

# Vyapar AI - Voice-First Business Assistant for Small Enterprises

\*Darshan Nitin Bhanushali, \*Yash Amit Bhavsar, \*Janvi Kishor Pandav

Ganesh Wadmare -Faculty Guide

Department of Artificial Intelligence and Data Science

K.J. Somaiya Institute of Technology

Sion (E), Mumbai - 400 022, India

**Abstract:** India has a large number of micro, small, and medium enterprises (MSMEs) over 63 million in total, almost all of which use manual processes and paper records for their day-to-day activities. Adoption of technologies is hindered by lack of digital literacy, cumbersome user interfaces, lack of adequate support for regional languages, and dependence on internet access. Vyapar AI tackles these problems by offering a voice-first, offline-capable mobile application for Android based on natural voice commands in Hinglish (Hindi + English) allowing owners of small businesses to handle their inventories, invoicing, client data, payment history, and business intelligence. The software combines React Native CLI frontend, Python FastAPI backend, and on-device speech recognition with Vosk engine, thus completely removing dependence on internet connectivity. Speech is analyzed with a deterministic rule-based NLP system, enabling prompt and predictable interactions without resorting to cloud-based APIs. Invoices can be shared and payment reminders sent directly from the application through deep linking for WhatsApp, thus avoiding use of external API services altogether. All business intelligence, such as sales summary, best-selling products, and stock shortages, is processed by the AI engine offline. This paper presents design considerations, architecture description, mathematical models, implementation details, and evaluation of features provided by Vyapar AI along with a comparison of closely-related applications and directions for further research. *Index Terms:* Voice-First Mobile Applications, Offline Speech Recognition, MSME, Rule-Based NLP, Hinglish, Android, FastAPI, WhatsApp Integration, Business Intelligence, Vosk.

**Keywords:** MSME Digitization, Voice-First Applications, Offline Speech Recognition, Hinglish NLP, Rule-Based Natural Language Processing, Android Application Development, React Native, FastAPI, Vosk Speech Engine, Offline AI Systems, Business Intelligence, Inventory Management, Invoice Automation, WhatsApp Integration, Edge Computing, Small Business Automation, Regional Language Support, Mobile Computing, AI for MSMEs, Offline-Capable Applications

## I. INTRODUCTION

There are over 63 million MSMEs in India, which contribute a total of about 30% to the country's GDP and provide employment to over 110 million people. While the economic impact is significant, over 95% of these businesses do not use any form of digital management technology. The MSME owner, usually a shopkeeper, trader, or small-scale manufacturer, uses a manual register, manual calculations, and other rudimentary forms of accounting that are prone to errors and inefficiencies.

The reasons for the lack of penetration of management software among these businesses are clear. Factors such as low digital literacy levels, complex UI/UX design, insufficient support for Hindi and regional languages, and dependence on consistent internet connectivity hinder the adoption of business management software by this demographic. Current solutions targeting larger businesses assume a level of technological proficiency and internet availability that is unavailable in MSMEs.

This paper proposes Vyapar AI, a voice-based mobile application allowing small business owners to manage their operations end-to-end using natural spoken Hindi and English

commands without requiring the internet for its core functionalities.

The proposed application is built using a state-of-the-art full-stack development framework using a React Native CLI front-end (Android) and a Python FastAPI back-end. The on-device speech-to-text conversion is accomplished using the Vosk offline speech-to-text engine, eliminating the need for the internet for speech recognition. Recognized text is parsed using a rule-based NLP engine to identify user intent and extract entities such as customer names, product quantities, and monetary amounts. The back-end executes the business logic to perform tasks such as managing inventory, creating invoices, managing customers, tracking payments, generating AI-driven insights, and integrating with WhatsApp using deep links.

The following are the contributions made by this paper:

- (1) Voice-first MSME management mobile application that supports Hinglish commands;
- (2) Deterministic rule-based NLP engine for recognizing nine categories of business intent;

- (3) WhatsApp integration using deep links and no external API calls;
- (4) End-to-end solution with features validation in Android.

