



Medical Chatbot Integrated with Natural Language Processing and Voice Assistance

G. B. Veeresh¹, M.C.A., (Ph.D.), Assistant Professor

Veeramreddy Myna, Shaik Zubair Ahmed, Jeerla Madhu, Varra Venkata Lokeshwar Reddy, Kancharla Hemanth

Department of Computer Science and Engineering, K.S.R.M. College of Engineering, Kadapa

Corresponding author 1: veeresh@ksrmce.ac.in

Abstract—This paper describes a medical chatbot built on machine learning and natural language processing techniques to deliver health information, assess reported symptoms, and offer preliminary guidance to users. Rather than treating each query in isolation, the system retains a record of prior interactions and uses that history to shape later responses. Voice interaction is not an add-on but a core requirement, since it opens access to elderly users, people with disabilities, and anyone for whom typing on a small screen is impractical. The chatbot is built to connect with Electronic Health Record systems, so that its guidance can draw on a patient's actual medical history instead of returning the same generic answer to everyone. It further functions as a triage layer: by estimating how urgent a complaint is and pointing the user toward an appropriate level of care, it shortens the time to useful advice while easing pressure on services meant for more serious cases. Access is through a web interface usable from any device without installation, and three interaction modes, text-to-text, text-to-voice, and voice-to-voice, were implemented and tested. The goal throughout is to help users make better-informed decisions and seek care sooner, not to replace a clinician's assessment.

Index Terms—medical chatbot, natural language processing, voice assistance, triage, electronic health records, healthcare accessibility.

I. Introduction

Someone who notices a new symptom is often unsure what to do next. What matters is not whether a symptom is dangerous as an abstract fact, but whether this instance, for this particular person, calls for a same-day visit, a phone call to a doctor, or simply riding it out. Conventional healthcare settings are not set up to answer that question on short notice, and it is in the interval between noticing a symptom and getting advice that both needless anxiety and dangerous delay tend to appear.

The system described here, a medical chatbot, uses machine learning to interpret what a user types or says and to generate health information, a symptom assessment, and preliminary guidance in response. Its approach is personalized: each new interaction builds on what the chatbot has already learned about that individual, instead of starting over with no context.

Voice support is not a secondary feature bolted onto a text system; it is central to the design. A text-only interface quietly assumes that the user can read fluently, has fine motor control, and is comfortable typing on a small screen. Each of these assumptions leaves someone out: older adults, people with motor impairments, and those with limited formal education. Adding voice interaction removes that dependence on typing, extending access to precisely the groups whose health needs tend to be greatest.

The second design commitment is integration with Electronic Health Record (EHR) systems. Access to a patient's existing records allows the chatbot to answer in light of that individual's medical history rather than in generic terms, a distinction that separates useful guidance from a symptom checker that hands everyone the same list of possibilities. In helping a user gauge how urgent a complaint is and pointing them toward the right resource,

the chatbot also operates as a triage aid, cutting the time to appropriate care while reducing unnecessary visits to services intended for more severe cases.

II. Problem Definition

The system responds to a set of interconnected gaps in how people currently access healthcare.

- 1) Reliable medical information is hard to reach quickly: people with a health concern often cannot get timely, trustworthy advice, and conventional care settings are not built for immediate response.
- 2) Symptom assessment tends to be slow, so people are often left to decide their next step before any guidance arrives.
- 3) Support is rarely personalized. Generic information overlooks an individual's specific circumstances, which breeds uncertainty and hesitation about whether to seek care at all.
- 4) Typed input is itself a barrier for users with limited typing ability or a disability, effectively excluding them from any service delivered purely through text.
- 5) Health records remain fragmented across systems with limited interoperability, leaving neither the patient nor the provider with a complete picture.

This last gap makes every other one worse. When a patient's history sits scattered across systems that do not talk to each other, each new interaction starts from incomplete information, and it is the patient, at precisely the moment they are least able to manage it, who ends up responsible for piecing that history back together.

III. Proposed System and Design

A. Proposed Solution

The proposed chatbot relies on machine learning models to interpret what a user reports, assess the symptoms described, and return medical information along with preliminary guidance. Because it learns continuously from past interactions, its responses adapt over time to what it already knows about that individual. Voice input and output sit alongside text so that users who cannot type comfortably reach identical functionality. EHR integration supplies the patient data needed to ground guidance in a person's actual history and current condition, and a web interface, usable from any device, serves as the single point of access for entering symptoms, retrieving information, and receiving recommendations.

