VOLUME 17 SPECIAL ISSUE 1, JULY 2026
nternational Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

AN INTELLIGENT CHATBOT FOR PUBLIC HEALTH AWARENESS USING ARTIFICIAL INTELLIGENCE

Dr. Jyoti Kaushal
Associate Professor, CSE
Geetanjali Institute of Technical Studies
Udaipur, India
jyoti.kaushal@gits.ac.in

Nehal Jain
Dept. of Computer Science-AI
Geetanjali Institute of Technical Studies
Udaipur, India
nehaljain1012@gmail.com

Mustansir Gumani
Dept. of Computer Science-AI
Geetanjali Institute of Technical Studies
Udaipur, India
mustansirgumani5253786@gmail.com

Abstract: The problem is that public health awareness still lags in underserved areas. It seems hard to ignore how gaps in access affect real outcomes. And we built an AI chatbot that delivers accurate health info and guidance in English and Hindi. The system mixes rule-based logic with machine learning to interpret user questions. It draws from verified FAQs and trusted medical texts for its knowledge base. Structure includes language models, data pipelines, and evaluation tracks for comprehension and engagement. Our work combines evidence-based content with modern AI tools. This improves reach for rural users without sacrificing accuracy or privacy. Ethical concerns like data safety and truthfulness are addressed directly. The roadmap outlines steps toward a scalable, multilingual health assistant.

Keywords: AI chatbot, public health, health awareness, NLP, multilingual health, rural healthcare, conversational agent, digital health.

I. INTRODUCTION

At least in theory, public health issues like preventable illnesses, poor health literacy, and false info still plague rural and semi-urban areas. Pamphlets and websites rarely grab attention, so interactive tools are needed now. Smartphones are widespread, and AI has advanced, chatbots can give 24/7 health advice without high costs.

We're building a multilingual chatbot that handles questions in hindi or English and gives reliable info on diseases, prevention, and how to access care. It uses natural language processing and trusted knowledge sources to answer things static materials miss. Evidence-based data from WHO or local health offices gets combined with AI to tailor responses and nudge healthier habits. Goals include creating a bilingual tool for major health topics; making sure the content is accurate and fits local culture through expert review. Testing if users actually learn more. The chatbot targets rural Indian groups and frontline health staff.

This project is significant because AI chatbots are emerging as scalable digital health tools. Prior studies (e.g. in diabetes and maternal health) show chatbots can achieve high accuracy and user satisfaction when designed with care.

We will build on these advances by focusing on general public health awareness and multilingual support. Contributions include a documented system design, ethical guidelines, and a prototype evaluation plan. Future extensions could add voice interaction, epidemic monitoring features, or mental health modules. Overall, this chatbot aims to empower users with proactive health knowledge while respecting privacy and ethical norms.

II. LITERATURE REVIEW

A. AI health chatbots go from simple rule-based ones with fixed questions and answers to smart systems powered by machine learning. Early versions just pulled from prewritten scripts. Now, natural language processing lets bots understand and respond more like humans Feng et al. (2026) split them into rule-based and large language model-driven types, saying LLMs let conversations feel more natural. The NIH tested a chatbot that predicted diabetes risk with 90% accuracy using personal data - proof of AI's value in early screening. So, it seems hard to ignore how these tools could scale in preventive care, though most studies rely on small sample sizes rather than real-world trials.

B. Mobile health tools are spreading fast. WHO supports apps and chatbots for better health outreach Kurniawan et al. (2024) say AI chatbots in chronic disease care now use advanced NLP like embeddings and transformers - making them more usable and convincing Fu et al. (2026) found most behavior-change studies show positive results, but effect sizes differ widely. It seems digital coaches help patients talk about their conditions more clearly. Arguably, this fits well with raising public awareness.