## II. BACKGROUND AND RELATED WORK

### A. MSME Landscape in India

The Indian MSME industry, with an estimated 63 million businesses, is responsible for about 30% of India's gross domestic product and provides employment to around 110 million people [13][15]. However, despite its significance, the use of digital business management software by the MSME sector is not yet widespread due to factors such as language, computer literacy, and access to the internet. According to studies in similar economies like sub-Saharan Africa, voice-based systems play a significant role in easing the technology adoption process for traders and shopkeepers [13][14].

### B. Voice-Based Billing and Accounting

Studies into voice-driven billing processes show that the automatic tokenization and bill generation process is not only technically possible but also saves a significant amount of time [1]. Extensions to multilingual voice driven bill generations targeting local Indian languages have been proven to be effective at deployment by shopkeepers [2]. The studies investigating voice-driven accounting processes including GST calculation and ledgers support the demand for voice-first accounting solutions for SMEs in India [3].

### C. Offline and Privacy-Preserving Voice Assistants

The study of offline voice assistants provides strong evidence for the feasibility of building speech-to-text functionality entirely on the device level. It was found that rule-based natural language processing systems using solely on-device processing were capable of providing fast and privacy-preserving interactions without the need for any connection [4][5][6]. The systematic study of offline speech recognition limitations in areas with poor connectivity was used as a source when deciding on the on-device speech engine [7].

### D. Voice-Controlled Inventory Systems

A series of empirical studies on voice-controlled stock control shows that it helps save time and improves the accuracy of stock control. One such study notes the 87% saving in time spent manually entering data and 94% increase in stock control accuracy [8]. Additionally, voice controlled stock inquiries and updates as well as prior art patents for voice activated retail stock management provide further evidence for the development of the inventory system in Vyapar AI [9][10][11].

### E. Voice-First Systems in Emerging Markets

The relevant literature to consider here in terms of context is centred around the development of voice-first mobile assistants for small businesses that function in contexts where connectivity is poor, and levels of literacy are low. Additionally, industry reviews of voice-first technology in India reveal that there is a significant yet untapped potential in voice for the MSME market segment within India.

Research on voice-driven transaction systems has proven that voice can successfully log monetary figures and client codes.

## III. LITERATURE REVIEW AND COMPARATIVE ANALYSIS

Table I presents a structured comparison of Vyapar AI against the most closely related systems identified in the literature. The comparison evaluates each system across voice modality, offline capability, backend architecture, and key distinguishing strengths and gaps. Vyapar AI is the only system in this comparison that combines full offline voice interaction in Hinglish, self-contained Android APK deployment, WhatsApp deep-link communication, and local AI-driven business insights within a single application.

TABLE I  
 Comparative Analysis of Related Systems

System / Reference	Voice Modality	Offline Support	Backend / NLP	Key Strengths	Gaps vs Vyapar AI
Dukawalla [13]	Voice-first	Limited – cloud-based	Cloud generative	Field-tested in Africa SMBs	Not fully offline; no local analytics
Visual + Voice Inventory [8]	Voice + CV	Hybrid	Custom backend + STT	87% less manual entry	No invoicing or WhatsApp
Offline Android Assistant [4]	Voice-first	Fully offline	Pattern-match rules	Full offline; lightweight	No business domain; no invoices
Offline AI Voice Asst. [5]	Voice-first	Mainly offline	Rule-based NLP	Privacy-focused; low-resource	Not tailored to SMB operations
Vyapar App (commercial) [20]	GUI-first	Online-oriented	Cloud-based	Rich feature set; WhatsApp	Not voice-first; not fully offline
Vyapar AI (this work)	Fully voice-first (Hinglish)	Core features fully offline	Rule-based NLP + local analytics	Full-stack APK; no external APIs	User-study benchmarks ongoing

## IV. RESEARCH GAP ANALYSIS

### A. Absence of an Integrated Voice-First MSME Solution

There have been many efforts toward voice assistants in the context of small businesses; however, none of them integrate the requirements of voice-first, offline capabilities, Hinglish language, inventory management, invoicing, customer management, payments tracking, and business insights into one integrated Android application. Vyapar AI solves that problem.

