

DawAI: AI Powered Medical Chatbot

¹Dr. Abdul Khadeer

¹Associate Professor, Department of Computer Science and Engineering
Deccan College of Engineering and Technology, Aghapura, Hyderabad, 500001 Osmania University

¹Mohammed Nabeel Uddin, ²Mohammed Abrar Ahmed, ³Bilal Ahmed Maaz

^{1,2,3}UG Scholar, Department of Computer Science and Engineering
Deccan College of Engineering and Technology, Aghapura, Hyderabad, 500001 Osmania University

Abstract

DawAI is an AI-powered medical chatbot designed to provide preliminary healthcare assistance using multimodal inputs such as speech and images. The system integrates advanced technologies including Whisper for speech-to-text conversion, a large language model for contextual understanding, and text-to-speech systems for generating human-like voice responses. Users can describe symptoms verbally or upload medical images, and the chatbot analyzes this information to deliver informative, context-aware responses. Built with a simple and interactive interface, DawAI aims to improve accessibility to basic medical guidance, especially in remote or resource-limited areas. The system serves as an efficient, scalable solution for early-stage health consultation and awareness.

INTRODUCTION:

Healthcare systems often face issues such as limited accessibility, long waiting times, and lack of immediate guidance. Existing digital solutions are mostly text-based and do not support voice or image inputs, limiting their real-world effectiveness.

DawAI addresses this gap by providing a multimodal AI-powered medical chatbot that accepts speech and image inputs to generate informative responses. The system integrates speech recognition, image analysis, and natural language processing within a simple interface.

It serves as a lightweight and accessible solution for preliminary medical assistance, especially in resource-constrained environments.

1. LITERATURE SURVEY:

Existing research in AI-based healthcare systems highlights the growing use of multimodal technologies for improving medical assistance. Studies on vision-language models demonstrate that combining image and text inputs can enhance diagnostic understanding, though challenges remain in accuracy and real-time processing.

Speech recognition models such as Whisper have been widely used in healthcare chatbots to convert patient voice input into text. While effective, they face limitations in handling medical terminology and diverse accents. Similarly, text-to-speech systems like gTTS and neural voice models improve user interaction by generating human-like responses, but often depend on stable internet connectivity.

User interface frameworks like Gradio enable rapid development of interactive AI applications by integrating multiple components such as audio, text, and image inputs into a single platform. However, customization and scalability can be limited.

Overall, current systems either lack multimodal integration or fail to provide seamless, real-time interaction. DawAI builds upon these approaches by combining speech, image, and language processing into a unified and efficient healthcare chatbot system.

2. EXISTING SYSTEM:

Current healthcare chatbot systems are primarily text-based and rely on predefined rules or basic machine learning models. These systems can handle simple queries but struggle with complex or real-world medical scenarios.

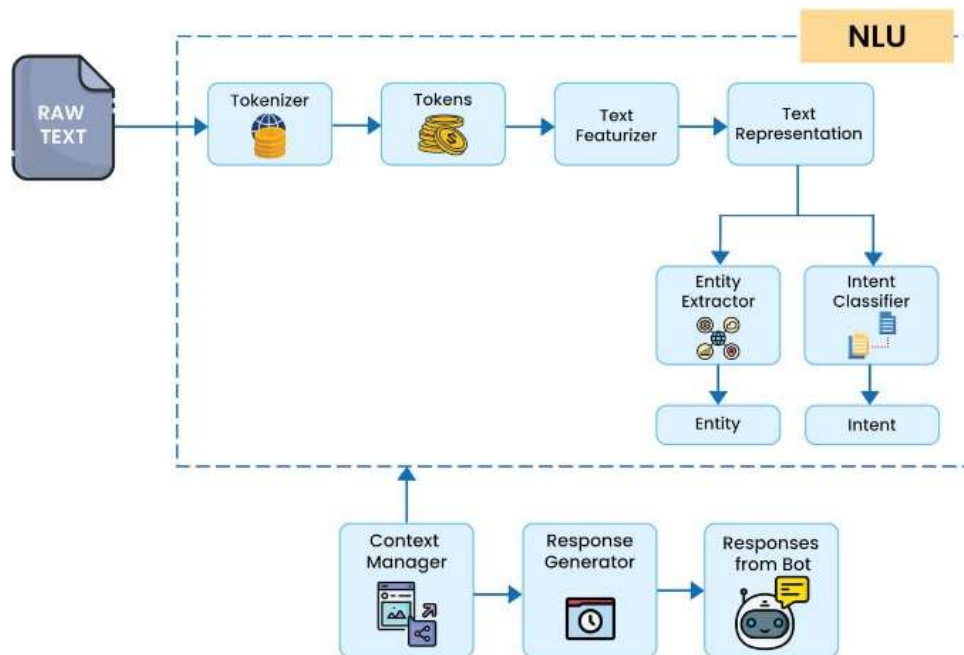
Most existing solutions do not support multimodal interaction, meaning they cannot

process voice input or analyze medical images. Additionally, many platforms depend on third-party services, limiting customization and raising data privacy concerns.

