A Human-Aligned Agentic AI System for Agricultural Decision Making Using Multimodal Data and RAG

Ashish Newar[a], Krishiv Geriani[b], Rachit Choudhary[c] and Sachidanand Navik[d]

[a,b,c,d]*Final year undergrads, IIT Kanpur*

Exploring and Building Agentic AI Solutions for a High-Impact Area of Society:
textitAgriculture

Abstract—Agriculture in India faces uncertainties from weather, crop diseases, market fluctuations, and complex policies, making reliable decision support critical. We present *Farmnaxx*, a human-aligned agentic AI advisor that combines *multimodal inputs* (text, voice, images) with *retrieval-augmented generation (RAG)* to deliver grounded, explainable, and multilingual answers. Farmnaxx integrates fine-tuned LLaVA models for disease detection and crop recommendation, Whisper for speech queries, and APIs for weather, prices, and government schemes. By dynamically selecting tools and grounding outputs in factual data, the system reduces hallucinations and improves trust. Designed for accessibility, Farmnaxx empowers farmers to make informed decisions through a natural, farmer-first interface.

Contents

1. Introduction

Agriculture remains the backbone of the Indian economy, directly or indirectly supporting over half of the country's population. Yet, farmers across India continue to face persistent challenges such as unpredictable weather, recurring crop diseases, volatile markets, and complex government schemes. These uncertainties often leave farmers with incomplete or unreliable information, making agricultural decision-making highly risky. Even small mistakes in irrigation timing, seed selection, or market timing can have devastating effects on livelihoods.

In recent years, artificial intelligence (AI) and machine learning (ML) have shown promise in providing data-driven support for farmers. However, most existing solutions remain limited: they are trained on narrow datasets, lack adaptability to noisy real-world queries, and often hallucinate when asked about unfamiliar contexts. Furthermore, many are designed only for English text input, overlooking India's multilingual and low-digital-literacy population. Farmers often rely on **voice queries in local dialects** or **visual cues such as diseased leaves**, which these systems cannot handle effectively.

To address these gaps, we present **Farmnaxx**, a human-aligned, agentic AI advisor for agriculture. Farmnaxx is built on **LLaVA 1.5 (7B)**, a vision-language model fine-tuned with domain-specific datasets including the **CDDM plant disease dataset**, **crop recommendation data** (NPK soil parameters, weather variables), **farmer call transcripts**, and **district-level crop yield data**. The system incorporates **Whisper** for speech recognition in multiple Indian languages, **Google Translate/Deep Translator** for multilingual query handling, and **Retrieval-Augmented Generation (RAG)** to ground responses in government policy documents and agricultural portals.

Unlike traditional chatbots, Farmnaxx follows an **agentic workflow**. Incoming queries are first classified into domains such as weather, disease, price, or policy. The system then routes the query to the appropriate tool: APIs for weather and market data, fine-tuned adapters for crop disease or recommendation, or RAG pipelines for government schemes. This design ensures that answers are both **context-specific and grounded in factual data**, significantly reducing the risks of hallucination.

By combining **multimodal capability** (text, voice, image), **multilingual accessibility**, and **agent-based decision flows**, Farmnaxx offers farmers an interface that feels natural while delivering technically robust, explainable results. Beyond agriculture, this work demonstrates how integrating fine-tuned multimodal models with retrieval and agentic reasoning can create **human-aligned AI systems for high-stakes, real-world applications**.

2. Datasets

The effectiveness of any AI system is strongly dependent on the quality and diversity of its datasets. For this work, we curated and utilized multiple agriculture-focused datasets that cover structured production data, crop recommendation inputs, farmer queries, and plant disease images. Together, these datasets enable both language and vision models to address real-world agricultural decision-making tasks.

2.1. CDDM Plant Disease Dataset [1]

It provides paired *image + text* data of plants and their corresponding disease labels, enabling training of models that can understand both *visual symptoms* and *textual descriptions*.

The CDDM dataset contains 137000 plant and leaf images across different crops, annotated with disease labels. It also includes healthy leaf samples, making it suitable for both disease detection and health classification tasks. Each sample is organized in subfolders named after the plant and disease type (e.g., `Apple/Alternaria`

Blotch/plant_69422.jpg). This dataset was essential for fine-tuning multimodal models (such as LLaVA) to handle visual disease identification.

