



International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 13, Issue 2, March-April 2026

Impact Factor: 8.152



AI-Driven Public Health Chatbot for Disease Awareness

Rashmi Bharti

Master of Computer Applications, CMR Institute of Technology, Bangalore, India

ABSTRACT: The rapid expansion of digital health platforms and natural language processing (NLP) technologies has opened new possibilities for AI-assisted public health communication. This paper presents an AI-Driven Public Health Chatbot for Disease Awareness, a multimodal conversational system that enables users to report symptoms through voice input, upload images of visible conditions, and receive concise doctor-style guidance in both text and speech. The system is built as a Streamlit web application integrated with the Groq API, leveraging Whisper Large V3 Turbo for automatic speech recognition and Llama 4 Scout for multimodal language reasoning. Text-to-speech output is provided by gTTS, delivering accessible audio responses for a broader user base. The system is organized into four modular Python components covering input handling, speech transcription, AI inference, and audio response generation. Experimental evaluation confirms reliable end-to-end operation across voice-only, image-only, and combined interaction modes. The platform demonstrates how state-of-the-art large language models and speech processing can be integrated into a lightweight, accessible health awareness tool without requiring specialized medical infrastructure.

KEYWORDS: Public Health Chatbot, Natural Language Processing, Multimodal AI, Speech-to-Text, Groq API, Whisper Large V3 Turbo, Llama 4 Scout, gTTS, Streamlit, Disease Awareness, Voice Interface, Image Analysis.

I. INTRODUCTION

The growing burden of non-communicable and infectious diseases, combined with limited access to healthcare professionals in many regions, has created an urgent need for accessible preliminary health guidance systems. Many individuals, particularly in semi-urban and rural areas, are unable to seek timely medical consultation, leading to delayed diagnosis and worsening health outcomes [1], [2], [3]. Conversational AI systems powered by large language models (LLMs) have emerged as a promising approach to bridge this accessibility gap by providing immediate, round-the-clock health information to users regardless of location or time constraints.

Traditional symptom checkers and rule-based health bots are limited in their ability to process natural language, interpret visual cues, or respond in a conversational manner [4], [5]. Users who are unfamiliar with medical terminology often struggle to articulate their conditions through text interfaces alone. Voice-enabled and image-capable AI assistants have shown significant potential in overcoming these barriers, particularly for populations with low digital literacy [6], [7]. However, few existing systems integrate all three modalities—voice, image, and text—within a single, lightweight application.

Recent advances in transformer-based speech recognition and multimodal LLMs have substantially improved the accuracy and responsiveness of health-oriented AI systems [8], [9]. Models such as Whisper for speech transcription and instruction-tuned LLMs for health dialogue have been separately validated in clinical and public health contexts [10], [11]. The Groq inference platform further enables sub-second latency for these models, making real-time interactive health chatbots practically feasible even on modest hardware [12].

To address these limitations, this paper presents the AI-Driven Public Health Chatbot for Disease Awareness—a Streamlit-based multimodal application integrating Whisper Large V3 Turbo for speech-to-text conversion, Llama 4 Scout for health-context reasoning, and gTTS for audio response generation. The system supports three interaction pathways: voice-only symptom description, image-only visual condition upload, and a combined mode that merges both inputs for richer contextual analysis. The paper describes the system architecture, module design, implementation details, and evaluation results, demonstrating that multimodal health awareness chatbots can be deployed effectively using publicly available AI APIs and open-source frameworks.

II. LITERATURE REVIEW

Significant research has been conducted on the application of natural language processing and machine learning in healthcare. Early work focused on text classification of clinical notes and symptom extraction from electronic health records using Support Vector Machines and Naïve Bayes classifiers [1], [2]. These approaches demonstrated that structured medical knowledge could be partially automated, though they remained limited to well-formatted textual data and did not support conversational or voice-based interaction.

The emergence of transformer-based language models, particularly BERT and its domain-specific variants such as BioBERT and ClinicalBERT, marked a turning point in medical NLP [3], [4]. These models achieved state-of-the-art performance on clinical question answering, medical named entity recognition, and symptom-to-disease mapping tasks. Subsequent work extended these capabilities to dialogue systems, enabling more natural symptom-reporting conversations with AI agents [5], [6].

