IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205262 |

Hotline AI - Voice Integrated Multilingual Chatbot to Assist National Welfare Schemes

Bharath S[1], Kaushik M[2], Vijayalakshimi A[3], Vidhya V[4]

U.G. Student, Department of I.T., Sri Venkateswara College of Engineering, Sriperumbudur, Tamil Nadu, India[1,2,3]

Head of the Department & Professor, Department of I.T., Sri Venkateswara College of Engineering, Sriperumbudur, Tamil Nadu, India[4]

ABSTRACT: A significant portion of the population remains unaware of the national welfare programs offered by the government, resulting in underutilization of these schemes and a failure to deliver their benefits to those in need. These schemes are essential for providing services to marginalized and underserved populations. To address this issue, there is an urgent need to raise awareness about these government schemes and resources. The HOTLINE AI – voice-integrated chatbot aims to tackle the problem by developing a conversational agent that educates citizens about the various schemes provided by the government, including details such as eligibility criteria and application processes. The chatbot is designed by leveraging advanced technology like Retrieval-Augmented Generation (RAG), which combines semantic search capabilities to analyze the query and content for a better understanding of the user's search intent and focuses on different parts of the text depending on the context. The HOTLINE AI employs multilingual transcription, allowing users to understand the answers in their own language. This initiative aims to empower citizens through accessible dialogue, fostering participation and maximizing the utilization of government resources for societal betterment.

KEYWORDS: Retrieval-Augmented Generation, Cohere large language model (LLM), Semantic word embeddings, Whisper -large audio transcription model.

I. INTRODUCTION

Natural Language Processing (NLP) is a dynamic and transformative field within Artificial Intelligence (AI) that focuses on enabling computers to understand, interpret, and generate human language. By combining insights from linguistics, computer science, and Machine Learning (ML), NLP powers a wide array of applications. These range from developing chatbots for real-time customer support to creating automated translation services that bridge language barriers. In the context of chatbots, NLP functions as a crucial translator, interpreting user inputs and generating natural language responses. This capability allows chatbots to engage in meaningful conversations, thereby improving user experience and operational efficiency. Chatbots, leveraging the power of NLP, have become prevalent across various sectors, providing seamless interactions and assistance. In customer service, they offer instant support and information on e-commerce platforms, enhancing user satisfaction. In the healthcare industry, chatbots assist with scheduling appointments and answering basic medical questions, streamlining patient care. Educational institutions also benefit from chatbot technology, using it to support students with enrollment processes and course-related inquiries. This widespread adoption of chatbots demonstrates their versatility and effectiveness in meeting the needs of different industries.

II. RELATED WORK

[1] The paper "GPTFX" proposes a novel GPT-3 based framework for mental health detection and explanation, outperforming traditional methods like LIME and SHAP in both classification accuracy and explanation quality. Fine-tuning GPT-3 models, particularly GPT Curie and GPT Davinci, provides more informative and comprehensive explanations. The method is cost-effective, utilizing API calls without additional infrastructure. Experimental results highlight its superiority in enhancing the understanding of mental health conditions. [2] The paper "Reinforcement Learning for Optimizing RAG for Domain Chatbots" proposes an RL-based approach to optimize token usage in RAG-based chatbots, improving performance and reducing costs. It trains an in-house retrieval model using infoNCE loss, outperforming public pre-trained models. The bot's answer quality is evaluated using GPT-4, with factual accuracy ratings converted into numeric rewards. This approach enhances chatbot accuracy and optimizes cost compared to the standard RAG pipeline. [3] The paper "Abstractive Text Summarization Using Transformer Based Approach" proposes