### B. Rule-Based NLP for Deterministic Business Commands

Current solutions utilize intent recognition models that leverage machine learning algorithms, which require significant computational power and rely on cloud APIs. A rule-based NLP model used in the current study offers a much faster response time, as it returns a deterministic result based on a predefined set of commands. It does not require any GPU hardware nor Internet connection, providing a unique value proposition for the given use case.

### C. WhatsApp as Default Communication Channel

The current literature does not take into consideration the fact that WhatsApp messenger plays a prominent role in MSMEs' daily communication routines in India. This project incorporates invoice sending and payment reminders through WhatsApp deep linking. It leverages an existing communication channel without requiring any third-party API integration, which is a new addition to the field of research.

### D. Offline-First Architecture for Low-Connectivity Markets

Although cloud-based solutions exist for managing small businesses, their operation relies on stable internet connectivity, which cannot be guaranteed in certain regions. An offline-first approach adopted by Vyapar AI sends transcriptions only to a local backend server, thus minimizing dependence on network infrastructure.

## V. PROPOSED SYSTEM: VYAPAR AI

### A. System Overview

Vyapar AI uses a layered and offline-first client-server system architecture designed to reduce dependency on networks and optimize local computation. The entire system functionality is expressed as follows:

$$F(\text{audio}) = \text{TTS}(\text{BL}(\text{NLP}(\text{STT}(\text{audio}))))$$

Here, STT converts the input audio signal into text using the pretrained offline acoustic model; NLP then translates the transcription into {intent, entities}; BL applies business logic to the input and generates a corresponding response; finally, TTS produces speech output based on the response locally.

### B. Layer 1: On-Device Speech Recognition

The speech recognition process is performed on the device itself, with Vosk offline engine. The only information sent to the server is the transcribed text string. The latency of STT does not exceed 500 ms at all times, effectively bypassing any network round-trip for voice capturing and operating without connectivity.

### C. Layer 2: Rule-Based NLP Engine

The NLP engine uses keyword pattern matching with no machine learning dependencies, enabling  $O(n)$  intent detection relative to command length with response times under 100 ms. Nine intent categories are supported: create\_invoice, check\_stock, add\_product, update\_stock, add\_customer, send\_reminder, get\_sales, get\_top\_products, and get\_low\_stock. Table II illustrates example Hinglish voice commands for each intent.

TABLE II

NLP Intent Categories and Example Commands

Intent	Example Voice Command
create_invoice	"Ramesh ko 500 ka invoice bana do"
check_stock	"Maggi ka stock kitna hai"
add_product	"Rice add karo 20 quantity"

Intent	Example Voice Command
update_stock	"Tata Tea ka stock 50 karo"
add_customer	"Naya customer add karo Ramesh"
send_reminder	"Ramesh ko payment reminder bhejo"
get_sales	"Total sales kitna hai"
get_top_products	"Top product kaunsa hai"
get_low_stock	"Kaunse products ka stock kam hai"

### D. Layer 3: FastAPI Backend and Business Logic

The business logic is implemented using the Python framework FastAPI in modular fashion with a dedicated layer per functionality. The FastAPI backend includes five layers of business logic covering Inventory Management, Invoice Creation & PDF Generation, Customer Management, Payment Tracking, and Sales Analytics respectively. Asynchronous design of FastAPI allows efficient processing of concurrent requests through non-blocking I/O. FastAPI backend can be run on local devices as well as local network server.

Total invoice amount calculation formula:  $\text{Total}(I) = \sum (Q_i \times P_i)$ ;  $i = 1$  to  $n$  (where  $Q_i$  - quantity,  $P_i$  - unit price of each line item).

Inventory is decreased automatically upon generating an invoice. An alert is triggered once  $\text{Stock} < \text{threshold}$ .

Month sales calculation:  $\text{Sales}(\text{month}) = \sum \text{Total}(I_j)$  over all invoices in that month.

Exposure of pending payments:  $\text{Pending} = \sum \text{Total}(I_j)$  over all invoices with  $\text{payment\_status} = \text{'unpaid'}$ .

### E. Layer 4: React Native CLI Frontend

The Android frontend is implemented in React Native CLI (no Expo) and delivers eight core screens: Dashboard, Voice Input, Inventory, Invoice, Customer, Payment Tracking, Analytics, and Settings. The Voice Input screen provides a microphone button with real-time waveform visualisation, displays transcribed text and detected intent, and speaks the backend response via Android Native TTS. The Dashboard presents summary cards for total sales, pending payments, product count, and low-stock alerts.

