Voice First Ai Vernacular Assistant for Government Schemes and Entitlements

Ms. M. Janani, M.Sc[1]., Mr.A.J. Somesh Kumar[2], Mr.A. Sanjay[3], Mr.M. Vishwa[4]
Mr.G.Vivek Gandhan[5]
[1,3,4,5]*Under Graduate Students, Department of Bachelor of Artificial Intelligence and Data Science Salem college of Engineering and Technology, Salem, Tamilnadu , India*
[2]*Assistant Professor, Department of Artificial Intelligence and Data Science, Salem college of Engineering and Technology, Salem, Tamilnadu, India*

***Abstract*—Accessing government schemes is often hindered by digital vernacular speech to connect citizens with relevant government schemes. By prioritizing local languages and robust error-literacy gaps, language barriers, and complex user interfaces. This paper presents a Voice-First AI Vernacular Assistant designed to democratize access to government entitlements. Developed as a Single Page Application (SPA) using React 19 and Vite, the system leverages the Web Speech API and a custom Natural Language Processing (NLP) engine to process multi-lingual speech. Key innovations include regular expression-based speech disfluency cleaning for English, Hindi, and Tamil, alongside dynamic phonetic language detection. A Profile Extraction Engine derives structured user data from unstructured speech, fueling a precise, point-based scoring algorithm. To ensure reliability, strict confidence guardrails prevent AI hallucination. The resulting platform offers a highly accessible, intuitive, and accurate solution for seamless scheme discovery.**

***Index Terms*—Voice Assistant, Vernacular AI, E-Governance, Natural Language Processing, Zero-Hallucination, React 19, Speech Disfluency Cleaning.**

I . INTODUCTION

The rapid digitalization of government services has significantly improved administrative efficiency. However, a digital divide persists for rural populations, the elderly, and non-English speakers who struggle with text-heavy, complex navigation structures. Voice-user interfaces (VUIs) present a natural, accessible alternative. This project introduces a multilingual, voice-first application capable of understanding natural, handling, the system bridges the gap between complex governmental frameworks and the everyday citizen.

II. RESEARCH PROBLEM

- Digital & Linguistic Barriers: Complex, text-heavy portals exclude citizens with low digital literacy, particularly those who communicate exclusively in vernacular languages rather than formal English.
- Natural Speech Processing Failures: Existing systems struggle to accurately interpret conversational speech, failing to filter out regional colloquialisms and disfluencies (e.g., "matlab", "appo").
- Unstructured Input Handling: Current search tools lack the capability to dynamically extract structured demographic data (e.g., age, occupation, social category) from unstructured voice queries.
- Risk of Misinformation (Hallucination): When processing ambiguous vernacular inputs, basic search systems or bots are highly prone to "hallucinating" or guessing incorrect scheme details, creating critical reliability issues for government entitlements.

III. OBJECTIVES

- To develop a highly responsive, voice-first Single Page Application (SPA) for discovering government entitlements.

- To engineer a custom NLP pipeline capable of cleaning localized speech disfluencies and dynamically detecting vernacular languages (English, Hindi, Tamil).
- To implement a reliable Profile Extraction Engine that maps conversational input to structured data (Age, Gender, Categories).
- To design a precise scheme-matching scoring algorithm safeguarded by strict confidence guardrails to eliminate false information.

IV. PROPOSED METHODOLOGY

The proposed methodology centers on developing a Single Page Application (SPA) using React 19 and Vite, integrated with the Web Speech API for real-time voice processing. The system's core is a custom Natural Language Processing (NLP) pipeline that utilizes regex-based disfluency cleaning to filter vernacular fillers across English, Hindi, and Tamil, alongside dynamic phonetic language detection. A dedicated Profile Extraction Engine parses these cleaned transcripts to extract structured demographic data—such as age, gender, and social category—which a point-based scoring algorithm then uses to rank relevant government schemes. To guarantee accuracy and prevent misinformation, strict zero-hallucination guardrails halt the matching process for low-confidence inputs, triggering safe fallback messages, while an embedded local metrics dashboard tracks system performance for continuous optimization.

V. METHODOLOGY

The architectural foundation of the proposed system is built upon a modern Single Page Application (SPA) framework using React 19 and Vite, ensuring high-performance client-side rendering and modular scalability. To facilitate seamless navigation across various interfaces—such as the Home, Scheme List, and Scheme Details pages—the system employs React Router DOM. Crucial to its multilingual capability is the global state management implemented via the Context API (Language Context), which efficiently handles the user's selected language and translations (t()) without the performance overhead of prop-drilling. The core voice interactions are encapsulated within custom React hooks, specifically the use Voice Assistant hook. This modular approach directly integrates the Web Speech API to manage continuous microphone listening, real-time speech-to-text transcription, and text-to-speech auditory feedback. The resulting interface is highly responsive, providing clear visual feedback through micro-animations and maintaining consistent visual theming via Vanilla CSS custom properties.

