International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 11 | Nov - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

VoxIntel: A Multimodal Intelligent Assistant for Tulu Language

P Shreya Shetty, Thanvi Rai N, Thaswi K J, Vaishnavi C, Dr. Akshaya D Shetty

[1]Department of Artificial Intelligence and Machine Learning,
Vivekananda College of Engineering and Technology, Puttur, Karnataka

[2]Department of Artificial Intelligence and Machine Learning,
Vivekananda College of Engineering and Technology, Puttur, Karnataka

[3]Department of Artificial Intelligence and Machine Learning,
Vivekananda College of Engineering and Technology, Puttur, Karnataka

[4]Department of Artificial Intelligence and Machine Learning,
Vivekananda College of Engineering and Technology, Puttur, Karnataka

[5]Assistant Professor, Department of Artificial Intelligence and Machine Learning,
Vivekananda College of Engineering and Technology, Puttur, Karnataka

Abstract - Tulu is a Dravidian language of coastal Karnataka and northern Kerala. It is normally written in Kannada script, and very little of it exists as digital text. Because of this, mainstream AI models tend to reply to Tulu questions in Kannada and to invent facts they have never seen. This paper presents VoxIntel: A Multimodal Intelligent Assistant For Tulu Language, supporting conversation, translation, lessons and speech on a standard laptop without any cloud AI service or API key. The system combines a Retrieval-Augmented Generation (RAG) knowledge base of 2,069 native-reviewed records with a locally hosted Gemma 3 4B model fine-tuned with QLoRA on a 12,029-entry Tulu dataset. When a user enters a query, the RAG knowledge base is searched first using embedding similarity. If a matching record is found, its reviewed answer is returned; only when no record matches does the Gemma 3 4B model generate the response. A Kannada detector, an automatic rewrite step and an output guard keep generated replies in Tulu. On 119 unseen replies the fine-tuned model scored 45.7 chrF against 20.2 for the earlier model, and its word-level F1 rose from 0.154 to 0.482. Repeated-word loops disappeared. The router chose the correct answer source for 88 of 100 test messages, and 1,602 automated tests pass.

Key Words: *Tulu Language, Low-Resource Language, Offline Assistant, RAG, Gemma 3 4B, QLoRA.*

1. INTRODUCTION

Tulu is a Dravidian language spoken in Dakshina Kannada and Udupi in Karnataka and in Kasaragod in Kerala. It is written in Kannada script, but many everyday words differ from Kannada. For example, the words for “what”, “how”, “and” and “water” all have their own Tulu forms. Very little Tulu text exists online, so large AI models saw almost no Tulu while they were being built.

As a result, a general chatbot that receives a Tulu question usually replies in Kannada. The reply looks correct on screen but is in the wrong language, and a native reader notices it at once. Such a model can also state facts it was never given, which makes it hard to trust. Every good Tulu sentence therefore has to be collected and checked by people.

This paper presents VoxIntel, an offline Tulu assistant for chat, translation, learning and voice that runs on an ordinary laptop with 8 GB of memory under one strict rule: no cloud AI service and no API key. Its main idea is to answer from a reviewed RAG knowledge base first and to call the local Gemma 3 4B language model only when the knowledge base has no matching answer. Native speakers approve all Tulu content, and a Kannada detector, a rewrite step and an output guard keep generated replies in Tulu.

The paper compares the fine-tuned model with the earlier one and reports how accurately the router chooses the right answer source. The remaining sections cover the literature survey, the proposed system, methodology, results, future scope and conclusions.

2. LITERATURE SURVEY

Prabhu et al. [1] developed a spoken language identification system for the coastal Karnataka languages using transfer learning and multi-view learning. The work shows that speech technology for the region’s under-resourced languages is feasible.

Sani et al. [2] studied emotion-semantic-aware neural machine translation between Indo-Aryan and Dravidian languages using transfer learning. Telrandhe and Kadu [3] reviewed the techniques, challenges and future scope of machine translation for low-resource Indo-Aryan languages. Both studies highlight how scarce data limits language technology for such languages.

Balipa and colleagues [4] developed a natural language processing model for Hindi-to-Tulu translation, one of the few efforts that support Tulu through computational methods. Balipa et al. [5] proposed a rule-based machine translation framework for low-resource language pairs. Shetty et al. [6] built TuluAgri Assist, a voice assistant for agriculture, showing that Tulu voice assistance is useful but serving a narrow domain.