### F. Layer 5: WhatsApp Integration

Vyapar AI integrates WhatsApp via wa.me deep links without requiring the WhatsApp Business API. When a user requests a payment reminder or invoice share, the backend generates a pre-filled message string and the app opens WhatsApp with the customer's phone number and message pre-populated. This approach leverages the most widely used communication channel in the Indian MSME market at zero API cost.

### G. Tech Stack Summary

Table III summarises the complete technology stack deployed in Vyapar AI

**TABLE III**  
**Technology Stack**

Component	Technology
Frontend	React Native CLI (Android only, no Expo)
Backend	FastAPI (Python, async, modular)
Database	SQLite (development) / PostgreSQL (production)
Speech-to-Text	Vosk (offline, on-device, ~40 MB model)
Text-to-Speech	Android Native TTS (on-device)
NLP Engine	Rule-based intent detection + entity extraction
PDF Generation	ReportLab / WeasyPrint (backend)
WhatsApp Integration	wa.me deep links (no API required)
Deployment	Self-contained Android APK

## VI. RESULTS AND FEATURE VALIDATION

The application was implemented and tested on Android hardware. All twelve core features were successfully implemented and validated. Speech processing runs entirely on-device with STT latency under 500 ms. Rule-based NLP operates under 100 ms for all intent categories. Table IV summarises the feature validation status for the current implementation.

**TABLE IV**  
**Feature Validation Summary**

Feature	Status	Offline Capable
Voice-controlled inventory management	Implemented	Yes
Invoice creation and PDF generation	Implemented	Yes
Customer management	Implemented	Yes
Payment tracking (paid / unpaid)	Implemented	Yes
WhatsApp invoice sharing (deep link)	Implemented	Requires WhatsApp
WhatsApp payment reminders (deep link)	Implemented	Requires WhatsApp
AI business insights – total sales	Implemented	Yes
AI business insights – top products	Implemented	Yes
Low-stock alerts	Implemented	Yes
Hinglish voice commands	Implemented	Yes
On-device STT (Vosk)	Implemented	Yes
Android TTS response output	Implemented	Yes

## VII. CHALLENGES AND OPEN ISSUES

### A. NLP Generalisation

While the NLP rule-based system performs quickly and deterministically, it is limited to a predetermined number of intents and lexical patterns. The ambiguity in some phrases or those requiring contextual interpretation could result in a classification error, requiring the use of workarounds or reprompts from the user. Using a simple on-device sequence classifier model is seen as an improvement within reach.

### B. Scalability for Multi-Device Deployments

While the commands in Hinglish (Hindi + English) are covered completely, India is known to have a multitude of languages. Hence, a significant share of the MSMEs that constitute our user base use Marathi, Gujarati, Bengali, Tamil, or Telugu as their mother tongue. Training an acoustic model that covers such languages is no small feat.

### C. Language Coverage

While Hinglish (Hindi + English) commands are fully supported, India's linguistic diversity means that a large share of the MSME target audience speaks Marathi, Gujarati, Bengali, Tamil, or Telugu as their primary language. Extending acoustic model support to these languages is a significant undertaking requiring region-specific training data.

### D. GST and Regulatory Compliance

There is currently no functionality regarding the computation of GST, the preparation of GSTIN based invoices, or summary statements for monthly returns. Compliance with India's Goods and Services Tax policy is a must for formal adoption in business. This feature is a high priority for the next phase of development.

### E. Formal User Evaluation

User studies with relevant participants and latency benchmarks are underway. Without a formal usability test conducted, the extent to which the voice system eases the workload for MSME owners compared to current GUI systems is uncertain.

## VIII. CONCLUSION

The Vyapar AI system provides a proof-of-concept that voice is a viable primary means of interacting with business tools in informal markets. Through offline speech recognition, deterministic and rule-based natural language processing (NLP) techniques, and a fully featured business management backend system, this project has made cutting-edge business management solutions powered by artificial intelligence available to less digitally literate individuals who do not have reliable internet connectivity and prefer to interact in Hinglish.