Following the initial auditory capture, the unstructured transcript undergoes rigorous processing through a custom Natural Language Processing (NLP) pipeline tailored for complex vernacular speech. An intelligent text-cleaning function applies localized Regular Expressions to strip conversational disfluencies and regional fillers (e.g., "um", "matlab", "appo") across English, Hindi, and Tamil inputs, operating alongside a dynamic phonetic language detection module. The cleaned data is then fed into a specialized Profile Extraction Engine (profileAnalyzer.js), which utilizes multi-language trigger banks to systematically extract structured demographic parameters like Age, Gender, and Social Category. These data points drive a point-based scoring algorithm that prioritizes direct category matches to rank government schemes. To guarantee data integrity and eliminate AI misinformation, the system relies on strict zero-hallucination guardrails based on calculated confidence scores. If the engine evaluates an input with "Low" confidence, it safely halts the matching process and triggers a pre-defined fallback prompt rather than hallucinating an inaccurate recommendation, while an embedded local metrics dashboard tracks these interactions for continuous system refinement.

5.1 System Overview:

- Frontend Architecture: Built as a modern Single Page Application (SPA) utilizing React 19 and Vite for highly optimized client-side rendering.
- Navigation: Routing is handled seamlessly via React Router DOM.
- State Management: The Context API globally manages language selection and translations, effectively eliminating prop-drilling.
- Voice Integration: Core functionalities—such as Web Speech API interactions, microphone

listening, and text-to- speech playback—are cleanly encapsulated within custom React hooks (e.g., useVoiceAssistant).

5.2 System Concept & Application Model:

- Conversational Interface: Operates on an intuitive input- output model where users speak their queries naturally.
- Real-Time Processing: The incoming speech stream is instantly captured, transcribed, and fed into the Natural Language Processing (NLP) pipeline.
- Dynamic Profiling: Instead of relying on rigid exact- keyword matches, the system analyzes the transcript to build a comprehensive "User Profile," which is then cross- referenced against a scheme database for highly targeted retrieval.

5.3 Technology & Functional Features:

- Intelligent Speech Cleaning: Utilizes localized regular expressions to actively strip speech fillers (e.g., "um", "matlab", "appo") across multiple languages, significantly improving extraction accuracy.
- Profile Extraction Engine: Parses the cleaned transcripts using multi-language trigger banks to extract structured demographic data (e.g., student, farmer, senior citizen).
- Scoring & Guardrails: Employs a point-based ranking algorithm that prioritizes direct category matches over fallback keywords. It utilizes Confidence Guardrails (High,

Medium, Low) to prevent AI hallucination; if confidence is "Low," it triggers a safe fallback message.