As a result, these systems provide limited accuracy, reduced user interaction, and are not fully effective for comprehensive healthcare assistance.

3. LIMITATIONS:

Current healthcare chatbot systems have several limitations that reduce their effectiveness. They are mostly text-based and cannot handle voice or image inputs, limiting real-world usability. Many systems rely on predefined rules, making them less flexible and unable to understand complex or varied user queries. They often lack accuracy when dealing with medical terminology or uncommon symptoms. Dependence on third-party services increases privacy risks and reduces control over data. Additionally, these systems may not provide personalized or context-aware responses, resulting in a less engaging and less reliable user experience.



EXISTING SYSTEM ARCHITECTURE:

4. PROPOSED SYSTEM:

The proposed system, DawAI, is a multimodal AI-powered medical chatbot designed to provide preliminary healthcare assistance through an interactive platform. It allows users to input symptoms using voice or upload medical images, enabling a more natural and real-world interaction.

The system integrates speech-to-text conversion,

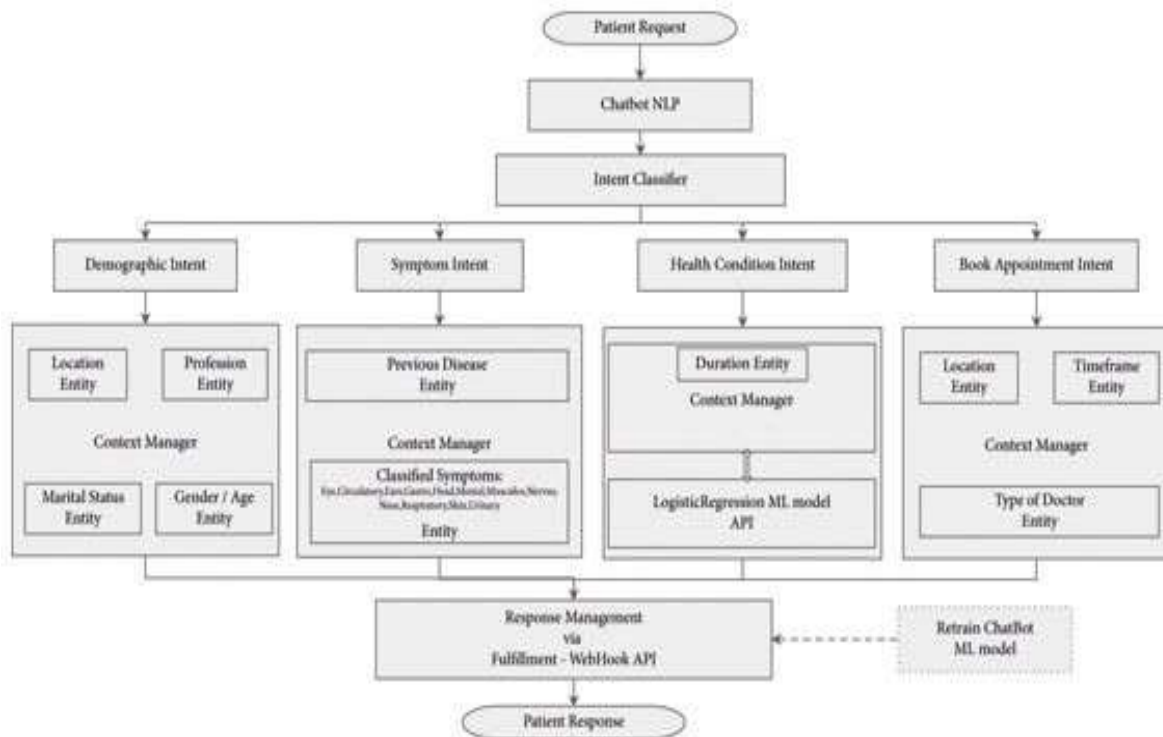
image analysis, and natural language processing to understand user queries and generate informative responses. It also includes text-to-speech functionality to deliver responses in a human-like voice.

DawAI is designed to be lightweight, accessible, and easy to use, making it suitable for academic purposes and basic healthcare support. It offers a more flexible and user-friendly solution compared to existing systems by combining multiple AI capabilities into a single platform.