Using offline speech recognition on the device removes the limitation of lack of internet connectivity, while the deterministic nature of NLP used guarantees quick response time with high accuracy of results, making it explainable and predictable, all without needing cloud AI services [4][5][6]. Integration with WhatsApp through deep links allows for

easy invoicing and payment reminders without paying any third-party API fees [13][15].

FastAPI microservices allow for incorporation of more business features such as GST filing, expense tracking, and supplier management in subsequent iterations, thereby making this solution more robust than existing offerings. Overall, this project has shown that a thoughtful application of artificial intelligence technology and engineering design can go a long way in improving business efficiency through voice interfaces.

## IX. FUTURE SCOPE

Future improvements to the Vyapar AI include:

- (1) incorporating GST into CGST/SGST/IGST calculations and invoice formats based on GSTIN number.
- (2) supporting additional languages such as Marathi, Gujarati, Bengali, Tamil, and Telugu.
- (3) using intent detection with machine learning techniques with an easy-to-implement on-device classifier.
- (4) multi-device synchronization through an optional cloud layer managed by the user.
- (5) using barcodes and QR codes for quick updates to inventory status.
- (6) providing UPI payment deep link functionality combined with WhatsApp for end-to-end invoice-to-payment processing.
- (7) developing a supplier management system for closing the loop in procure-to-pay processes.
- (8) providing predictive analysis of low inventory based on sales velocity trends.
- (9) developing a voice-enabled onboarding process.
- (10) iOS support using React Native's cross-platform capabilities.

## X. LIST OF ABBREVIATIONS

1. MSME – Micro, Small and Medium Enterprise
2. NLP – Natural Language Processing
3. STT – Speech-to-Text
4. TTS – Text-to-Speech
5. API – Application Programming Interface
6. APK – Android Package Kit
7. GDP – Gross Domestic Product
8. GST – Goods and Services Tax
9. UPI – Unified Payments Interface
10. UI/UX – User Interface / User Experience

## XI. CONFLICT OF INTEREST

The authors declare no conflict of interest. No external funding was received for this study.

## XI. ACKNOWLEDGEMENT

The authors gratefully acknowledge the guidance and support of the faculty at the Department of Artificial Intelligence and Data Science, K.J. Somaiya Institute of Technology, Mumbai. All three authors contributed substantially to the design, implementation, analysis, and reporting of the system presented in this paper.