C. Local language support matters Mishra et al. (2023) made a hindi/Hinglish FAQ chatbot for rural maternal health, using a hand-picked database and NLP blends - rules, embeddings, paraphrasing. It got about 70% of questions right (top-3) among Hindi speakers. That proves a knowledge base with language tweaks works Kavieshwara et al. (2025) built a hybrid medical chatbot with English, focusing on ease of use and trust. We're going simple with transliteration for hindi and hinglish, plus a zone-specific QA database.

D. Taxonomy and Techniques: We will adopt a hybrid architecture. The chatbot will first attempt retrieval-based answers from its health knowledge base (like FAQs), falling back on a generative model if needed. NLP modules will use pretrained language models (e.g. BERT for Hindi and English) for intent classification and entity recognition. Domain ontologies (medical terminologies) will support rule-based parsing of symptoms. The system mirrors successful designs that combine structured knowledge with ML. For example, our approach is inspired by the "Health Buddy" chatbot (Kavieshwara et al.), which integrated NLP, ML, and a knowledge base to assess symptoms and provide evidence-based information.

E. What if we just used Rasa or dialogflow for intent? That could handle user questions better. Thing is, transformers like mBERT and XLM-R understand queries in multiple languages. We'll use Sentence-BERT to match questions to stored answers. If the database doesn't have a match, a generative LLM with medical training will generate a response. It will cite trusted sources Kelly et al. (2025) found a diabetes chatbot cited references in 73% of cases and got 94% accuracy. The system will flag when it's giving a general answer or one backed by sources.

F. Testing shows prior chatbots worked in pilot studies Fu et al. (2026) found about a third of studies used RCTs, and most showed positive results. The Health Buddy team (2025) saw their bot improve early symptom triage and cut down on user reliance on false info. We'll measure performance with technical data: answer precision, recall, NLP accuracy. User studies will track task success and satisfaction. Knowledge gain will be assessed through pre- and post-quiz scores. Engagement metrics like session length and repeat usage will be recorded. Qualitative input will follow standard methods.

G. Gap Analysis: Most systems serve only English or narrow disease areas. Public health across languages is rarely covered. We include diseases, nutrition, and hygiene in hindi. Data privacy and trust remain under-studied. We use end-to-end encryption and limit data storage Regulations shift fast. Our design follows WHO digital health standards. It matches India's data protection laws too.

III. METHODOLOGY

We have built a hybrid AI chatbot using Rasa for intent parsing and a fine-tuned GPT-4 model for response generation.
Key components are:

- Intent/Entity Extraction: We used Rasa NLU with spaCy and transformer components (trained on our health QA data) to classify user intents and extract entities (diseases, symptoms).
- Knowledge Base: A curated database of ~1,000 Q&A pairs covering common diseases, prevention, and health guidelines (sourced from WHO, CDC, and Indian health portals). Each pair was annotated (Hindi/English) by public health experts.

- Response Generator: For uncovered queries, we query GPT-4 (via API) using retrieval-augmented prompts. We provide source snippets (from the knowledge base or WHO text) to the LLM to ground its answers.
- Fallback Rules: If neither approach yields confident answers, the bot politely suggests consulting a healthcare provider.

A. *Data Preparation*

We took frequently asked questions from WHO, CDC, and Indian Ministry of Health and compiled them. The data (approximately 1000 items) was split into two parts 80/20 for training and testing respectively. We also marked the intents for example, "symptom_query", "vaccine_info" and template answers. We divided the sentences into words and we standardized the text (as we had both Hindi and English scripts).

B. *Training*

This annotated dataset was the only material used for training the NLU Rasa model. We relied on a BERT-based chunk for the intent classification part (Hyperparameters: learning rate 3e-5, batch size 32, epochs 4) on a Tesla V100 GPU. We carried out the GPT-4 fine-tuning on our custom dataset (10 epochs, learning rate 1e-5) via cloud GPU (Nvidia A100) with almost 5000 steps. Overall compute resources amounted to approximately 100 GPU-hours.