2.2. Agriculture Production Dataset [5]

The AgricultureDataset.csv contains structured records of crop production across Indian states and districts. This dataset was designed for statistical analysis and grounding responses with real agricultural figures. Key features include:

- **State_Name** – Name of the state
- **District_Name** – Name of the district
- **Crop_Year** – Year of production
- **Season** – Agricultural season (e.g., Kharif, Rabi)
- **Crop** – Type of crop
- **Area** – Cultivation area (in hectares)
- **Production** – Crop production (in tonnes)

2.3. Crop Recommendation Dataset [2]

The Crop_recommendation.csv dataset provides soil and climate features for crop recommendation. This dataset is widely used to train supervised learning models that suggest optimal crops for a given set of environmental conditions. Columns include:

- **N, P, K** – Nitrogen, Phosphorus, and Potassium content in soil
- **temperature** – Temperature in °C
- **humidity** – Relative humidity (%)
- **ph** – Soil pH value
- **rainfall** – Rainfall (mm)
- **label** – Recommended crop

2.4. Farmer Query–Response Dataset [3]

The farmer_call_query_dataset.csv is a conversational dataset of real farmer queries paired with expert-provided responses. It is highly valuable for fine-tuning chatbot-style language models for agricultural advisory systems. Its structure includes:

- **questions** – Farmer's natural language query
- **answers** – Expert's response

3. Fine-Tuning Methodology

We highly recommend to go through *Fine-Tuning Llava 1.5 7B on Agricultural Datasets* [4], to get a better a understanding of the fine-tuning process.

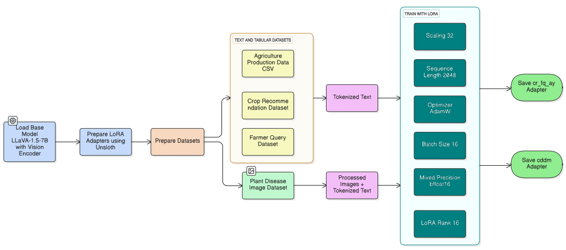

Figure 1. Our Methodology to fine-tunned Llava 1.5 7B using unsloth

To adapt the base **LLaVA-1.5-7B** vision-language model for agricultural decision-making tasks, we performed fine-tuning using the unsloth framework. This approach enabled memory-efficient training with low-rank adaptation (LoRA) modules while preserving the generalization ability of the base model.

3.1. Training Framework

The unsloth library was employed to integrate LoRA adapters into the pre-trained LLaVA backbone. LoRA reduces the number of trainable parameters by restricting updates to low-rank matrices, allowing fine-tuning of large models on limited compute resources. We configured:

- **LoRA Rank:** 16
- **Scaling Factor:** 32
- **Target Modules:** Attention and feed-forward layers
- **Maximum Sequence Length:** 2048 tokens
- **Optimizer:** AdamW with weight decay
- **Batch Size:** 2 (with gradient accumulation)
- **Learning Rate:** Tuned between $2e^{-5}$ and $5e^{-5}$

3.2. Datasets Used

We integrated multimodal agricultural datasets described in Section 2.

1. **AgricultureDataset.csv:** Structured crop production data (state, district, crop type, area, and yield).
2. **Crop_recommendation.csv:** Soil and climate parameters mapped to optimal crops.
3. **Farmer_call_query_dataset.csv:** Query-response pairs for conversational fine-tuning.
4. **CDDM Plant Disease Dataset:** Images of plant and leaf samples labeled with disease type or healthy condition.

The text-based datasets were tokenized and aligned with prompts in instruction–response style, while the image dataset was processed with paired captions and disease labels.

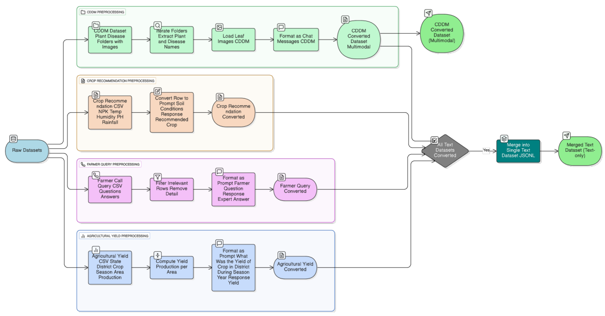

Figure 2. Preprocessing and preparation of dataset for training

3.3. Training Process

To specialize the LLaVA 1.5 backbone for agricultural tasks, we adopted a **parameter-efficient fine-tuning (PEFT)** strategy using **Low-Rank Adaptation (LoRA) adapters**. This approach enabled us to adapt the model's knowledge to the agricultural domain without incurring the prohibitive computational cost of updating all 7 billion parameters of the base model. Instead, LoRA introduces additional trainable matrices at specific layers, thereby allowing domain adaptation while keeping the majority of parameters frozen.

