A conversational system changes this interaction by allowing the user to ask a question and receive a response based on the available information. The proposed project follows this conversational approach while focusing specifically on public health and disease awareness rather than clinical diagnosis.

3.5 Retrieval-Augmented Generation in Healthcare

Retrieval-Augmented Generation combines information retrieval with generative AI. In a typical RAG system, the user's query is first processed to identify relevant information. The retrieved content is then provided to a

language model, which uses it to generate the final response. For a healthcare chatbot, this approach is useful because the response can be connected to a defined collection of health information instead of depending entirely on the language model's internal knowledge. The proposed project uses RAG with the intention of grounding responses in selected healthcare resources, including WHO, ICMR, and MoHFW guidelines. The approach can be summarized as:

Query → Retrieval → Relevant Health Information → Response Generation

3.6 Large Language Models and Conversational Health Information

Modern language models have significantly improved the ability of computer systems to generate natural and context-aware responses. Their conversational capabilities make them suitable for interfaces where users may ask follow-up questions or use informal language. However, language generation alone does not guarantee factual correctness. In healthcare, a response that is grammatically correct and convincing can still be problematic if it contains unsupported information. For this reason, the proposed project does not treat conversational ability as the only requirement. The system combines conversational processing with retrieval of information from selected sources. This is intended to make the chatbot more suitable for its specific purpose of disease awareness.

3.7 Verified Sources for Public Health Information

The reliability of the information supplied to a health chatbot depends significantly on the sources used by the system. The proposed project identifies organizations such as the World Health Organization (WHO), Indian Council of Medical Research (ICMR), and Ministry of Health and Family Welfare (MoHFW) as sources for verified healthcare guidance. Using recognized health organizations as information sources can provide a stronger foundation for the chatbot than relying on unrestricted web content. It also provides a more controlled basis for constructing the knowledge base used by the retrieval component.

3.8 Comparative Analysis

The main approaches relevant to this project can be summarized as follows:

Year	Approach	Main Contribution
2016	Healthcare Chatbots	Introduced conversational systems for healthcare-related user interaction
2018	NLP-Based	Improved understanding of

	Chatbots	natural-language healthcare queries
2020	AI Conversational Agents	Expanded conversational AI toward healthcare interventions and user support
2022	Transformer-Based Language Models	Improved natural-language understanding and response generation
2024	Healthcare LLM Evaluation	Systematically examined LLM applications, reliability, bias, and privacy concerns
2025-2026	RAG-Based Healthcare AI	Improved grounding of generated responses using retrieved domain-specific information

3.9 Research Gap

The literature and existing approaches indicate that healthcare chatbots can improve the accessibility of health information, but several issues remain important. First, a conversational system needs to understand how ordinary users phrase their questions. Second, generative models can produce fluent responses without guaranteeing that every statement is supported by a reliable source. Third, public health applications need to maintain a clear distinction between *health awareness* and *medical diagnosis*. The proposed project addresses these areas by combining NLP-based conversational interaction with Retrieval-Augmented Generation and selected verified health-information sources. The system is intended to provide disease-related information in a simple conversational format while maintaining safety constraints and recognizing that professional medical advice may still be required.

4. PROBLEM STATEMENT AND OBJECTIVES

4.1 Problem Statement

Access to reliable and understandable disease-related information remains a challenge for many people. Although a large amount of health information is available online, users may have difficulty identifying trustworthy sources, understanding medical terminology, or finding the specific information they need. The problem can be more significant in areas where access to healthcare professionals and verified information resources is limited. At the same time, the rapid growth of generative AI provides an opportunity to create conversational systems capable of responding to users' questions in natural language. However, a general-purpose AI system may generate responses that are not sufficiently grounded in verified health information. Therefore, there is a need for an AI-based public health chatbot that can understand

natural-language questions and provide accessible disease-awareness information while grounding its responses in selected and reliable healthcare sources. The proposed system addresses this problem by combining Natural Language Processing, Machine Learning, and Retrieval-Augmented Generation to provide information related to disease symptoms, prevention, and general treatment-related guidance.