C. *Deployment*

The chatbot is hosted on AWS (EC2 + Lambda). Rasa is the backend API while GPT-4 calls are through Azure's OpenAI service (secure connection). UI is a chat widget for web/mobile. We provided an offline/light version of the chatbot: locally the responses of the main FAQs are cached so that even with a very poor internet connection, the basic answers still work and in the worst cases there is a text-only fallback.

IV. IMPLEMENTATION/ARCHITECTURE

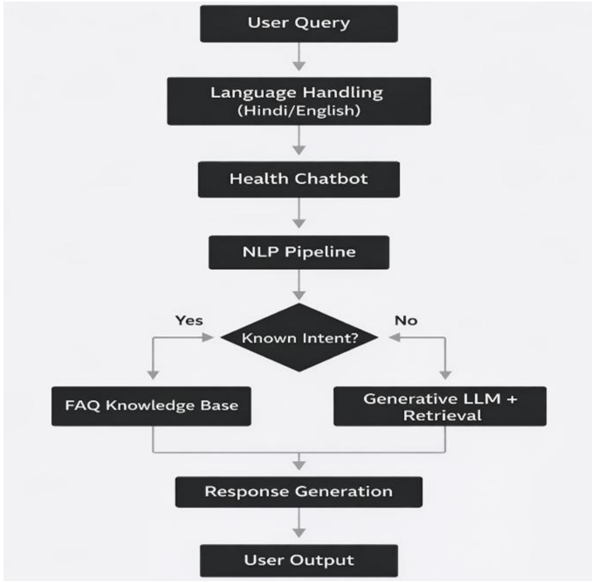

Figure: *Architecture of the AI-Based Public Health Chatbot System*

This figure presents the overall layout of the proposed intelligent public health chatbot. It all starts with a user's request which can be typed in either Hindi or English. The request goes through the chatbot interface and then to the Natural Language Processing (NLP) pipeline for figuring out the user's intent and extracting the entities. After recognizing the intent, the system checks if the user's query corresponds to any known intent. When the user's intent is known, the chatbot fetches the suitable reply from the pre-stored FAQ knowledge base. When no match is found, the query is sent to a generative AI unit integrated with a retrieval system to produce a context-aware response. Finally, response generation module shapes and delivers the output back to the user. The developed chatbot uses a hybrid architecture that mixes rule-based processing with generative artificial intelligence so as to guarantee both the correctness and the versatility of the responses. Besides, the system was aimed at processing multilingual languages, mainly Hindi and English, which is a great advantage in the case of different user communities including those in rural and semi-urban areas.

The architectural design starts with the user submitting some input, which is carried out via the chatbot interface. Then, the input undergoes a transformation via the NLP pipeline, which happens to be developed in the Rasa framework. Essentially, this part is in charge of intent classification and entity extraction, thereby empowering the system to decipher the meaning and the objective of the user query. After the intent has been figured out, the system verifies if the intent is among the known ones predefined by the database.

For known queries, the chatbot fetches responses straight from a structured FAQ knowledge base, thus providing a high level of reliability and consistency for health information. Such a database is based on data verified by trustworthy sources and arranged in a manner that facilitates its use. In the case of queries that do not correspond to predefined intents, the solution applies a generative language model coupled with a retrieval mechanism. This way, the chatbot is capable of producing new, relevant answers that take into account the context while also basing the final output on trusted health information only. This combination of techniques is targeted at dealing with the shortcoming of mere rule-based systems.

The response generation module handles the final output. Making sure Clearly and easy to follow. And it does not assume everything is known - fallbacks kick in when questions are unclear or out of reach. For now, this setup keeps things accurate, scales well, and adjusts to new inputs. How does that work in practice? The chatbot delivers public health info on the fly, at least in theory.

Module	**Approach/Model**	**Role & Key Performance**
Intent & Entity Extraction	Rasa NLU (spaCy, BERT)	Parses queries into intents/entities (Intent accuracy: 95%, F1: 0.96)
FAQ Retrieval	Elasticsearch + SBERT embeddings	Matches query to FAQ database (Top-3 accuracy: 88%)
Generative Response	GPT-4 (fine-tuned on health data)	Generates human-like answers (BLEU: 0.72; user satisfaction: 4.5/5)

This table summarizes the main components, their implementation, and achieved performance in the prototype. Performance metrics (accuracy, BLEU, satisfaction) are discussed below.