## XIII. REFERENCES

- [1] IRJET. Voice Based Billing System. International Research Journal of Engineering and Technology, Vol. 6, Issue 3. Available: <https://www.irjet.net/archives/V6/i3/IRJET-V6I31136.pdf>
- [2] IJARST. Voice Based Billing System – Multilingual, Voice-Activated Billing for Shopkeepers. Available: <https://ijarst.co.in/Paper25159.pdf>
- [3] AI Accountant. Voice-Based Accounting Entry: The Game-Changer for Indian Finance Teams. Available: <https://www.aiaccountant.com/blog/voice-based-accounting-entry-india>
- [4] Academia.edu / IJRASET. Design and Development of a Privacy-Preserving Offline AI Assistant for MSMEs. Available: <https://www.ijraset.com/best-journal/design-and-development-of-a-privacy-preserving-offline-ai-assistant>
- [5] IJRASET. Building an Offline Virtual Voice Assistant using AI and NLP. Available: <https://www.ijraset.com/best-journal/building-an-offline-virtual-voice-assistant-using-ai-and-nlp>
- [6] IJSAT. Enhancing User Autonomy and Privacy: An Offline Virtual Voice Assistant. Available: <https://www.ijer.org/research-paper.php?id=9830>
- [7] SciTePress. Offline Speech Recognition Development – Systematic Review. Available: <https://www.scitepress.org/papers/2018/67880/67880.pdf>
- [8] IJIRT. An AI-Based Visual and Voice-Controlled Inventory System. Available: <https://ijirt.org/Article?manuscript=180458>
- [9] IJIRT. Visual and Voice-Assisted Inventory Automation. Available: [https://ijirt.org/publishedpaper/IJIRT188367\\_Paper.pdf](https://ijirt.org/publishedpaper/IJIRT188367_Paper.pdf)
- [10] Polsri e-prints. Inventory Management System using AI Voice. Available: <http://eprints.polsri.ac.id/22452/>
- [11] IJERT. Voice-Based Indexing System in Warehouse Management. Available: <https://www.ijert.org/voice-based-indexing-system-in-warehouse-management>
- [12] IJRASET. An Intelligent Smart Retail System for Voice-Guided Stock Monitoring. Available: <https://www.ijraset.com/best-journal/an-intelligent-smart-retail-system>
- [13] arXiv. Dukawalla: Voice Interfaces for Small Businesses in Africa. Available: <https://arxiv.org/pdf/2505.05170.pdf>
- [14] Engineering for Change. AI Voice Assistant for a Textile Marketplace App in Rural India. Available: <https://www.engineeringforchange.org/projects/voice-recognition-feature-for-rural-textile-producers/>
- [15] PwC India. Voice First – Research Insights Report. Available: <https://www.pwc.in/assets/pdfs/research-insights/2019/voice-first.pdf>
- [16] IJRPR. Voice-Recognition-based Money Transaction System Survey, Vol. 4, Issue 4. Available: <https://ijrpr.com/uploads/V4ISSUE4/IJRPR11454.pdf>
- [17] ThinkDebug. From Swiping to Speaking: Why Voice-First Apps Are the Future. Available: <https://thinkdebug.com/from-swiping-to-speaking-why-voice-first-apps-are-the-future-in-2025/>
- [18] Vosk. Offline Speech Recognition API for Python, Android, and More. Available: <https://alphacephei.com/vosk/>
- [19] FastAPI. Modern, Fast Web Framework for Building APIs with Python. Available: <https://fastapi.tiangolo.com/>
- [20] Vyapar App. Business Billing and Accounting Software for SMBs. Available: <https://vyaparapp.in/>
- [21] VoiceInvoicer. Voice-Driven Invoice Creation App. Available: <https://apps.apple.com/>

- [22] B. Henkels, C. D. Schultz, A. De Keyser, and D. Mahr, "The sound of progress: AI voice agents in service," *J. Service Manage.*, vol. 37, no. 1, pp. 1–32, Feb. 2026, doi: 10.1108/JOSM-06-2025-0269.
- [23] "Does your voice assistant remember? Analyzing conversational context recall and utilization in voice interaction models," arXiv preprint arXiv:2502.19759, 2025.
- [24] Y. Mei, Y. Zheng, D. Xu, and Y. Long, "SHNU multilingual conversational speech recognition system for INTERSPEECH 2025 MLC-SLM Challenge," in *Proc. Interspeech 2025*, pp. 1–5, Jul. 2025, arXiv preprint arXiv:2507.03343.
- [25] "NTU Speechlab LLM-based multilingual ASR system for Interspeech MLC-SLM Challenge 2025," arXiv preprint arXiv:2506.13339, Jun. 2025.
- [26] T. Alumäe and A. Fedorchenko, "TalTech systems for the Interspeech 2025 ML-SUPERB 2.0 Challenge," in *Proc. Interspeech 2025*, pp. 1–4, Jun. 2025, arXiv preprint arXiv:2506.01458.
- [27] "The multimodal information based speech processing (MISP) 2025 Challenge: Audio-visual diarization and recognition," in *Proc. Interspeech 2025*, pp. 1–5, May 2025, arXiv preprint arXiv:2505.13971.
- [28] "CompanionCast: A multi-agent conversational AI framework with spatial audio for social co-viewing experiences," arXiv preprint arXiv:2512.10918, Dec. 2025.
- [29] "Enhancing speech emotion recognition with graph-based multimodal fusion and prosodic features for the Speech Emotion Recognition in Naturalistic Conditions Challenge at Interspeech 2025," in *Proc. Interspeech 2025*, pp. 1–5, Jun. 2025, arXiv preprint arXiv:2506.02088.
- [30] "TriageSim: A conversational emergency triage simulation framework from structured electronic health records," arXiv preprint arXiv:2603.10035, 2025.
- [31] "RelayS2S: A dual-path speculative generation for real-time dialogue," arXiv preprint arXiv:2603.23346, 2025.