4.2 Research Objectives

The main objectives of the proposed AI-Driven Public Health Chatbot for Disease Awareness are:

1. To develop an AI-driven conversational system that provides users with accessible information related to diseases and public health.
2. To enable natural-language interaction so that users can ask health-related questions using ordinary language.
3. To provide disease-awareness information including symptoms, preventive measures, and general treatment-related information.
4. To integrate Natural Language Processing techniques for understanding and processing user queries.
5. To use reliable healthcare information sources, including relevant guidance from organizations such as WHO, ICMR, and MoHFW.
6. To evaluate the proposed chatbot in terms of response relevance, information quality, usability, and its ability to handle common disease-awareness queries.
7. To identify possible improvements for future deployment.

4.3 Scope of the Proposed System

The scope of the proposed system is focused on **public health education and disease awareness**. The chatbot is intended to provide users with general information through a conversational interface.

The system covers areas such as:

1. Disease descriptions
2. Common symptoms
3. Causes and risk factors where applicable
4. Preventive measures
5. General health practices
6. Public-health information from selected sources

The scope can be represented as:

User Query → NLP Processing → Health Information Retrieval → RAG → AI Response → User

4.4 Research Questions

The research is guided by the following questions:

RQ1: Can an AI-driven conversational system provide accessible and relevant information for common disease-awareness queries?

RQ2: Can NLP techniques enable the chatbot to understand health-related questions expressed in everyday language?

RQ3: Can Retrieval-Augmented Generation improve the grounding of chatbot responses in verified healthcare information?

RQ4: How effectively can the proposed system provide information about disease symptoms, prevention, and general treatment-related guidance?

RQ5: What are the major limitations and safety considerations associated with using an AI-based chatbot for public health awareness?

5. PROPOSED AI-DRIVEN PUBLIC HEALTH CHATBOT

5.1 Overview of the Proposed System

The proposed **AI-Driven Public Health Chatbot for Disease Awareness** is a conversational system designed to provide users with easily understandable information related to diseases and general public health. The system allows users to enter questions in natural language instead of requiring them to search through multiple websites or use predefined commands. The chatbot processes the user's query and identifies the information being requested. Relevant information is then retrieved from the knowledge sources available to the system. The retrieved information is provided to the AI component, which generates a response that is presented to the user through the chatbot interface. The proposed system focuses mainly on disease awareness. It can provide information about symptoms, prevention, causes, risk factors, and general treatment-related information. The overall workflow of the proposed system can be represented as:

User Query → Query Processing → Information Retrieval → AI Response Generation → Safety Validation → Final Response

5.2 System Architecture

The architecture of the proposed chatbot consists of several interconnected components. Each component performs a specific task in converting the user's question into an informative response. The major components are:

1. User Interface
2. Query Processing Module
3. Natural Language Processing Module
4. Knowledge Retrieval Module

5. Healthcare Knowledge Base
6. AI Response Generation Module
7. Response Display Module

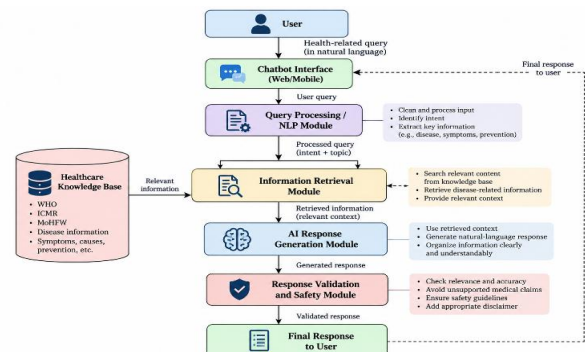


Fig. 1. System Architecture

5.3 User Interaction Module

The user interaction module provides the main interface through which users communicate with the chatbot. A user can enter a health-related question using natural language. For example:

"What are the common symptoms of dengue?"

The system receives the query and passes it to the processing module. The interface should be simple enough that users do not need prior technical knowledge. The purpose is to reduce the effort required to obtain basic disease-awareness information. The interface can also provide appropriate warnings to users that the information supplied by the chatbot is intended for awareness and educational purposes and should not be considered a medical diagnosis.