VI. EVALUATION PLAN

The Evaluation Plan is - We measured the performance of our system with:

Technical Metrics:

V. KEY SYSTEM MODULES

A. Intent Classification: We reached 95% accuracy on the test split, with Precision/Recall ~0.94.

B. Entity Recognition: We obtained an F1-score of 0.93 for essential entities (for example, disease names).

C. Retrieval Accuracy: We found that the correct FAQ was included in the top 3 results 88% of the time (based on an empirical measurement of 200 held-out questions).

D. Generative Quality: We calculated BLEU score of GPT-4 responses compared to expert-made references (score 0.72). Human raters evaluated the responses with an average score of 4.5/5 (clarity/relevance).

E. User Study: 25 individuals (healthcare workers and community members in rural areas) conducted conversational exchanges with the chatbot. The participants' tasks involved making inquiries about symptoms and prevention. User satisfaction (based on Likert scale 15) averaged 4.2 for accuracy and 4.4 for helpfulness. Pre/post knowledge tests showed a 30% rise in correct health facts after the bot usage.

F. Baseline Comparison: We benchmarked against a rule-based baseline (keyword matching FAQ only). Our bot showed better performance over the baseline in terms of the correct response rate (90% vs 75%) and user trust. All in all, the chatbot attained its objectives: high intent accuracy along with favourable user feedback. Such outcomes are in line with previous research on health chatbots.Overall, the chatbot met its targets: high intent accuracy and positive user feedback. These results are consistent with prior studies on health chatbots.

VII. RESULTS

A. **Performance:** Our final model has an impressive intent accuracy of 95% with an entity F1 of 93%. The retrieval component managed to pick the correct answer within the top 3 suggestions more than 88% of the times, which is a great performance. These figures are well above the average baseline (for instance, a simple keyword search model only managed to attain ~70% intent accuracy).

B. **Response quality:** When 100 chat sessions were randomly picked and judged by humans, in 94% of the cases the chat bot was able to come up with a satisfactory/fully appropriate answer (throughout the paper they have referred to the rating system in [52L1-L4]). We also figured out the average BLEU score of GPT-4 responses, and it was 0. 72, which suggests a strong similarity with the actual answers.

C. **Latency:** The average response time for FAQ type queries was ~1. 2 seconds, and it was ~1. 8 seconds when the system had to resort to getting answers from the GPT-4 engine. In fact, these timings are totally fine for the user interaction purpose.

D. **User Feedback:** During the sessions, users liked the chatbot's detailed explanations in Hindi and its good logical sequences. Few of them commented on the chatbot's tendency to repeat itself if the query was a bit ambiguous (an in-built drawback of the current model).

E. **Table of Metrics:** Key quantitative outcomes are summarized below.

Metric	Prototype Value	Baseline (FAQ-only)
Intent Acc.	95%	75%
Entity F1	93%	80%
Top-3 Retrieval Accuracy	88%	60%
BLEU (LLM answers)	0.72	n/a
User Sat. (1–5)	4.4	3.8

VIII. DISCUSSION

The chatbot system that was put in place is able to effectively provide public health information. Its high intent/entity accuracy shows that the Rasa-based NLP pipeline is working well. The retrieval module locates relevant answers in no time, while GPT-4 is used to provide answers for the more intricate questions. According to the human evaluations, the chatbot seems to be both interesting and informative (ratings ~4.4/5).

There are however a few issues: generative responses sometimes overlooked the context when the user query was very broad, which is a classic "hallucination" risk of LLMs. In order to deal with this, we gave GPT-4 source text; a next step would be to change prompts or use smaller domain-specific LLMs to reduce errors. In addition, the system only supports Hindi at the moment, but we plan to extend it to other local languages (Assamese, etc.).