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205262 |

a transformer-based model, TLGA, for generating accurate and concise summaries of financial news. Utilizing a transformer encoder-decoder architecture, it outperforms baseline models on datasets like CNN/Daily Mail and XSum. The evaluation focuses on faithfulness and informativeness, highlighting TLGA's utility for investment decisions. The paper suggests that a deeper analysis of the model's challenges, strengths, and evaluation metrics would enhance understanding. [4] The paper "SHADE: Speaker-History-Aware Dialog Generation Through Contrastive and Prompt Learning" addresses the challenge of maintaining consistency in dialogue generation. The proposed SHADE system enhances personalized interactions by using contrastive learning and prompt learning to ensure a consistent speaking style and leverages external knowledge from Wikipedia to avoid factual errors. The model's performance is evaluated using ROUGE and METEOR metrics. While the approach shows promise in generating consistent and informative responses, it lacks comprehensive experimental analysis and comparison with existing methods. Further research is needed to validate its effectiveness and address potential ethical concerns and biases. [5] The paper "Feasibility of a voice-enabled automated platform for medical data collection: CardioCube" evaluates the use of a voice-enabled platform on Amazon Echo for collecting medical data from cardiovascular patients. In a study with 22 participants, CardioCube accurately gathered cardiovascular risk factors and medical history with 97.5% accuracy and current medications with 95.8% accuracy. The results affirm the platform's feasibility and effectiveness in integrating with electronic medical records (EMR) systems. The study concludes that CardioCube is a reliable tool for medical data collection in cardiovascular care. [6] The paper "Speech-Oriented Sparse Attention Denoising for Voice User Interface Toward Industry" highlights the benefits of voice user interfaces (VUI) in Industry 5.0 over traditional graphical interfaces. It introduces a novel speech-oriented sparse attention denoising scheme to enhance VUI performance in IoT and IIoT by reducing computational complexity while preserving speech signal dependencies. The paper effectively motivates its research but lacks quantitative evidence and detailed technical explanations, hindering comprehensive assessment. Additionally, it does not thoroughly discuss related works, impacting the understanding of its novelty. Providing quantitative metrics, technical details, and a literature review would improve clarity and credibility. [7] The paper "MSStyleTTS: Multi-Scale Style Modeling With Hierarchical Context Information for Expressive Speech Synthesis" introduces a model enhancing expressiveness in paragraph-based speech synthesis. It incorporates multi-scale style representation and context information, validated through ablation studies and subjective evaluations. While the model addresses coherence issues between sentences, a more extensive comparison with existing methods could provide clearer insights. Further exploration of how the model accommodates style variation in human recordings would enrich the study's impact. [8] The paper "Cognitive Hexagon-Controlled Intelligent Speech Interaction System" by Himanshu Chaurasiya explores the role of cognitive hexagon in ISI systems, emphasizing speech quality dimensions and AI integration. While acknowledging AI advancements and interdisciplinary approaches, it lacks critical analysis of limitations, reliability, and ethical implications. A more balanced discussion on positive aspects and concerns surrounding AI integration is suggested for future research.

III. MODULE DESCRIPTION

3.1 DATA COLLECTION:

In this process, the data was collected from the national welfare schemes website where the data for every scheme is being retrieved as text. The data is being classified according to their respective beneficiary domain like healthcare, education, agriculture, rural empowerment, etc. Each domain consists of list of schemes and the data related to the respective schemes. The data is incorporated into document in order to perform the Retrieval Augmented Generation (RAG) with the text documents or Pdf documents.

3.2 DATA PROCESSING MODULE:

The Data Processing Module involves cleaning and transforming raw data into a suitable format for model training, aiming to improve data quality, handle missing values, and ensure compatibility with analysis techniques. Text extraction utilizes the PyPDF library to read and extract text from PDFs, enabling efficient handling of document data for further processing. The extracted text is then split into smaller chunks using a TextSplitter to enhance model building and information retrieval efficiency. Embeddings convert text into numerical vectors, representing semantic similarity in a high-dimensional space, enabling the retrieval of relevant results. This process leverages methods like GloVe and Transformer-based models to capture semantic connections and improve NLP task performance.the cohere embeddings models is used for embedding generation withthe text.

3.3 VECTOR DATABASE MODULE:

Storage units, typically vector stores, play a crucial role in efficiently managing and retrieving large volumes of data by holding encoded text representations for rapid semantic similarity searches. Vector databases, optimized for storing and querying vector embeddings, excel at finding data based on semantic relevance rather than exact matches. This makes

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205262 |

them particularly useful for recommendation systems, image and document retrieval, and natural language processing, where understanding contextual relationships is essential. The embeddings from the cohere embedding model is being stored in the vector database.

3.4 SEMANTIC WORD EMBEDDINGS MODULE:

Sentence embeddings convert the meaning of sentences and paragraphs into fixed-size vectors, improving information retrieval (IR) tasks like semantic search, document retrieval, and question answering. By embedding queries and context paragraphs into vectors, and using cosine similarity to measure their distance, systems effectively retrieve relevant content. This approach enhances accuracy over traditional keyword searches, especially in handling nuanced user intent and extensive, unstructured text collections.