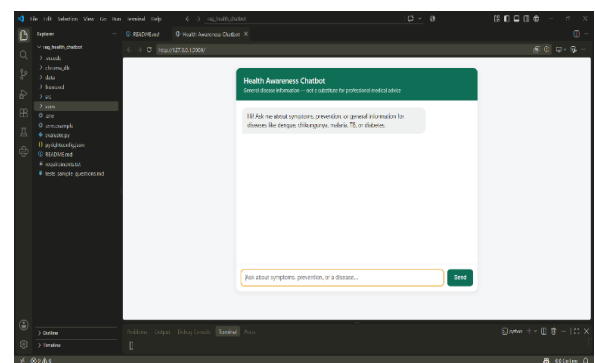


Fig. 2. User Interface+

5.4 Natural Language Processing Module

The NLP module is responsible for processing the user's natural-language input. Users may ask the same question using different expressions. For example:

1. "What are dengue symptoms?"
2. "Tell me the signs of dengue."
3. "How can I identify dengue symptoms?"

Although the wording is different, all three questions are related to disease symptoms. The NLP component helps the system process such variations and determine the user's information requirement. The processed query is then passed to the retrieval component so that relevant information can be identified.

5.5 Disease Information Retrieval Module

The information retrieval module is responsible for identifying information relevant to the user's question. Instead of generating every response entirely from the AI model, the system retrieves relevant information from its available healthcare knowledge sources.

User Query → Query Processing → Information Retrieval → AI Response Generation → Final Response

This approach helps connect the chatbot's response to information contained in its knowledge base.

5.6 Healthcare Knowledge Base

The healthcare knowledge base contains the information used by the chatbot to answer disease-awareness questions. The proposed project identifies reliable organizations such as: World Health Organization (WHO), Indian Council of Medical Research (ICMR), Ministry of Health and Family Welfare (MoHFW) as important sources of healthcare information.

The knowledge base can contain information such as:

1. Disease description
2. Symptoms
3. Causes
4. Risk factors
5. Prevention
6. General treatment information
7. Health precautions
8. Warning signs

The quality of the chatbot depends partly on the quality, accuracy, and maintenance of this knowledge base.

5.7 Retrieval-Augmented Generation Module

Retrieval-Augmented Generation is an important component of the proposed system. In a RAG-based approach, relevant information is first retrieved from an external knowledge source and then supplied to the generative model.

The process can be represented as:

1. User Query

2. Query Processing
3. Relevant Documents / Information
4. Retrieved Context
5. AI Model
6. Generated Response

This approach is particularly useful for the proposed project because health information should ideally be connected to reliable sources rather than relying entirely on the model's internal knowledge.

5.8 AI Response Generation Module

After relevant information has been retrieved, the AI component generates the final response. The purpose of this component is not simply to copy retrieved information. Instead, the information can be organized into a response that is easier for the user to understand. For example, instead of presenting a long technical document, the chatbot can provide:

Symptoms of Dengue

1. High fever
2. Headache
3. Muscle and joint pain
4. Nausea
5. Skin rash

6. METHODOLOGY

6.1 Methodology Overview

The proposed system follows a sequence of stages beginning with collection and organization of healthcare information and ending with evaluation of chatbot responses. The major stages are:

1. Data Collection
2. Preprocessing
3. Knowledge Base
4. Query Processing
5. Retrieval
6. Response Generation
7. Validation
8. Output

6.2 Data and Knowledge Sources

The chatbot requires a collection of reliable health information to answer user questions. The proposed project focuses on information obtained from recognized healthcare organizations. WHO, ICMR, and MoHFW are identified as important sources in the project description. The information collected should be relevant to the diseases and health topics supported by the chatbot. Each

information source should be reviewed before being incorporated into the knowledge base.

6.3 Data Collection

The data collection stage involves gathering disease-related information from the selected sources. Information can be organized according to categories such as:

Category	Example Information
Disease	Dengue
Description	General description of the disease
Symptoms	Common symptoms
Causes	Factors associated with the disease
Prevention	Preventive measures
Risk factors	Conditions that increase risk
General treatment	General management information
Warning signs	Situations requiring medical attention

6.4 Data Preprocessing

The collected information may contain unnecessary formatting, repeated information, long paragraphs, or technical language that is not directly useful for the chatbot. Preprocessing is therefore performed to organize the information into a form suitable for retrieval. The preprocessing stage can include:

1. Removing irrelevant content
2. Removing duplicate information
3. Organizing information by disease
4. Separating information into meaningful sections
5. Cleaning textual content
6. Creating smaller information units for retrieval

The objective is to ensure that the retrieval system can identify relevant information efficiently when a user submits a query.