B. System Architecture

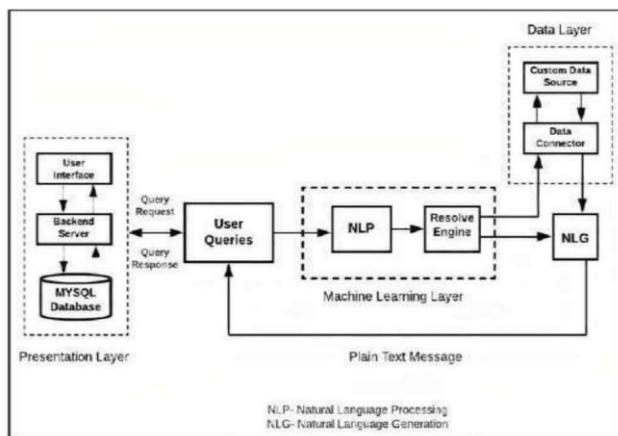


Fig. 1. System architecture of the medical chatbot.

As Fig. 1 shows, the pipeline begins with input capture in either text or speech form. Spoken input first passes through speech recognition, after which natural language understanding extracts the symptoms and intent contained in the message. The extracted information is then handed to the assessment model, and a response is generated, returned as text or as synthesized speech depending on the mode in use. Keeping speech recognition and speech synthesis as separate stages on either side of one shared core is what lets all three interaction modes reuse a single assessment engine instead of each being built independently.

C. Model Selection and Evaluation

A range of algorithms suited to natural language processing were evaluated, spanning established classifiers such as Naive Bayes and Support Vector Machines through to neural network-based approaches. Comparing candidates systematically, rather than committing early to a single architecture, was a deliberate choice: the right model depends as much on how much good training data is on hand as on the nature of the task, and an elaborate model trained on a modest medical corpus typically generalizes worse than a simpler one would.

IV. Implementation

The system is implemented in Python and deployed as a web application. Natural language processing performs tokenization, normalization, and intent extraction on each incoming message; the assessment model then maps the extracted symptoms to candidate conditions and an urgency level; and a separate response-generation stage composes the reply. Speech recognition converts spoken input into text at the front of the pipeline, and

speech synthesis converts the generated response into audio at the end. Confining speech handling to these two endpoints is precisely what lets the text-to-voice and voice-to-voice modes exist without a second copy of the reasoning logic sitting in between.

User accounts store the interaction history that personalization depends on. The design also anticipates EHR integration through a defined interface, so that, where an institutional system is available, patient history can be drawn from it directly rather than being built up solely from chat records.

The response policy is worth stating explicitly. What the system offers is information paired with an urgency rating; diagnosis is deliberately left out of scope. When the assessment indicates urgency, the response directs the user toward appropriate care rather than offering management advice, and every reply carries a reminder that it is meant to complement medical consultation, not substitute for it. This is treated as a design constraint rather than a legal disclaimer, since a system that behaves as though it is diagnosing will be trusted as though it is, whatever its terms of use happen to say.

V. Results and Discussion

The completed system was tested across all three interaction modes. Fig. 2 shows the login page through which a user establishes a session and links to their interaction history.

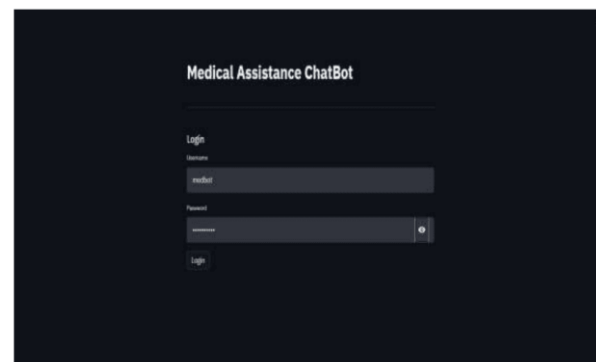


Fig. 2. Login page.

Fig. 3 shows the text-to-text mode, where a typed query receives a typed reply. This mode serves as the baseline against which the other two are compared, since it isolates the assessment engine from the speech components entirely.



