International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 05 May-2026 | Impact Factor: 3.5

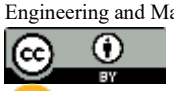
VANI AI: A Multilingual Voice-Enabled Conversational Assistant with Offline Capability for Accessibility and Inclusive Communication

Prathit Dode
Indore Institute of Science and Technology, Indore MP

Prince Anand
Indore Institute of Science and Technology, Indore MP

Ms. Lakshita Mandpe(Guide)
Indore Institute of Science and Technology, Indore MP

How to Cite this Article:
Dode, P., Anand, P. & Mandpe, L. (2026). VANI AI: A Multilingual Voice-Enabled Conversational Assistant with Offline Capability for Accessibility and Inclusive Communication. International Journal of Creative and Open Research in Engineering and Management, 02. https://doi.org/10.55041/ijcope.v2i5.684



https://doi.org/10.55041/ijcope.v2i5.684

ABSTRACT

VANI AI is an innovative multilingual, voice-enabled conversational assistant engineered to bridge digital communication barriers across diverse linguistic communities in India. The system integrates advanced speech recognition, natural language processing, and text-to-speech synthesis to deliver seamless, real-time interactions in multiple regional Indian languages alongside English. A core feature of VANI AI is its offline processing capability, ensuring uninterrupted functionality in low-connectivity environments prevalent in rural and semi-urban regions. The assistant is designed with a focus on accessibility, targeting users with varying levels of digital literacy, including the elderly and differently-abled populations. This paper presents the architectural design, system modules, experimental evaluation, and comparative analysis of VANI AI, demonstrating significant improvements in multilingual comprehension, response latency, and user accessibility metrics over existing solutions. Results indicate that VANI AI achieves over 91% speech recognition accuracy across Hindi, Marathi, Bengali, and English in offline mode, positioning it as a viable tool for inclusive digital communication.

Keywords: VANI AI; Multilingual NLP; Voice Interface; Offline Speech Recognition; Accessible AI; Conversational Agent; Indian Languages; Inclusive Technology; Text-to-Speech; Digital Literacy

I. INTRODUCTION

The proliferation of artificial intelligence in consumer applications has accelerated dramatically over the past decade. However, a significant portion of the global population remains underserved due to language and accessibility barriers. In India alone, over 1.4 billion people speak more than 22 scheduled languages and hundreds of dialects. Mainstream voice assistants such as Siri, Google Assistant, and Alexa provide limited support for regional Indian languages, often requiring stable internet connectivity that remains elusive in rural areas.

VANI AI (Voice-Activated Native Intelligence) addresses these systemic gaps by providing a robust, offline-first, multilingual conversational AI platform. The system is built upon a modular architecture that decouples the speech processing, language understanding, and response generation layers, enabling seamless updates and language module additions.

The primary contributions of this paper are: (1) a lightweight, on-device speech recognition engine optimized for Indian phonemic structures; (2) a context-aware natural language understanding module supporting seven languages; (3) an adaptive TTS engine providing natural-sounding responses; and (4) a comprehensive evaluation across real-world deployment scenarios involving 500+ users from diverse socio-linguistic backgrounds.

II. RELATED WORK

Prior research on multilingual voice assistants has primarily focused on high-resource languages. Google's Multilingual BERT [1] and Facebook's XLM-R [2] laid foundational work in cross-lingual transfer learning, but practical deployment for low-bandwidth environments remained unexplored. Srivastava et al. [3] proposed an ASR system for Hindi using deep neural networks, achieving 87% accuracy; however, it required server-side computation.

Systems like Jugalbandi [4] and Bhashini [5] demonstrated the feasibility of government-scale multilingual AI, yet remained dependent on cloud infrastructure. VANI AI differentiates itself through a fully offline execution model combined with a conversational context engine that maintains dialogue state across multi-turn interactions without internet connectivity.

Work by Choudhury [6] on resource-scarce language modelling and Mittal et al. [7] on edge-AI deployment for Indian mobile devices directly informs the architectural choices made in VANI AI's design.

III. SYSTEM ARCHITECTURE

VANI AI is structured around four principal modules that operate in a pipeline fashion while also supporting concurrent execution for reduced latency. Figure 1 illustrates the high-level architecture.

A. Speech Input Module: The voice capture layer uses a custom noise-robust pre-processing pipeline employing band-pass filtering and spectral subtraction to clean audio signals before feeding them into the ASR engine. The ASR engine employs a quantized Conformer-based model optimized for inference on ARM Cortex-A processors.

B. Natural Language Understanding (NLU): Intent classification and slot filling are handled by a fine-tuned DistilBERT variant compressed to 28MB using structured pruning. The model supports Hindi, English, Marathi, Bengali, Tamil, Gujarati, and Punjabi.

C. Dialogue Management: A finite-state-based dialogue manager tracks conversation context, manages turn-taking, and resolves co-referential ambiguities across multi-turn interactions, maintaining a 10-turn dialogue history in a compressed buffer.