Lau [7] examined the language options of Alexa, Google

Assistant and Siri and argued that mainstream voice assistants reproduce monoglossic language ideologies, favouring a few dominant languages. Mannonov et al. [8] discussed bridging speech and text through multimodal AI for language understanding. Related Dravidian-language work includes multimodal hate speech detection using text and audio features [9] and machine-learning-based Kannada-to-English translation [10].

Across these studies, two gaps stand out. Most Tulu work focuses on translation or a single domain rather than open conversation, and mainstream assistants do not support Tulu at all. These gaps motivate VoxIntel, a multimodal, multi-domain conversational assistant for Tulu.

3. PROPOSED SYSTEM

Most chat assistants and translators are built for widely used languages, and Tulu has almost no dedicated tools. A general AI model given a Tulu question usually answers in Kannada, because the two languages share a script, and it can state facts it was never given. The earlier VoxIntel model showed the same weaknesses: it was trained only on single questions, and on 119 test replies it got stuck repeating one word 8 times, used English letters 6 times, leaked tags or bullet lists 6 times and replied in another script 5 times.

The proposed VoxIntel system therefore prefers reviewed answers over open-ended generation. A React and Vite web app provides the chat, translate, learn and voice screens, and a Python FastAPI backend coordinates the processing. Ollama runs the Gemma 3 4B model and the search embeddings, and all programs communicate only on the user's laptop.

When the user enters a query, the backend first searches the RAG knowledge base of 2,069 reviewed records. If the query matches a record, the reviewed answer is returned as written. If the query does not match any record, the fine-tuned Gemma 3 4B model generates the response, which is then passed through language control before it is shown (Fig -1).

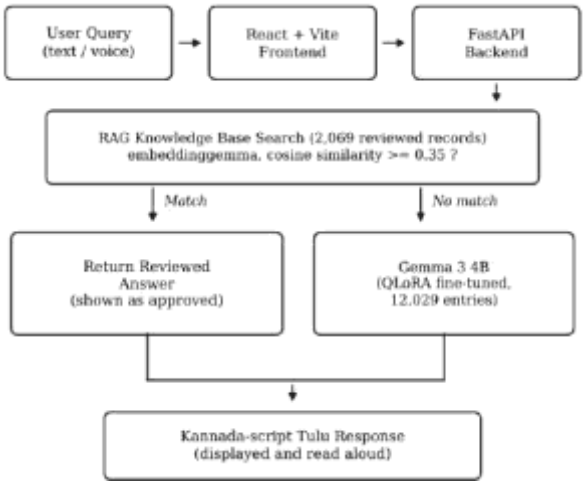


Fig -1: Architecture of the Proposed System.

4. METHODOLOGY

The proposed VoxIntel system follows a sequence of stages that prepare reviewed data, search the RAG knowledge base, generate a reply only when no match is found, and control the language of the output.

4.1 Dataset Preparation and Preprocessing

Two data resources are used. A Tulu dataset of 12,029 entries is used to fine-tune the Gemma 3 4B model for language understanding and response generation, and 2,069 records form the RAG knowledge base that supplies reviewed answers. The collected data undergoes data collection and organisation, followed by cleaning and removal of duplicate or incorrect entries. The Tulu text is then reviewed and corrected to improve quality and consistency, and the RAG records are structured for retrieval.

4.2 Query Processing Pipeline

Each query is received by the FastAPI backend, which keeps the last 12 messages of the chat as memory. Messages that must never be generated, such as a crisis message, the current time or a calculation, are handled first by fixed reviewed replies or computed tools. All other queries are searched in the RAG knowledge base. If a match is found, the stored answer is returned; otherwise the query goes to the Gemma 3 4B model (Table -1). Calculator results, weather readings and retrieved facts are not replayed as chat history, so old data never enters a new conversation.

Table -1: Response paths

Path	Condition	Answer source
RAG match	Cosine ≥ 0.35	Reviewed answer as written
Local model	No RAG match	Gemma 3 4B (QLoRA)
Tools	Time, weather, calculation	Computed, not guessed

4.3 RAG Knowledge Base Retrieval

The RAG knowledge base stores 2,069 reviewed Tulu records with Tulu and English question forms. Search uses the embeddinggemma model with 768-dimension vectors. The cosine similarity between a query vector q and a stored question vector d is:

$$cos(q, d) = (q \cdot d) / (\|q\| \|d\|)$$

A record is accepted when cos(q, d) ≥ 0.35, and its approved answer is returned exactly as written, so the model is not involved and the result is always the same.