3.5 RAG CHAT MODEL

The Cohere LLM model, integrated into Retrieval-Augmented Generation (RAG) chatbots, significantly enhances their ability to provide accurate and contextually relevant responses. In the RAG framework, the chatbot retrieves relevant information from a large dataset and generates a response based on this augmented context. The process begins with question condensation, where previous user questions are condensed to retrieve relevant nodes from the vector store. The retrieved nodes are then used by the Cohere model to generate a well-informed and contextually appropriate response. The chat engine orchestrates this flow by condensing questions, retrieving information, and generating responses, ensuring accuracy and relevance. By combining retrieval with generation, the Cohere LLM ensures detailed and accurate answers, enhancing the chatbot's performance and user interaction.

3.6 SPEECH RECOGNITION MODULE:

Speech recognition technology, also known as Speech-to-Text (STT), converts spoken language into written text through a multi-step process involving signal processing and pattern recognition. Initially, audio input is captured, digitized, and processed to remove noise and enhance quality. The system extracts features from the audio, identifying characteristics like frequency and amplitude, which are segmented into phonemes. Advanced STT systems employ deep learning models, such as recurrent neural networks (RNNs) or transformer-based models, to decode these phonemes into words. Using a language model, the system constructs coherent text, correcting for context and grammar. The Whisper-Large model enhances this process with its multilingual support and high-quality speech recognition capabilities.

3.7 SPEECH SYNTHESIS MODULE:

Speech synthesis technology, also known as Text-to-Speech (TTS), converts written text into spoken words through a series of complex processes. The system begins with text normalization, cleaning and preparing the input text. A linguistic analysis module then parses the text to identify syntactic and prosodic features, breaking it into phonemes and determining pitch and duration. A synthesis module, often leveraging deep neural networks, generates natural-sounding speech by predicting and producing waveforms directly from the text. The final output is a continuous audio stream that mimics human speech. The Whisper-Large model is chosen for its advanced capabilities, offering multilingual support and producing high-quality, natural-sounding speech output.

3.8 MULTILINGUAL TRANSLATION MODULE:

Multilingual translation for text and voice facilitates communication across different linguistic groups by converting written or spoken content from one language to another using advanced machine learning models. For text translation, the Google Translate model provides near-instantaneous translations, leveraging large multilingual datasets to ensure accuracy and contextual relevance. For voice translation, the Whisper-large model is used, combining automatic speech recognition (ASR), machine translation (MT), and text-to-speech (TTS) synthesis. The Whisper model transcribes spoken words into text, translates the text into the target language, and uses TTS to convert the translated text back into spoken form. The process involves loading and standardizing the audio, converting it into a log-Mel spectrogram for frequency analysis, and decoding the spectrogram to produce transcribed text. The gTTS (Google Text-to-Speech) is then used to generate natural-sounding speech from the text, ensuring the translated content is accessible in audio format. This integrated approach allows for seamless and accurate multilingual communication.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205262 |