Fig. 3. Text-to-text interaction.

Fig. 4 shows the text-to-voice mode, in which a typed query returns a spoken response, useful for a user who can type but struggles to read a screen. Fig. 5 shows voice-to-voice, the fully hands-free mode built for users who cannot type at all.

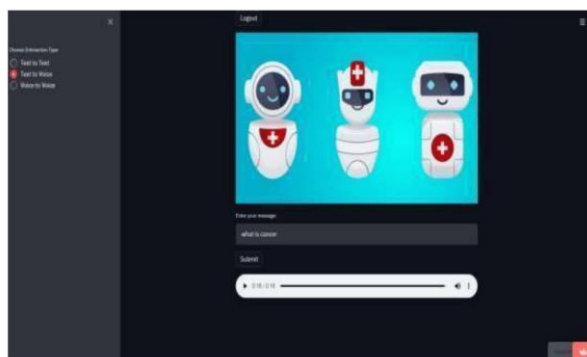


Fig. 4. Text-to-voice interaction.



Fig. 5. Voice-to-voice interaction.

All three modes ran correctly against the same assessment engine, which confirms that separating speech handling from the reasoning core achieved its intended purpose. In the spoken modes, the accuracy of voice recognition set the ceiling on performance, especially with accented speech and clinical vocabulary that general-purpose recognition models are not tuned to handle. A symptom described using a regional term may fail to transcribe at all, and once a mis-transcription occurs, it propagates silently into the assessment that follows.

Three limitations bear directly on any real deployment. First, the assessment is only as reliable as the data used to train it; any symptom pattern the training data never saw will be handled badly, with no flag to the user that this occurred. Second, the chatbot cannot examine the patient directly; it works entirely from self-reported symptoms, and the accuracy of that reporting varies with the user's health literacy. Third, and most seriously, any system that offers a preliminary assessment risks giving false reassurance: a user told their symptoms are probably minor may put off care they actually needed. For this reason, the design resolves uncertainty toward escalation rather than reassurance, and the output that matters is the urgency assessment, not the specific condition suggested.

VI. Conclusion and Future Enhancements

This paper has presented a medical chatbot that pairs natural language processing with voice assistance to provide health information, an assessment of reported symptoms, and triage guidance. Three interaction modes built on one shared assessment engine, text-to-text, text-to-voice, and voice-to-voice, put the same capability within reach of users who cannot type as of those who can. The system is designed to integrate with Electronic Health Records so that its responses can draw on patient history, and it operates as a triage aid that directs users to an appropriate level of care rather than attempting to replace clinical judgment.

Future work should first tackle recognition errors on Indian-accented speech and on regional medical terms, since this is

exactly where today's accessibility gains break down in practice. Adding multilingual support, particularly Telugu given the system's regional context, would extend its reach considerably; a voice interface in a language the user cannot speak comfortably provides no benefit at all. Completing EHR integration against an actual institutional system would turn personalization from a design intention into working practice. Finally, clinical validation of the urgency assessment against real patient outcomes, rather than against a held-out test set, is a precondition for patient-facing use: a triage tool has to demonstrate that it escalates the cases that genuinely needed escalation before it can responsibly be handed to someone deciding whether to seek care.

References

- [1] L. Laranjo et al., "Conversational agents in healthcare: a systematic review," *J. American Medical Informatics Association*, vol. 25, no. 9, pp. 1248–1258, 2018.
- [2] A. Palanica, P. Flaschner, A. Thommandram, M. Li, and Y. Fossat, "Physicians' perceptions of chatbots in health care," *J. Medical Internet Research*, vol. 21, no. 4, 2019.
- [3] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed. draft, 2023.
- [4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
- [5] H. Xu, S. Stenner, S. Doan, K. Johnson, L. Waitman, and J. Denny, "MedEx: a medication information extraction system for clinical narratives," *J. American Medical Informatics Association*, vol. 17, no. 1, pp. 19–24, 2010.
- [6] Health Level Seven International, *HL7 FHIR Release 4 Specification*. [Online]. Available: <https://www.hl7.org/fhir/>
- [7] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. Sebastopol, CA: O'Reilly Media, 2009.
- [8] World Health Organization, *Ethics and Governance of Artificial Intelligence for Health*. Geneva: WHO, 2021.