4.4 Fine-Tuned Language Model

When the knowledge base has no matching record, the query is sent to Gemma 3 4B Instruct, fine-tuned with QLoRA on the 12,029-entry Tulu dataset, including multi-turn conversations. Fine-tuning trains 29.8 million parameters (0.69 percent). Training was carried out on free Kaggle T4 GPUs, and the model was converted with llama.cpp to a 4-bit file of 2.50 GB that runs in Ollama. The model receives the Tulu system prompt, up to 12 earlier messages and any matching dictionary words. The earlier model remains installed as a backup.

4.5 Language Control

Every model reply is checked by a Kannada detector that uses 68 Kannada words and 20 word endings. If N is the number of words in the reply and N_K the number of Kannada markers among them, the reply is rewritten once when:

$$R = N_K / N \geq 0.25$$

An output guard then removes English letters and false claims that the assistant runs in the cloud. Replies are always shown in Kannada-script Tulu.

4.6 Native-Speaker Review and Terminology

All Tulu content is approved by native speakers, and every new Tulu line added to the app, such as labels, safety replies and recall replies, is listed for native review. Reviewer corrections applied to the approved data comprise 13 word fixes, 6 answer replacements and 2 withdrawn answers, each with recorded evidence. They follow patterns such as the Kannada h sound becoming p in Tulu, so the approved data contains no Kannada markers. Two typing problems are fixed automatically: an invisible joiner character after the virama, and the Kannada digit zero typed in place of the anusvara.

4.7 Voice Interaction and Web Interface

Users can speak a question with the microphone and have replies read aloud. The React 19 and Vite 8 frontend provides chat, translate, learn and voice screens, and Supabase handles the optional login with approval and an admin portal. A one-click Windows setup installs Python, Node.js and Ollama, downloads the model and starts the app.

5. RESULT AND DISCUSSION

VoxIntel was evaluated as a complete system covering routing, reviewed answers, tools, the language model, login and voice. Queries answered from the RAG knowledge base or by a computed tool always give the same result because the model is not involved. Only queries with no match reach Gemma 3 4B, and its replies pass the Kannada check and the output guard before they are shown.

5.1 Implementation Setup

The key configuration used in the tests is summarised in Table -2.

Table -2: Key system parameters

Parameter	Configuration / Value
Frontend	React 19 + Vite 8
Backend	FastAPI (Python 3.10+)
Local inference	Ollama (loopback only)
RAG knowledge base	2,069 reviewed records
Embeddings	embeddinggemma (768-dim), threshold 0.35
Language model	Gemma 3 4B, QLoRA fine-tuned, 4-bit GGUF, 2.50 GB
Training data	12,029-entry Tulu dataset
Trainable parameters	29.8 M (0.69 %)
Voice	Microphone input, replies read aloud
Login	Supabase (optional)

5.2 Model Comparison

The fine-tuned model was compared with the earlier model on 119 unseen replies using character n-gram F-score (chrF) and word-level F1. Word-level F1 is computed from precision P and recall R over matching words:

$$F1 = 2PR / (P + R)$$

Table -3: Earlier model vs. new model

Metric	Earlier	New model
chrF	20.2	45.7
Word-level F1	0.154	0.482
Repeated-word loops	8	0

The new model more than doubles the chrF score and triples the word-level F1. The earlier model also used English letters 6 times, leaked tags or bullet lists 6 times and replied in another script 5 times on the same test replies. The gain is attributed to training on genuine multi-turn conversations containing only approved Tulu.

5.3 Router Accuracy

The router chose the correct answer source for 88 of 100 test messages, an accuracy of 88 %. Answers from the RAG knowledge base are shown exactly as approved, which keeps them repeatable.

5.4 Automated Testing

A total of 1,602 automated tests pass: 1,495 backend, 75 frontend and 32 database tests. Browser tests also cover sign-up, login, approval and the admin portal. The tests check routing order, that reviewed replies are shown exactly as approved, that approved data has no Kannada markers, that training conversations contain only approved Tulu, and that no test question leaks into training. They also verify that the training prompt equals the app's prompt, memory limits, name memory, the crisis reply, dictionary near-matches and the login approval rules.