IV. RESULTS

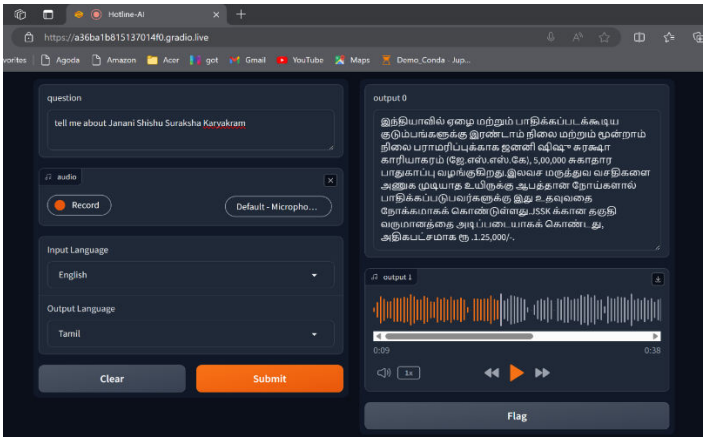

Fig 4.1 Language Transcription from English to Tamil

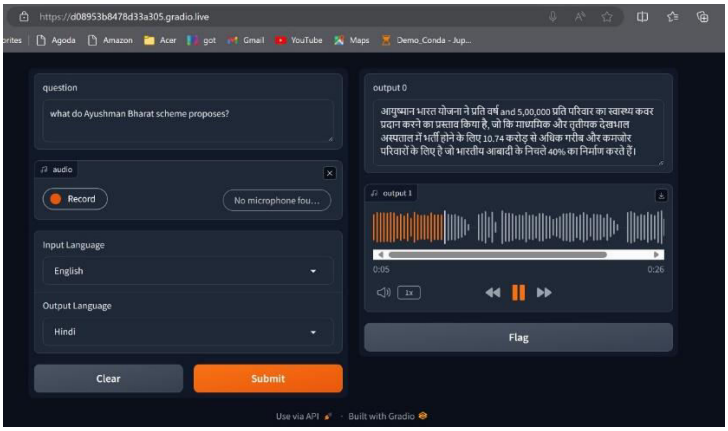

Fig 4.2 Language Transcription from English to Hindi

V. CONCLUSION AND FUTURE WORK

The HOTLINE AI voice-integrated chatbot significantly advances in bridging the information gap regarding government welfare programs. By leveraging Retrieval-Augmented Generation (RAG) technology, the chatbot provides accurate and contextually relevant information, ensuring users' queries are thoroughly understood and addressed. The integration of semantic search capabilities allows the system to provide precise details about various schemes, including eligibility criteria and application processes. This detailed guidance is crucial for marginalized and underserved populations, who often face barriers to accessing such information.The use of multilingual transcription ensures inclusivity, allowing users from diverse linguistic backgrounds to interact with the chatbot in their preferred language. This enhances the accessibility of information and fosters a sense of empowerment among citizens, encouraging greater participation in welfare programs. Overall, the HOTLINE AI chatbot is a powerful tool for maximizing the utilization of government resources. Future work for the HOTLINE AI initiative includes refining the conversational agent's capabilities, expanding its reach across diverse devices and interfaces, integrating with popular messaging platforms, providing real-time updates on welfare schemes, and implementing personalized recommendation systems. Rigorous evaluation and user feedback mechanisms will be essential to continually improve user satisfaction and effectiveness. Overall, ongoing innovation and optimization aim to maximize the impact of HOTLINE AI in raising awareness and empowering citizens to access government resources effectively.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205262 |

REFERENCES

1. Alzubaidi L, Zhang J, Humaidi A.J. (2021), 'Review of deep learning: concepts, CNN architectures, challenges, applications, future directions J Big Data', pp. 53.
2. Bateman, J. A, Hois J, Ross R, & Tenbrink T. (2010), 'A linguistic ontology of space for natural language processing' – Artificial Intelligence, 174(14), pp. 1027–1071.
3. Bhardwaz S., & Godha R. (2023), 'Svelte.js: The Most Loved Framework Today' – 2nd International Conference for Innovation in Technology (INOCON), pp. 1-7.
4. B. N. D. Santos, R. M. Marcacini, S. O. Rezende (2021), 'Multi-Domain aspect extraction using bidirectional encoder representations from transformers' – IEEE Access Vol. 9, pp. 91604–91613.
5. Chandrasekaran, D., & Mago, V. (2021), 'Evolution of Semantic Similarity—A Survey' – ACM Computing Surveys, 54(2), pp. 1–37.
6. Chang, Y. (2023), 'A survey on evaluation of large language models' – arXiv.org, Computer Science, Computation and Language.
7. Chiara Valentina Misischia, Flora Poecze, Christine Strauss (2022), 'Chatbots in customer service: Their relevance and impact on service quality' – Procedia Computer Science, vol. 201, pp. 421-428, ISSN 1877-0509.
8. David Vilar and Marcello Federico (2021), 'A Statistical Extension of Byte-Pair Encoding' – Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT), pp. 263–275, Bangkok, Thailand. Association for Computational Linguistics.
9. Dettmers T., Pagnoni A., Holtzman A., & Zettlemoyer L. (2023), 'QLORA: Efficient Finetuning of Quantized LLMS' – arXiv.org, Computer Science, Machine Learning.
10. Dhingra B., Shallue C.J., Norouzi M., Dai, A.M. & Dahl G.E. (2018), 'Embedding Text in Hyperbolic Spaces'.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details