"KANNADA LANGUAGE GENERATIVE AI FOR FARMERS REVOLUTIONIZING AGRICULTURE WITH A VISION ENABLED MULTIMODEL USING OPEN SOURCED LLM"

Gopala[1], Bhumika C S[2], Divyashree Y[3], Lavanya T M[4] and Rachana J R[5]

Professor, Dept. of CSE, Sri Siddhartha Institute of Technology, Tumkur[1]

Students, Dept. of CSE, Sri Siddhartha Institute of Technology, Tumkur[2,3,4,5]

Abstract: The development of Large Language Models (LLMs) has significantly transformed the field of artificial intelligence, enabling machines to understand and generate human-like language. While most LLMs are trained on dominant global languages like English, there is a growing need to include regional languages such as Kannada to ensure linguistic inclusivity and cultural representation. This research focuses on the transformation and application of LLMs for the Kannada language. It explores data collection, preprocessing, tokenization, and fine-tuning strategies to adapt LLMs effectively. The study also addresses challenges such as limited datasets, script complexity, and semantic nuances unique to Kannada. By building or adapting Kannada LLMs, this work aims to enhance natural language processing (NLP) capabilities for Kannada speakers, supporting applications like translation, chatbots, sentiment analysis, and digital education. This transformation is a step towards democratizing AI access across linguistic boundaries.

I. INTRODUCTION

Agriculture is a cornerstone of the Indian economy, and effective communication of agri cultural practices and innovations is essential for empowering farmers. However, the majority of agricultural content is often not available in regional languages, limiting its accessibility. This project aims to develop a generative AI system that communicates with farmers in Kannada, leveraging vision-enabled multimodal capabilities and open-sourced large language models. This approach seeks to enhance the accessibility and impact of agricultural knowledge, ultimately revolutionizing farming practices.

II. METHODOLOGY

Data Collection and Preparation: Gather a comprehensive dataset of agricultural information, including text, images, and videos, in Kannada. Annotate the data to create a robust training set for both language and vision models.
Model Development: Language Model: Fine-tune an open-sourced large language model (e.g., GPT-3) on Kannada agricultural texts to enable fluent and contextually accurate generation.

Vision Model: Train a convolutional neural network (CNN) on agricultural images to recognize crops, diseases, pests, and other relevant visual information.

Integration and Multimodal Processing: Develop a multimodal interface that combines text and image inputs, allowing the AI to process and respond to queries holistically. Implement natural language processing (NLP) techniques to understand and generate responses in Kannada.

Personalization and Contextualization: Use machine learning algorithms to analyse user inputs and provide tailored advice based on specific farming conditions, such as soil type, weather patterns, and crop variety. Implement feedback mechanisms to continuously refine and improve the system's recommendations.

User Interface Development: Design a mobile-friendly application with voice and text input capabilities to facilitate easy interaction for farmers. Incorporate features like image upload and real-time chat for comprehensive support.

Testing and Deployment: Conduct field tests with farmers to evaluate the system's effectiveness and usability. Gather feedback to fine-tune the models and interface before large-scale deployment.

III.   OBJECTIVES

1. Voice-Based Farmer Assistance: Farmers can ask questions in Kannada using voice or text.
The LLM replies in Kannada with appropriate crop suggestions.
2. Chat bots and Virtual Agents in Kannada AI-powered chat bots trained in Kannada to answer agricultural queries 24/7. Useful for pest control tips, fertilizer usage, and sowing techniques.
3. Weather and Crop Alerts Kannada LLMs send weather warnings, rain predictions, and crop-specific alerts in simple Kannada messages. Helps farmers make timely decisions.
4. Access to Government Schemes explains agricultural schemes, subsidies, and application steps in Kannada. Reduces confusion and increases participation in welfare programs.
5. Market Price Information provides daily market prices for crops in Kannada. Helps farmers sell at the right time and place.
6. Translation of Scientific Data LLMs translate technical farming documents into farmer-friendly Kannada. Makes knowledge from universities and research labs accessible.
7. Training and Awareness Programs Kannada LLMs used in audio/video content creation for agri-training. Farmers learn modern methods through Kannada tutorials.
8. Mobile and IoT Integration Used in mobile apps and IoT devices to give Kannada-based voice instructions for smart irrigation, soil monitoring, etc.