6. FUTURE SCOPE

- **Larger Tulu Dataset:** Expanding the Tulu dataset and the RAG knowledge base, and further fine-tuning the language model.
- **Regional Dialects:** Supporting different regional dialects of Tulu.
- **Better Speech:** Enhancing voice recognition and text-to-speech for more natural interaction.
- **Learning Tools:** Adding learning materials, quizzes and pronunciation exercises.
- **More Platforms:** Extending the application to mobile devices and other operating systems.

7. CONCLUSIONS

VoxIntel demonstrates the use of artificial intelligence to build an offline conversational assistant for the Tulu language. By checking a native-reviewed RAG knowledge base of 2,069 records first and using a Gemma 3 4B model fine-tuned on a 12,029-entry Tulu dataset only when no record matches, the system provides improved Tulu responses through chat, translation, learning and voice interaction. The new model scored 45.7 chrF against 20.2

for the earlier version, raised word-level F1 from 0.154 to 0.482 and removed repeated-word loops, while the router selected the correct source for 88 of 100 messages.

Local processing also keeps conversations on the user's device, preserving privacy. Overall, VoxIntel provides a practical approach to supporting Tulu communication and promoting the digital preservation of regional languages.

ACKNOWLEDGEMENT

The authors thank Dr. Akshaya D Shetty for her guidance throughout this project, and the Department of Artificial Intelligence and Machine Learning, Vivekananda College of Engineering and Technology, Puttur, for the facilities and support provided.

REFERENCES

1. S. Prabhu, D. V, S. Kamath, M. H, K. K. Prakasha, and D. A. Dinesh, "Spoken Language Identification System for Detecting Coastal Karnataka Languages Using Transfer Learning and Multi-View Learning," Engineering Research Express, vol. 8, no. 1, p. 015203, 2026.
2. S. Sani, S. V. Gangashetty, K. Samudravijaya, and A. K. Dubey, "Emotion-Semantic-Aware Neural Machine Translation Between Indo-Aryan and Dravidian Languages via Transfer Learning," IEEE Access, vol. 14, pp. 4953–4969, 2026.
3. N. Telrandhe and R. Kadu, "Machine Translation Models for Low-Resource Indo Aryan Languages: A Comprehensive Review of Techniques, Challenges and Future Scope," in Proc. 3rd DMIHER Int. Conf. Artificial Intelligence in Healthcare, Education and Industry (IDICAIHEI), Wardha, India, 2025, pp. 1–6.
4. A. K, A. C, and M. Balipa, "Natural Language Processing Model for Hindi to Tulu Translation," in Proc. Int. Conf. Advances in Modern Age Technologies for Health and Engineering Science (AMATHE), Shivamogga, India, 2025, pp. 1–6.
5. M. Balipa, A. K, Ananya, and M. Murugappan, "A Rule-Based Machine Translation Framework for Low-Resource Language Pairs," in Proc. 4th Int. Conf. Sentiment Analysis and Deep Learning (ICSADL), Bhimdatta, Nepal, 2025, pp. 969–974.
6. S. Shetty, S. Shetty, and N. V. Bangera, "TuluAgri Assist: Personalized Voice Assistant for Agriculture and Farming Support," in Proc. IEEE Int. Conf. Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), Mangalore, India, 2025, pp. 554–557.
7. M. Lau, "Alexa, Google Assistant, and Siri's language options: how voice assistants reproduce monoglossic language ideologies," AI & Society, vol. 40, pp. 6831–6846, 2025.
8. A. Mannonov, L. H. Jasim, A. S. Anvarovna, W. Suryasa, and A. Nayak, "Bridging Speech and Text using Multimodal Artificial Intelligence for Next-Gen Language Understanding," in Proc. Int. Conf. Computational Innovations and Engineering Sustainability (ICCIES), Coimbatore, India, 2025, pp. 1–6.
9. M. N. C, R. Natarajan, and S. R, "Multimodal Hate Speech Detection in Dravidian Languages Using Text and Audio Features," in Proc. Int. Conf. Communication and Signal Processing (ICCSP), Melmaruvathur, India, 2025, pp. 148–152.
10. S. Sidramappa, M. V. Reddy, and S. K, "Machine Learning-Driven System for Kannada-to-English Text Translation," in Proc. 2nd Int. Conf. Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2025, pp. 1–6.