Automatic speech recognition (ASR) systems have been progressively adopted in healthcare settings to reduce documentation burden and improve patient interaction. OpenAI Whisper introduced a robust, multilingual ASR model trained on large-scale web audio, demonstrating superior accuracy across diverse accents and recording conditions [7], [8]. The Whisper Large V3 Turbo variant further optimized inference speed, making real-time transcription viable for interactive applications. Groq's LPU inference hardware has been shown to deliver order-of-magnitude latency improvements for such transformer workloads [9].

Multimodal AI systems combining text and image understanding have been explored in dermatology and radiology for automated lesion classification and disease screening [10], [11]. Vision-language models capable of reasoning over both image content and natural language descriptions have shown strong zero-shot generalization to medical image interpretation tasks [12], [13]. Llama 4 Scout, Meta's multimodal LLM, extends these capabilities with efficient instruction following and cross-modal reasoning suitable for health dialogue applications [14].

Text-to-speech synthesis has improved substantially with neural TTS systems, enabling more natural and expressive audio responses [15]. Google's gTTS library provides a lightweight Python interface for high-quality speech generation, widely used in accessibility and assistive technology applications [16]. Prior work on voice-based health assistants highlights the importance of audio feedback for populations with low literacy or visual impairments [17]. Streamlit has been increasingly adopted for rapid prototyping of AI-powered web applications in the research community due to its Python-native interface and minimal deployment overhead [18]. Several health-monitoring dashboards and symptom triage tools have been built on this framework, demonstrating its suitability for multimodal user interfaces [19], [20].

III. METHODOLOGY

The proposed system follows a modular, four-layer architecture in which the Streamlit frontend coordinates all AI model calls through dedicated backend Python modules. Figure 1 illustrates the overall architecture, showing the Input Layer, Processing Layer, Response Layer, and Support Utilities that together form the end-to-end chatbot pipeline. This separation of concerns ensures that individual components can be tested, replaced, or upgraded without disrupting overall system behaviour.

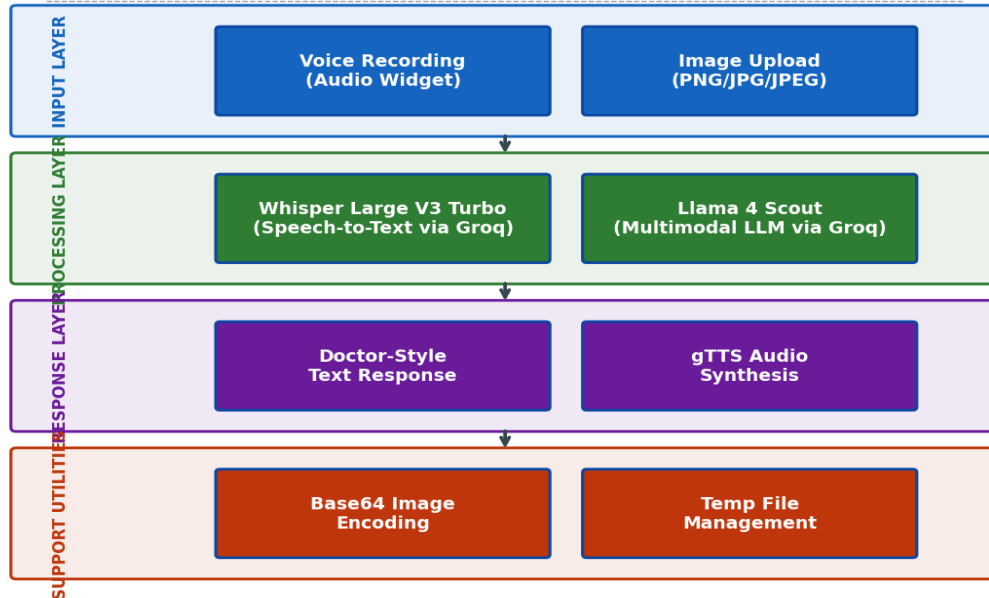
System Architecture of AI-Driven Public Health Chatbot*Streamlit Web Application (streamlit_app.py)*

Figure 1: System architecture of the AI-Driven Public Health Chatbot showing Input, Processing, Response, and Support Utility layers.