Kannada LLM Transforming Agriculture with AI
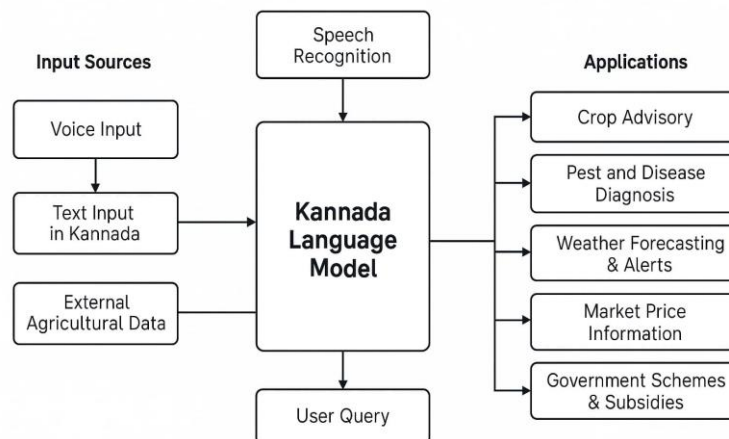
HARDWARE SPECIFICATIONS
- High-Performance GPU: NVIDIA A100, V100, or RTX 3090/4090
    At least 24 GB VRAM per GPU (multi-GPU setup preferred)
- RAM: Minimum: 64 GB
    Recommended: 128 GB or more
- Storage: SSD with at least 2 TB for storing datasets, checkpoints, and logs.
- CPU: Multi-core (8+ cores), e.g., Intel Xeon or AMD Threadripper
- Power Supply & Cooling:Reliable UPS, good ventilation, and cooling systems for stable operation.
- Edge Devices (for farmers): Smartphones (Android) with basic ML inference ability Low-power IoT devices with support for TinyML if needed
- Cloud Infrastructure: AWS, Google Cloud, or Azure with AI/ML GPU instance Containerization support (Docker, Kubernetes)

SOFTWARE SPECIFICATIONS
- Operating System: Ubuntu 20.04 LTS (Preferred for development)
Windows 10/11 (For basic testing)
Android OS (For end-user mobile app)

- Programming Languages: Python (Core language for AI/ML and NLP)

Java/Kotlin (for Android apps)

JavaScript (for web-based applications)

- AI/ML Frameworks: PyTorch or TensorFlow (Model training & inference)

Hugging Face Transformers (For Kannada LLM fine-tuning or using pre-trained models),ONNX (For cross-platform model deployment)

- NLP Tools for Kannada : IndicNLP Library, iNLTK, AI4Bharat Tools,Tokenizers for Kannada
- Development Tools: Jupyter Notebook or Google Colab, Git & GitHub (Version control), VS Code / PyCharm, Docker (for containerization)
- Database & APIs: PostgreSQL or MongoDB (for storing data)

IV. ADVANTAGES

1. Language Accessibility→ Helps farmers understand AI-based solutions in native Kannada without needing English knowledge.
2. Voice and Text Support→ Farmers can use voice input or type in Kannada to get solutions, making tech more inclusive.
3. Accurate Local Information→ Trained on regional data, the LLM provides context-aware responses based on local crops, seasons, and soil types.
4. Real-Time Assistance→ Farmers can ask real-time questions and get instant advice on pests, weather, or crop care.
5. Low Literacy Support→ Voice-based interfaces allow illiterate or semi-literate farmers to benefit from AI tools.
6. Customized Recommendations→ Based on location, crop, and season, Kannada LLM can give personalized advice for better yield.
7. Bridging Digital Divide→ Makes modern AI tools usable for rural Kannada-speaking communities.

V. CONCLUSION

The development of a Kannada language generative AI system with vision-enabled multimodal capabilities represents a significant step towards revolutionizing agriculture. By making advanced agricultural knowledge accessible in the local language and providing personalized, real-time support, this project aims to empower farmers, improve productivity, and promote sustainable farming practices. The integration of cutting-edge AI technologies with user-centric design ensures that the benefits of technological advancements reach the grassroots level, fostering a more prosperous and resilient agricultural community.

REFERENCES

1. Dr. Mitesh M. Khapra

Affiliation: IIT Madras, AI4Bharat Co-founder

Contribution: Developed IndicNLP and IndicBERT, foundational for Kannada LLMs

Relevant Work: Kakwani et al., 2020 – "IndicNLP Suite: NLP for Indian Languages"

Key figure in enabling AI models in Indian languages for real-world applications, including agriculture.

2. Dr. Pushpak Bhattacharyya

Affiliation: IIT Bombay

Contribution: Expert in multilingual NLP; his work supports development of regional language AI applications

Relevant to: Machine Translation, WordNet for Indian languages

3. Prof. Ponnurangam Kumaraguru ("PK")

Affiliation: IIIT Hyderabad

Area: Social computing and technology for rural development

Relevance: Advocates for AI in Indian languages and rural tech inclusion, including agriculture

4. Dr. Kalika Bali

Affiliation: Microsoft Research India

Contribution: Specialist in language technologies for Indian languages, including Kannada

Focus: Speech interfaces and conversational agents for low-literacy users in agriculture

5. Dr. Manish Shrivastava

Affiliation: IIIT Hyderabad, NLP & ML Lab

Work: Contributed to multilingual datasets and models like iNLTK, useful for Kannada NLP