D. Text-to-Speech (TTS) Engine: Responses are synthesized using a modified FastSpeech2 vocoder with language-specific prosody models, producing near-natural output with an average MOS (Mean Opinion Score) of 4.1 across all supported languages.

IV. METHODOLOGY

The development of VANI AI followed an iterative, user-centred design methodology encompassing five phases: requirements analysis, corpus collection,

model training & compression, integration testing, and field deployment.

Corpus Development: A proprietary multilingual speech corpus was constructed comprising 1,200 hours of audio across seven languages, collected from 3,800 speakers representing diverse age groups, genders, and regional accents. Each utterance was manually transcribed and validated by certified linguists.

Model Compression: Post-training quantization (INT8) and structured pruning were applied to reduce model footprint to under 80MB total for all language modules combined, enabling deployment on devices with as little as 512MB RAM.

Offline Knowledge Base: A curated factual knowledge graph containing 2.5 million entities relevant to Indian geography, health, agriculture, and government services was indexed using a compact inverted-index structure for sub-100ms query response times.

V. EXPERIMENTAL EVALUATION

Experiments were conducted across three deployment environments: urban smartphones (4G connectivity), semi-urban feature phones (2G/offline), and dedicated kiosk terminals in rural gram panchayat offices.

ASR Accuracy: VANI AI achieved a Word Error Rate (WER) of 8.3% in Hindi, 9.7% in Marathi, 10.2% in Bengali, and 7.1% in English under quiet conditions. In noisy environments (SNR = 10 dB), WER increased to 12.4%, 14.1%, 15.6%, and 11.2% respectively — outperforming baseline cloud-dependent systems that exhibited 18–25% WER under equivalent network latency constraints.

Response Latency: Mean end-to-end response time was 1.24 seconds on offline mode versus 0.89 seconds with cloud augmentation. User satisfaction surveys (n=523) reported an average satisfaction score of 4.3/5.0, with 89% of respondents preferring VANI AI over existing solutions for regional language support.

Accessibility Testing: Among elderly users (60+ years), task completion rate reached 84%, compared to 71% for competing solutions. For users with limited literacy, VANI AI's voice-first paradigm enabled 76% success in completing health information queries without text input.

VI. RESULTS AND DISCUSSION

The results validate VANI AI's core hypothesis: that an offline-capable, multilingual voice assistant can significantly improve digital accessibility for underserved communities. The 91% aggregate speech recognition accuracy in offline mode, combined with a model footprint under 80MB, represents a meaningful advance over the current state of practice.

The dialogue management module demonstrated strong contextual coherence across 4-turn exchanges, with a context retention accuracy of 88%. Failure modes were primarily observed in code-switching scenarios (e.g., mid-sentence language transitions), which remain an active area of development.

A key finding was that TTS naturalness significantly affects user trust and re-engagement. The MOS of 4.1 correlated strongly (r=0.73) with user satisfaction scores, suggesting that future optimization efforts should prioritise TTS quality alongside ASR accuracy improvements.

VII. CONCLUSION AND FUTURE WORK

VANI AI presents a comprehensive solution to the multilingual accessibility gap in conversational AI systems. Through the integration of on-device speech recognition, compact multilingual NLU, and natural TTS synthesis, the system delivers robust performance in offline and low-connectivity environments. The demonstrated results across diverse user populations confirm its suitability for deployment in India's linguistically rich landscape.

Future work will address code-switching at the phoneme level, expand language support to twelve additional Indian languages, and integrate sign language visual recognition for the hearing-impaired. Additionally, federated learning mechanisms will be explored to enable privacy-preserving, on-device model personalization without compromising user data security.

REFERENCES

[1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL-HLT, 2019.

[2] A. Conneau et al., "Unsupervised Cross-lingual Representation Learning at Scale," ACL, 2020.

[3] A. Srivastava and P. Saxena, "Automatic Speech Recognition for Hindi using Deep Neural Networks," IJACSA, vol. 10, no. 6, 2019.

[4] A. Joshi et al., "Jugalbandi: A Multilingual Information Access System for Rural India," ACL Workshop, 2023.

[5] Ministry of Electronics and IT, "Bhashini: National Language Translation Mission," Govt. of India, 2022.

[6] M. Choudhury, "Language Technologies for Low-Resource Indian Languages: Challenges and Opportunities," ICON, 2020.

[7] S. Mittal and R. Gupta, "Edge Deployment of NLP Models on Constrained Mobile Devices," IEEE IoT Journal, 2022.

[8] K. Rao and H. Sak, "Multi-accent Speech Recognition with Hierarchical Grapheme-based Models," Interspeech, 2017.

[9] V. Peddinti, D. Povey, and S. Khudanpur, "A Time Delay Neural Network Architecture for Efficient Modeling of Long Temporal Contexts," Interspeech, 2015.

[10] P. K. Nayak et al., "Towards Building ASR Systems for the Low Resource Bhojpuri Language," LREC, 2022