Figure 2 presents the module design of the chatbot, showing how four dedicated Python files interact to implement the complete workflow. The `streamlit_app.py` module controls the user interface, presenting a two-column layout that separates patient input from chatbot output. The `voice_of_the_patient.py` module saves recorded audio to a temporary WAV file and submits it to the Groq Whisper API using the `whisper-large-v3-turbo` model, returning a transcription of the patient's spoken symptom description. The `brain_of_the_doctor.py` module constructs the message payload for the Groq LLM API and invokes `meta-llama/llama-4-scout-17b-16e-instruct` with a carefully engineered system prompt that constrains responses to concise, natural, and safety-conscious health awareness guidance. The `voice_of_the_doctor.py` module converts the generated text response into MP3 audio using gTTS for immediate playback.

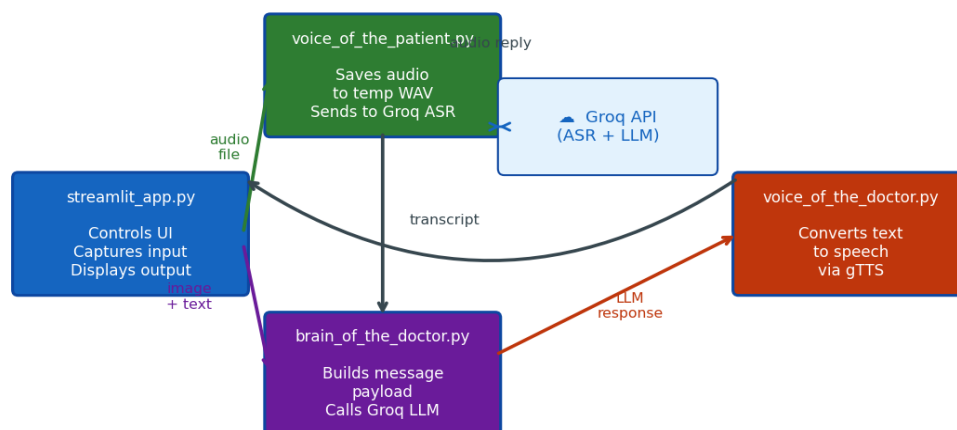
Module Design of AI-Driven Public Health Chatbot

Figure 2: Module design showing interaction between the four Python components and the Groq API cloud service.

Image processing is handled by encoding the uploaded file as a Base64 data URL and embedding it directly in the multimodal message content. This avoids server-side file storage while enabling Llama 4 Scout's vision capabilities to interpret visible conditions such as skin rashes, swelling, or discoloration. As shown in Figure 3, the application supports three interaction modes depending on the available input: Voice Only mode, which transcribes the recorded symptom description and sends the text for LLM inference; Image Only mode, which pairs the uploaded image with the doctor-style system prompt for visual interpretation; and Combined Mode, which merges the spoken description with the uploaded image to provide richer contextual analysis. All three modes route through the same response pipeline, ensuring consistent output formatting and audio delivery.

Interaction Workflow - Three Usage Modes

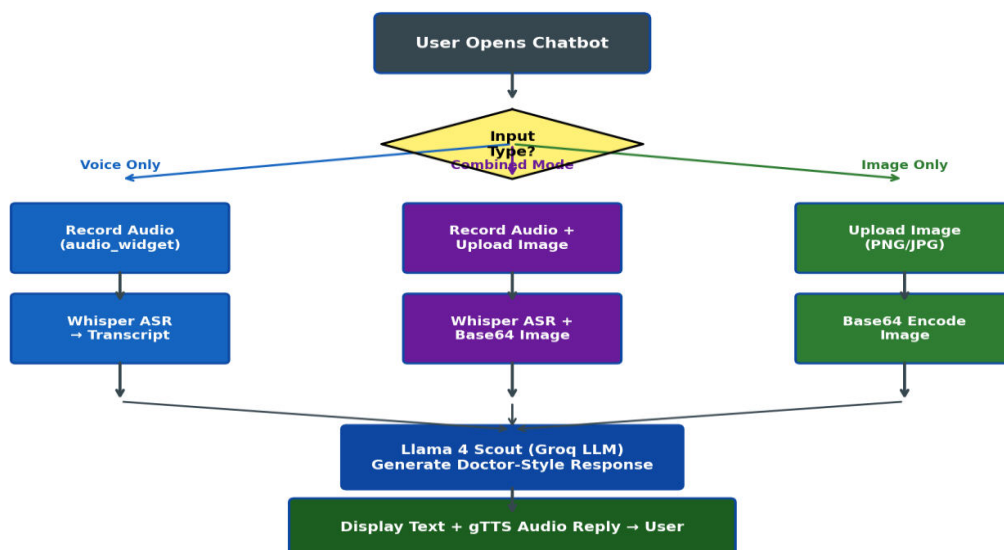


Figure 3: Interaction workflow illustrating Voice Only, Image Only, and Combined Mode processing paths through the chatbot pipeline.