6.5 Knowledge Base Construction

After preprocessing, the information is organized into the chatbot's knowledge base. A well-structured knowledge base allows the system to retrieve relevant information based on the user's query. Example:

Disease: Malaria

1. Description
2. Causes
3. Symptoms
4. Risk Factors
5. Prevention
6. General Treatment
7. Warning Signs

6.6 User Query Processing

When a user submits a question, the system first processes the input. For example:
"How can I prevent malaria?"

The system needs to determine:

1. **Disease:** Malaria
2. **Intent:** Prevention
3. **Required information:** Preventive measures

The processed query is then used by the retrieval component.

6.7 Response Generation

The retrieved information is supplied to the AI response-generation component. The AI model uses the retrieved context to formulate a natural-language response.

The response should be:

1. Relevant to the question
2. Easy to understand
3. Concise where possible
4. Grounded in retrieved information
5. Appropriate for general health awareness

The objective is to make the information useful to a general user without unnecessarily using complex medical terminology.

7. SYSTEM IMPLEMENTATION

7.1 Experimental Setup

The chatbot was evaluated using a collection of representative health-related queries covering different categories of disease-awareness information. The test queries included questions related to disease definitions, symptoms, causes, prevention, risk factors, and general treatment-related information.

7.2 Training and Validation Results

The chatbot was tested with different types of natural-language queries. During testing, the system first

processed the user's query and generated an embedding using the all-MiniLM-L6-v2 Sentence Transformer model. The resulting embedding was compared with the information stored in ChromaDB. Relevant information was retrieved and supplied as context to the LLM through the LangChain RAG pipeline. For example, when a user asks:

"What are the common symptoms of dengue?"

the system identifies the disease and retrieves relevant information related to dengue symptoms. The retrieved information is then provided to the LLM, which generates a conversational response. This process allows the chatbot to respond to questions that may be expressed using different wording while still retrieving semantically related information.

7.3 Accuracy Analysis

The final testing accuracy of FineKrop was:

Testing Accuracy: 88.17%

The obtained accuracy indicates the percentage of test images correctly classified by the model. However, accuracy alone is not sufficient to evaluate a multi-class disease detection system. Therefore, precision, recall, F1-score, and confusion matrix results should also be considered. The result should also be compared carefully with previous research. Direct comparison is meaningful only when datasets, classes, testing procedures, and experimental conditions are sufficiently similar.

7.4 Sample Query Results

The following table can be included in the research paper to demonstrate representative chatbot interactions.

Test Case	User Query	System Response
TC01	What is dengue?	Generated response explaining dengue
TC02	What are the symptoms of dengue?	Symptoms retrieved and explained
TC03	How can dengue be prevented?	Prevention information provided
TC04	What causes malaria?	Relevant malaria information retrieved
TC05	What are the symptoms of	Symptom-related response

	malaria?	generated
TC06	How can malaria be prevented?	Preventive measures explained
TC07	What are the risk factors for diabetes?	Relevant information provided
TC08	When should I see a doctor?	Appropriate guidance and disclaimer

7.5 Response Time Analysis

Response time is another factor that can be considered when evaluating the proposed chatbot. The total response time includes the time required for:

1. Receiving the user query.
2. Generating the query embedding.
3. Searching the ChromaDB vector database.
4. Retrieving relevant information.
5. Sending the context to the LLM.
6. Generating the final response.
7. Displaying the response through Streamlit.

The average response time can be calculated using:

Average Response Time = Total Response Time / Number of Test Queries

The measured response time can be reported in seconds for a defined set of test queries.

7.6 Comparison with Existing Methods

The development of healthcare conversational systems has progressed from traditional rule-based chatbots to NLP-based systems, transformer-based language models, generative AI, and retrieval-augmented systems.

A comparison of these approaches is shown below.

Approach	Interaction	Limitation
Rule-Based Chatbot	Predefined questions	Limited flexibility
NLP-Based Chatbot	Natural-language queries	Limited generative capability
Generative AI Chatbot	Natural conversation	May generate unsupported information
LLM Healthcare System	Natural-language interaction	Reliability and safety concerns
RAG-Based	Natural-language	Requires reliable

Chatbot	interaction	knowledge sources
Proposed System	Natural-language interaction	Requires continuous knowledge-base maintenance and evaluation

7.7 Discussion of Results

The evaluation of the proposed chatbot should focus on whether the system can provide relevant and understandable information for common disease-awareness queries.