Advantages:

1. Supports multimodal interaction by accepting both voice and image inputs, making the system more practical and user-friendly
2. Generates human-like responses using text-to-speech, improving accessibility and overall user experience
3. Lightweight and easy to deploy on basic systems without requiring high-end hardware or complex setup
4. Cost-effective as it uses open-source tools and APIs, reducing development and operational expenses
5. Modular architecture allows easy customization, maintenance, and future feature enhancements
6. Provides faster, context-aware, and more interactive responses compared to traditional text-based systems

5. IMPLEMENTATION:



The implementation of DawAI is based on integrating multiple AI components into a single workflow. Users interact with the system through a simple interface where they can provide input via voice or upload medical images. The audio input is processed using a speech-to-text model to convert it into text, while images are analyzed using a multimodal AI model.

The processed inputs are then sent to a language model, which generates context-aware response

System Integration & Local Hosting:

The system is designed to run in a local hosting environment using tools such as XAMPP or a Python-based server. It operates on localhost, allowing developers to test and interact with the application through a web browser without requiring internet deployment.

The platform can also be configured to run over a local area network (LAN), enabling access

8. CONCLUSION AND FUTURE SCOPE:

DawAI demonstrates how multiple AI technologies can be integrated to create an intelligent and accessible healthcare chatbot. By combining speech recognition, image analysis, and natural language processing, the system provides a more interactive and user-friendly approach to preliminary medical assistance. It addresses key limitations of existing systems by enabling multimodal

based on the symptoms provided. This response is further converted into speech using text-to-speech technology to create a more interactive experience.

The system is developed using Python and deployed through a web-based interface, enabling easy access and testing. Basic security measures such as input validation and session handling are implemented to ensure safe and reliable operation.

from multiple devices within the same network for testing or demonstration purposes. This allows seamless interaction across systems without external hosting.

The system integrates all components, including speech processing, image analysis, and response generation, within a unified environment. Temporary data such as audio and images are handled locally to ensure privacy and efficient processing.

interaction and delivering context-aware responses.

In the future, the system can be enhanced by integrating secure payment and teleconsultation features, adding real-time search capabilities, and developing an admin dashboard for better management. Further improvements may include multilingual support, advanced security mechanisms.

9. REFERENCES

1. Y. Chen, M. Kapoor, and R. Dey. “Advancing multimodal AI for image and text-based medical diagnosis.” *Journal of Applied AI in Medicine*, vol. 12, pp. 23–39, Mar. 2023.
2. A. Radford, J. W. Kim, T. Xu, et al. “Robust speech recognition using Whisper Large V3.” *OpenAI Research Blog*, vol. 8, pp. 1–11, Dec. 2022.
3. Google Cloud Team. “Text-to-speech conversion using Google gTTS API.” gTTS API Documentation, [Online]. Available: <https://pypi.org/project/gTTS/> [Accessed: May 12, 2025].
4. ElevenLabs Inc. “Neural TTS with human-like voice synthesis.” ElevenLabs Developer Portal, [Online]. Available: <https://www.elevenlabs.io/docs/> [Accessed: May 12, 2025].
5. A. Abid and D. Khan. “Gradio: Interactive interface for ML applications.” *Journal of Open Source Software*, vol. 6, no. 64, pp. 2341, Jun. 2021.
6. C. Yuan, H. Li, and Y. Wu. “Vision-language pretraining for medical image understanding.” in *Proc. MICCAI*, 2021, pp. 133–147.
7. A. Mishra, S. Raj, and K. Bandyopadhyay. “Explainable AI for medical diagnosis using attention mechanisms.” *IEEE Trans. Med. Imaging*, vol. 41, no. 3, pp. 1120–1131, Mar. 2022.
8. H. Tan, L. Singh, and B. Nagaraj. “Real-time virtual health assistants using AI pipelines.” in *Proc. ACM SIGAI Health Workshop*, 2021, pp. 89–97.
9. S. Holzinger, C. Biemann, and A. Pattichis. “Explainable AI in healthcare for clinical transparency.” *IEEE ITB*, vol. 24, no. 2, pp. 202–210, Apr. 2020.
10. T. Brown, B. Mann, N. Ryder, et al. “Language models are few-shot learners.” *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.
11. J. Devlin, M. Chang, K. Lee, and K. Toutanova. “BERT: Pre-training of deep bidirectional transformers for language understanding.” in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
12. Z. Zhang, P. Xie, and E. Xing. “BioBERT: A pre-trained biomedical language representation model.” *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020.
13. A. Esteva, B. Kuprel, R. Novoa, et al. “Dermatologist-level classification of skin cancer with deep neural networks.” *Nature*, vol. 542, pp. 115–118, 2017.