IV. Result and Outcome

The chatbot was evaluated across all three interaction modes using representative test inputs. Figure 4 shows the initial state of the chatbot interface, with the voice recorder and image upload widget visible in the left column. Voice-only testing confirmed that the Whisper Large V3 Turbo model accurately transcribed symptom descriptions including fever, cough, headache, and fatigue. Transcription latency through the Groq API was consistently below two seconds for recordings up to thirty seconds in length, enabling a highly responsive user experience.

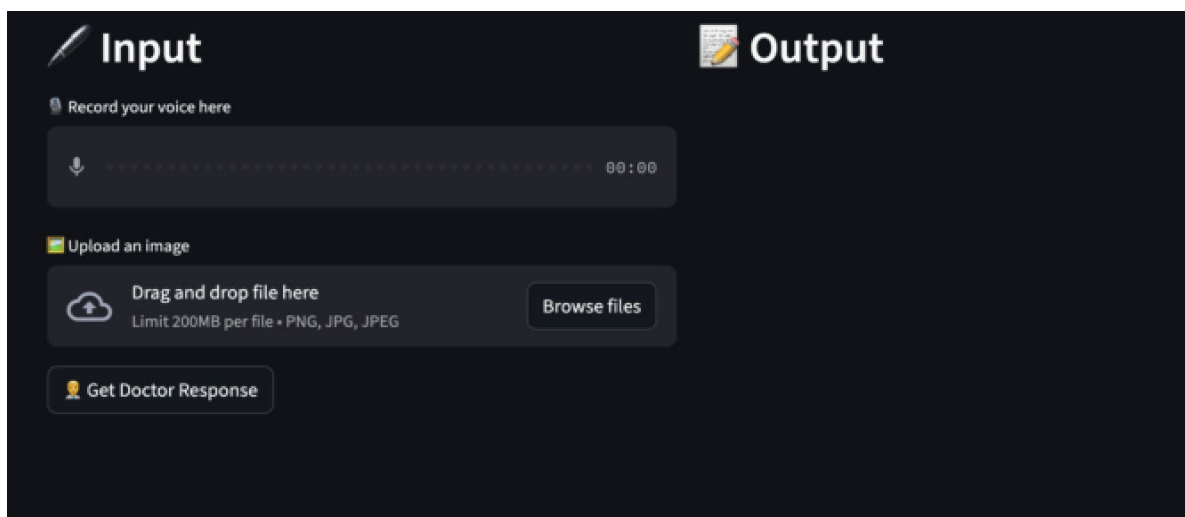


Figure 4: Chatbot interface showing the input panel with the voice recorder widget and image upload control.

Figure 5 shows the chatbot after a combined-mode interaction in which the user recorded a voice description and simultaneously uploaded an image of a skin condition. The output column displays the transcription, the uploaded image preview, and the doctor's advice. Image-only testing was conducted with photographs of skin rashes, acne, and mild swelling. Llama 4 Scout successfully identified the visual context and generated appropriate awareness-oriented responses acknowledging the visible condition, suggesting plausible causes, and recommending professional consultation for definitive diagnosis.