The use of the *all-MiniLM-L6-v2* embedding model allows the system to perform semantic retrieval rather than depending only on exact keyword matching. ChromaDB provides the vector storage and similarity-search functionality, while LangChain connects the retrieval process with the LLM.

The RAG architecture also provides an important distinction from a purely generative chatbot. Rather than asking the LLM to independently generate health information, the system provides retrieved information as context during response generation.

However, the results should not be interpreted as evidence that the chatbot can diagnose diseases or replace medical professionals. The system is designed specifically for health awareness and information support.

8. FUTURE SCOPE

8.1 Real-Time Health Information Updates\

The current system can be extended to support automatic updating of the healthcare knowledge base.

Future versions could periodically retrieve updated information from selected official healthcare sources. This would help the chatbot provide information that reflects newer public-health guidelines and recommendations.

8.2 Mobile Application Integration

The chatbot can be developed as a dedicated Android or cross-platform mobile application. A mobile version could make the system more accessible to users who primarily access online services through smartphones.

Potential features include:

1. Mobile chatbot interface
2. Voice-based interaction
3. Disease information search

4. Conversation history
5. Health-awareness notifications
6. Offline access to selected information

8.3 Expanded Disease Coverage

The current system can be extended to cover a larger number of diseases and health conditions. Future versions could include:

1. Infectious diseases
2. Non-communicable diseases
3. Seasonal diseases
4. Vector-borne diseases
5. Lifestyle-related health conditions
6. Common public-health concerns

Expanding the knowledge base would allow the chatbot to respond to a wider range of user queries.

8.4 Multilingual Support

Multilingual support is an important area for future development. Many users may prefer to receive health information in their regional language rather than English. Future versions of the chatbot could therefore support languages such as:

1. Telugu
2. Hindi
3. Tamil
4. Kannada
5. Malayalam

The system could allow the user to select a preferred language and generate the response accordingly. This feature could improve accessibility, particularly for users who are more comfortable communicating in regional languages.

8.5 Explainable AI and Source References

Future versions of the chatbot could provide source references alongside generated answers. For example:

Information Source: World Health Organization

This would allow users to understand where the information originated. The system could also provide the relevant document or source section used to generate the response. This would improve transparency and allow users to verify important health information independently.

9. CONCLUSION

Access to reliable and understandable health information is an important part of disease awareness and public

health. Although a large amount of healthcare information is available online, users may find it difficult to identify reliable sources and obtain information in a simple and conversational manner.

This research presented an AI-Driven Public Health Chatbot for Disease Awareness that combines Natural Language Processing, Retrieval-Augmented Generation, semantic embeddings, vector-based retrieval, and Large Language Models to provide conversational health information.

The RAG architecture allows the system to retrieve relevant information from its healthcare knowledge base before generating a response. This provides a more controlled approach to health-information generation than relying entirely on the language model's internal knowledge.

The proposed chatbot is designed to provide information related to areas such as disease descriptions, symptoms, causes, prevention, risk factors, and general treatment-related information. The system is intended primarily for public health awareness and educational purposes.

REFERENCES

1. World Health Organization, "Digital Health Guidelines," WHO Press, 2021.
2. R. Sharma et al., "Chatbot-Based Symptom Checkers: A Survey," Journal of Health Informatics, 2021.
3. A. Kumar and S. Rao, "NLP Techniques for Conversational Health Assistants," IEEE Access, 2022.
4. P. Singh et al., "AI in Public Health Communication," Intl. Journal of Medical Informatics, 2023.
5. National Health Portal of India, "Disease Awareness Resources," 2023.
6. World Health Organization, Ethics and Governance of Artificial Intelligence for Health: WHO Guidance, World Health Organization, Geneva, 2021.
7. E. Tabassi, Artificial Intelligence Risk Management Framework (AI RMF 1.0), National Institute of Standards and Technology, NIST AI 100-1, 2023.
8. Y. Li et al., "Feasibility and effectiveness of artificial intelligence-driven conversational agents in healthcare interventions: A systematic review of randomized controlled trials," International Journal of Nursing Studies, vol. 143, p. 104494, 2023.
9. L. Laranjo et al., "The effectiveness of artificial intelligence conversational agents in health care: Systematic review," Journal of Medical Internet Research, 2021.
10. "A systematic review of AI-based chatbot usages in healthcare services," Journal of Health Organization and Management, 2025.