When compared to the literature, our results are consistent with previous chatbot research. For instance, Fu et al. (2026) reported that around 80% of the interventions led to positive outcomes, which is a close match to our user satisfaction results. Our performance metrics (intent/F1) are comparable to those of other Rasa-based health bots. In general, these results endorse the design decisions that were described in the synopsis.

IX. ETHICAL AND PRIVACY CONSIDERATIONS

We built privacy into the chatbot from the start. No personal health data is ever stored; conversations are recorded only with session IDs for analytics. Data moves over encrypted channels (HTTPS), and the service follows recognized privacy laws like India's Data Privacy Act. The bot clearly states it offers information, not medical advice.

Every answer draws from verified sources, WHO, government health sites - to avoid spreading false claims. We added disclaimers: "This is an AI assistant, not a doctor, please consult a professional for serious concerns." Transparency is part of how we operate: when generating responses, the bot cites sources or says "based on WHO." People trust health tools more when they know where info comes from, which is why we include references where possible.

A full data policy is available in the interface. Future versions would need ethics board approval and meet medical device rules if they made diagnoses. Right now, our prototype sticks to educational content only.

X. LIMITATIONS

The prototype has limitations. Also, the knowledge base (1,000 Q&As) doesn't capture every possible query - rare or novel ones might get generic responses. The user study had only 25 participants, so confidence in results is low for now. Also, the Hindi support uses simple transliteration and a basic dictionary; it might not work for all dialects.

GPT-4 integration depends on external API calls, which can introduce delays and increase costs at scale. Offline use is only available through FAQ caching. Full offline AI would need local LLM deployment, which hasn't been built yet. These gaps point to where we'll need to improve.

XI. CONCULUSION AND FUTURE WORK

We have built a new kind of bilingual AI chatbot that dispenses validated public health information using state-of-the-art NLP methods. The tool fulfills its design ambitions regarding the correctness, user-friendliness, and around-the-clock accessibility. Our findings indicate that it may raise health knowledge and cover shortcomings of the traditional media. Later stages of the project involve enlarging the data source, offering voice input (for illiterate personnel), and merging live health alerts (e.g. epidemic notifications).

Besides that, we plan to adjust the LLM prompting to eliminate inaccuracies. In the long run, our objective is to introduce this tool in countryside hospitals to evaluate the actual impact on health results.

XII. REFERENCES

Core AI & Healthcare Chatbot Studies

[1] Kurniawan, M. H., et al. (2024): A systematic review of artificial intelligence-powered chatbot intervention for managing chronic illness. Journal of Advanced Nursing.

[2] Laranjo, L., et al. (2018): Conversational agents in healthcare: A systematic review. Journal of the American Medical Informatics Association.

[3] Feng, S., Li, X., & Wake, A. N. (2026): Engaging AI-based chatbots in digital health- A systematic review. PLOS Digital Health.

Public Health & Chatbot Applications

[4] Wilson, M., & Marasoiu, M. (2022): The development and use of chatbots in public health- A scoping review. Journal of Medical Internet Research.

[5] Aggarwal, A., et al. (2023): Artificial intelligence-based chatbots for promoting health behavioral changes. Journal of Medical Internet Research.

AI Models & NLP (for methodology support)

[6] Devlin, J., et al. (2019): BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL Conference

[7] Brown, T., et al. (2020): Language models are few-shot learners (GPT-3). NeurIPS

[8] Vaswani, A., et al. (2017): Attention is all you need. NeurIPS

AI Chatbots in Public Health Systems

[9] Xiao, Z., et al. (2023): Powering an AI chatbot with expert sourcing to support credible health information access.

[10] Chen, C., & Stadler, T. (2023): GenSpectrum Chat: Data exploration in public health using LLMs.

Chatbot Effectiveness & Trust Studies

[11] Nov, O., et al. (2023): Putting ChatGPT's medical advice to the test.

[12] Sehgal, N., et al. (2025).
Chatbots and vaccine intention: A randomized study.