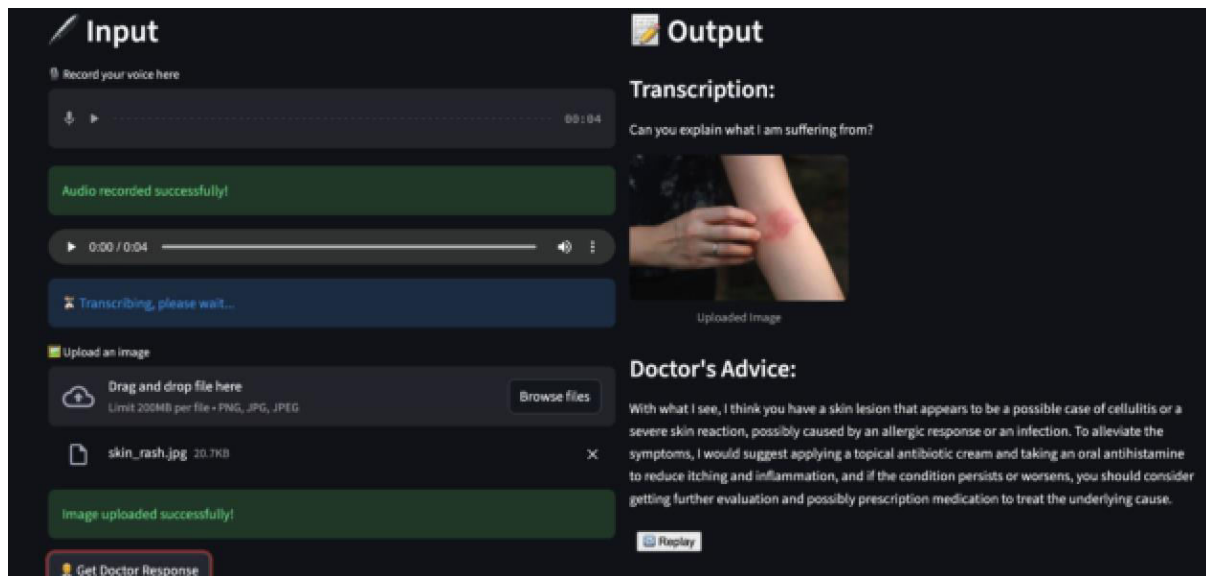


Figure 5: Combined mode result displaying voice transcription, uploaded skin image, and doctor's advice in the output column.

The Doctor's Advice output remained consistently bounded to three to five sentences, in line with the prompt engineering constraints used for safe public-health awareness messaging. Combined mode testing demonstrated that the system correctly merged voice transcriptions with uploaded images to produce richer, more contextually relevant guidance than either modality alone. Audio output via gTTS was generated within one second of text response completion in all tested scenarios, and the HTML audio block rendered correctly in the modern browsers used during evaluation.

Across complete testing sessions, the Streamlit interface performed consistently on both desktop and mobile browsers. Error handling was verified for edge cases including empty audio recordings, unsupported image formats, and network interruptions during API calls. In all cases, the application displayed informative error messages without crashing. Overall, the end-to-end pipeline demonstrated functional completeness and reliability suitable for public health awareness deployment.

V. CONCLUSION

Accessible and timely health awareness has become increasingly important in a world where many individuals still face barriers to immediate medical consultation, reliable health information, and early symptom interpretation. Delays in receiving even preliminary guidance can increase anxiety, encourage misinformation, and postpone professional care. In this context, intelligent conversational systems have the potential to serve as supportive front-line awareness tools by transforming complex AI capabilities into simple, user-friendly interactions for the general public.

This paper presented an AI-Driven Public Health Chatbot for Disease Awareness, a lightweight multimodal application that combines speech recognition, image-assisted reasoning, text generation, and audio synthesis within a unified Streamlit interface. The system enables users to describe symptoms verbally, upload images of visible conditions, and receive concise doctor-style awareness guidance in both textual and spoken form. By integrating these features into a single workflow, the platform addresses practical usability challenges that are often left unresolved in text-only symptom checker systems.

A central strength of the proposed chatbot lies in its multimodal design. Supporting voice-only, image-only, and combined interaction modes allows the platform to accommodate users with different communication preferences, literacy levels, and accessibility needs. This design is particularly valuable in public health settings, where users may not know the correct medical terminology to type, but may find it easier to speak naturally or share an image of a visible condition for contextual interpretation.

Another important contribution of the system is its modular architecture across four Python components responsible for interface control, audio transcription, reasoning, and audio reply generation. This separation of concerns improves maintainability, simplifies testing, and creates a clear foundation for future enhancement. The use of publicly accessible frameworks and APIs also demonstrates that meaningful AI-assisted health awareness systems can be developed without highly specialized hardware or complex deployment infrastructure.

Experimental observations showed that the chatbot operated reliably across all three interaction modes, with responsive end-to-end behavior and stable rendering of both textual and audio outputs. The integration of Whisper Large V3 Turbo, Llama 4 Scout, and gTTS through the Groq platform enabled practical real-time performance suitable for user-facing applications. These results indicate that modern multimodal AI services can be orchestrated effectively to deliver interactive health awareness experiences in a lightweight web-based environment.