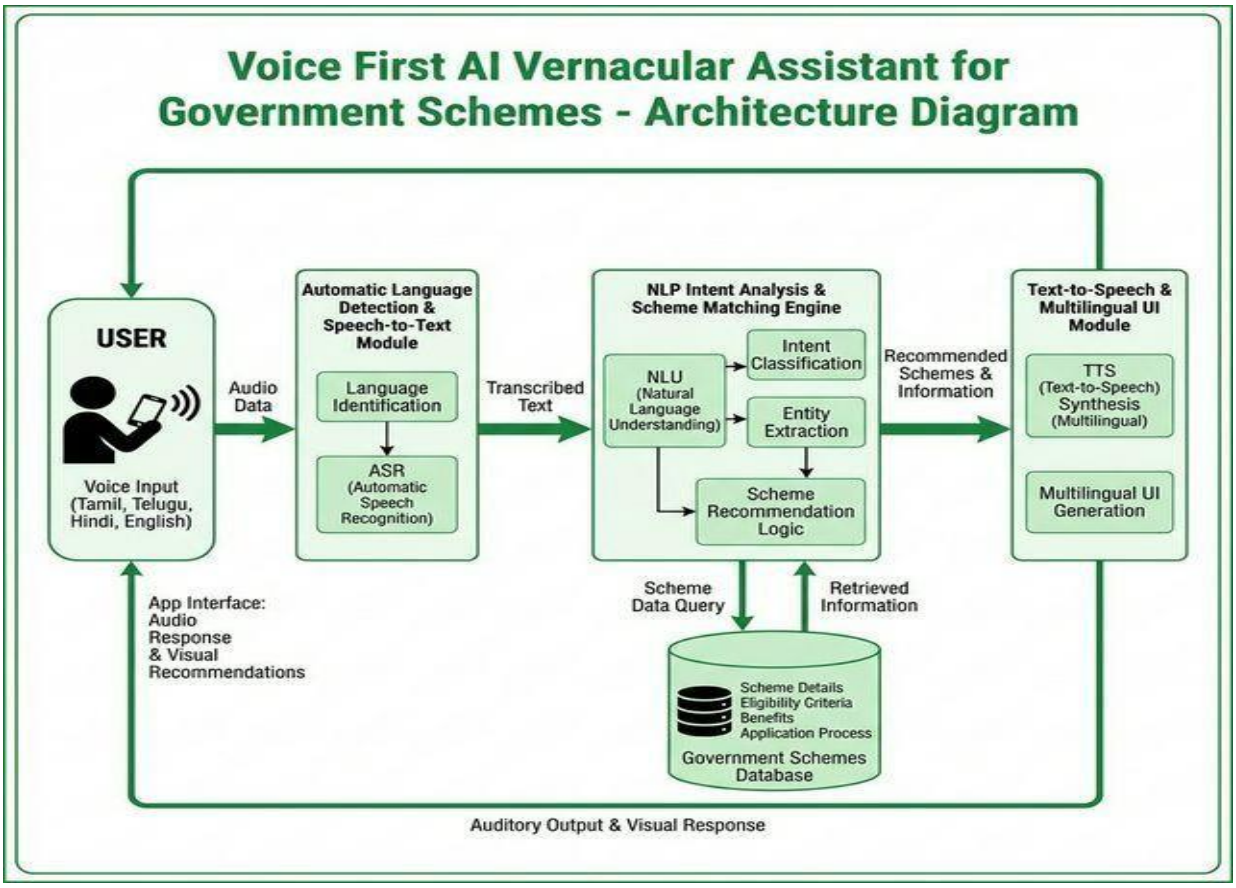
- UI/UX & Metrics: Features Vanilla CSS with custom properties for consistent theming and micro-animations (e.g., pulse effects for the microphone). An embedded Enterprise Metrics Dashboard tracks attempts, fallback rates, and success click for local optimization.

5.4 Benefits to Users & Local Stores (Service Centers):

- For Users: Provides a frictionless, typing-free experience that empowers marginalized or technologically inexperienced citizens to independently discover their entitlements in their native tongue.
- For Local Stores/CSCs: Reduces the initial inquiry burden on local e-Seva operators. Citizens arrive knowing their eligible schemes, allowing service centers to focus purely on the application submission process rather than manual scheme hunting.

5.5 Impact & Future Scope:

- Social Inclusion: Significantly lowers the barrier to entry for e-governance, promoting broader digital equity.
- Language Expansion: Future iterations will expand vernacular phonetic detection to encompass a wider variety of regional languages.
- LLM Integration: Plans include integrating advanced Large Language Models (LLMs) to process deeper conversational context.
- End-to-End Fulfillment: Future scope includes linking the assistant directly to government application submission APIs to create a complete, start-to-finish entitlement platform.



VI. CONCLUSION

The Voice-First AI Vernacular Assistant successfully demonstrates how modern web technologies and targeted NLP techniques can solve critical accessibility issues in e- governance. By moving away from rigid text searches and embracing natural, multi-lingual speech processing—fortified by intelligent disfluency cleaning and zero-hallucination guardrails—the application provides a highly accurate, user- centric tool. This system not only streamlines scheme discovery but also establishes a scalable framework for inclusive digital public infrastructure.

REFERENCES

[1] Gowtham Ramesh, et al. (2022). Jugalbandi: A Voice- First Multilingual Chatbot for Government Services (Technical Report/Case Study)

[2] Kaushal Bhogale, Abhigyan Raman, Tahir Javed, Sumanth Doddapaneni (2023). Vistaar: Diverse Benchmarks and Training Sets for Indian Language ASR

[3] Yash Madhani, Sumanth Doddapaneni, et al. (2023). IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for all 22 Scheduled Indian Languages

[4] AI4Bharat Team (Neeraj Varshney, et al.) (2024). Airavata: Introducing Hindi Instruction-tuned LLM

[5] Le Fang, et al. (2025). Speech LLMs in Low-Resource Scenarios: Data Volume Requirements and the Impact of Pretraining