From a safety and application perspective, the proposed system occupies an awareness-oriented role rather than a diagnostic one. The prompt engineering strategy was intentionally designed to constrain responses to short, cautious, and non-prescriptive guidance, while encouraging users to seek professional consultation for definitive diagnosis and treatment. This distinction is essential in healthcare-related AI systems, where responsible communication and clear scope boundaries are just as important as technical capability.

In contrast to clinical decision-support systems or disease prediction frameworks that often prioritize diagnostic depth, structured records, or specialized medical workflows, the proposed chatbot emphasizes accessibility, immediacy, and ease of interaction. Its contribution lies in bridging the gap between advanced AI models and everyday public-facing health communication. As a result, the system is especially relevant for awareness campaigns, first-level digital screening support, and preliminary guidance in settings where medical expertise may not be instantly available.

The broader significance of the platform lies in its potential social impact. A system of this kind can support public health outreach by extending basic awareness services beyond traditional clinical environments, helping users interpret symptoms more confidently and encouraging timely escalation when needed. Its web-based implementation also makes it suitable for wider deployment across educational institutions, community health programs, telehealth front ends, and other digital public service contexts.

Despite these strengths, the current implementation remains limited to awareness-level reasoning and does not incorporate verified electronic health records, specialist medical databases, or longitudinal patient context. Future work can therefore focus on multilingual interaction, domain-specific healthcare knowledge integration, structured symptom intake, conversation memory, secure authentication, and connections to telemedicine services or hospital directories. Such enhancements would increase both the inclusiveness and the practical utility of the system while preserving appropriate clinical safeguards.

In conclusion, the AI-Driven Public Health Chatbot for Disease Awareness demonstrates how state-of-the-art speech processing, multimodal reasoning, and audio response generation can be combined into an accessible and responsible public health support tool. By unifying voice, image, text, and speech output within a single streamlined application, the proposed system contributes to improved health awareness, broader digital accessibility, and the ongoing evolution of human-centered AI in healthcare communication.

REFERENCES

- [1] World Health Organization, "Noncommunicable diseases," WHO Fact Sheets, 2023.
- [2] A. K. Jha et al., "Use of electronic health records in U.S. hospitals," *New England Journal of Medicine*, vol. 360, no. 16, pp. 1628–1638, 2009.
- [3] D. Bates et al., "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," *Health Affairs*, vol. 33, no. 7, pp. 1123–1131, 2014.

- [4] P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv:1711.05225, 2017.
- [5] E. Alsentzer et al., "Publicly available clinical BERT embeddings," Proc. 2nd Clinical NLP Workshop (NAACL), 2019.
- [6] J. Lee et al., "BioBERT: a pre-trained biomedical language representation model," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020.
- [7] S. M. Laranjo et al., "Conversational agents in healthcare: a systematic review," JAMIA, vol. 25, no. 9, pp. 1248–1258, 2018.
- [8] A. Radford et al., "Robust speech recognition via large-scale weak supervision," Proc. ICML, 2023.
- [9] Groq Inc., "Groq LPU Inference Engine," Technical Overview, 2024. [Online]. Available: <https://groq.com>
- [10] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, 2017.
- [11] T. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017.
- [12] J. Li et al., "BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models," Proc. ICML, 2023.
- [13] OpenAI, "GPT-4 Technical Report," arXiv:2303.08774, 2023.
- [14] Meta AI, "Llama 4: Multimodal Foundation Models," Meta AI Research, 2024. [Online]. Available: <https://ai.meta.com>
- [15] Y. Wang et al., "Neural text-to-speech synthesis: a systematic review," IEEE Trans. Neural Networks and Learning Systems, vol. 34, pp. 1–20, 2022.
- [16] Google, "gTTS: Google Text-to-Speech Python Library," PyPI, 2023. [Online]. Available: <https://pypi.org/project/gTTS/>
- [17] T. Hoy, "OK Google, OKGoogle: Introducing new uses and users to voice assistants," Computers in Libraries, vol. 38, no. 1, pp. 26–29, 2018.
- [18] Streamlit Inc., "Streamlit: The fastest way to build data apps," 2024. [Online]. Available: <https://streamlit.io>
- [19] M. Singh and V. Sachan, "Development of AI-based health monitoring dashboard using Streamlit," IJACSA, vol. 12, no. 5, pp. 1–7, 2021.
- [20] E. Topol, Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. Basic Books, 2019.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 8.152