3.3.1. Data Conversion and Multimodal Prompting

The dataset was reformatted into a *chat-style prompt-response format*, which aligns with the conversational structure of large vision-language models. Each entry paired a plant leaf image with a natural language explanation, where the prompt mimicked a farmer's question (e.g., *"What is the content of this picture?"*) and the response provided the disease-aware interpretation (e.g., *"This image shows a tomato leaf affected by blight."*).

This conversion ensured that the model learned to interpret images in a dialogue-oriented setting, preparing it to handle real-world farmer interactions. The process was repeated across all crop and disease categories, resulting in a diverse multimodal training corpus that captured both visual patterns of plant diseases and the linguistic styles of practical agricultural advisory.

3.3.2. Model Initialization and LoRA Configuration

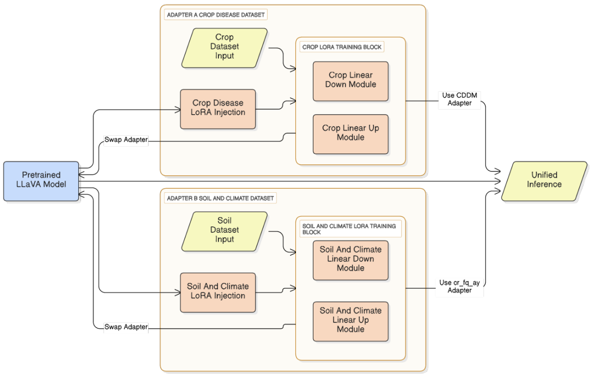

Figure 3. Adapter-Based Fine-Tuning of LLaVA Using Crop Disease and Soil–Climate Datasets for Unified Agricultural Inference

We initialized the training with the LLaVA 1.5 (7B) backbone, loaded in memory-efficient 4-bit quantization to make training feasible on limited GPU hardware (such as Kaggle and Colab environments). LoRA adapters were inserted into key modules of the model, including the vision encoder, transformer attention layers, and MLP blocks.

The configuration adopted was:

- LoRA rank $r = 16$, scaling $\alpha = 16$ for a balanced trade-off between capacity and efficiency.
- Dropout was disabled (`lora_dropout = 0`) to ensure stable convergence.
- Bias terms were excluded (`bias = none`) to avoid unnecessary overhead.
- A fixed random seed was used to maintain reproducibility across runs.

This setup allowed us to fine-tune the model efficiently while preserving the general-purpose reasoning abilities of the underlying LLaVA architecture.

3.3.3. Fine-Tuning Methodology

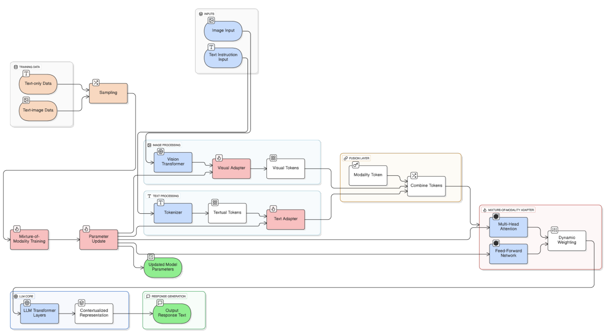

Figure 4. Low-Level Workflow of Fine-Tuning LLaVA with Text-Image Agricultural Datasets

Fine-tuning was conducted through **Supervised Fine-Tuning (SFT)**, a method that aligns model outputs with domain-specific ground truth labels. Instead of adjusting the full parameter set, only the LoRA adapters were updated, making the process significantly more lightweight.

The optimization was carried out using an 8-bit variant of the AdamW optimizer, with a weight decay of 0.01 to regularize the training. A learning rate of 2×10^{-4} with linear scheduling ensured gradual adaptation without catastrophic forgetting of the pre-trained knowledge.

Given hardware constraints, we used a per-device batch size of 2, coupled with gradient accumulation steps to simulate an effective batch size of 8. Training typically converged within 50–64 steps depending on the crop–disease subset, with a maximum sequence length of 1024 tokens to capture detailed multimodal interactions.

3.3.4. Training Configuration and Efficiency

The training configuration can be summarized as:

- **Batching:** Small batch size with gradient accumulation to balance efficiency and stability.
- **Optimizer:** AdamW (8-bit) with weight decay to prevent overfitting.
- **Learning schedule:** Linear warm-up followed by steady-state training.
- **Steps:** 50–64 depending on dataset coverage.
- **Context length:** Up to 1024 tokens, accommodating multimodal dialogue.

This design ensured training was **scalable, reproducible, and hardware-friendly**, while still achieving strong domain adaptation. The use of LoRA adapters allowed us to preserve the *general conversational abilities* of LLaVA while injecting targeted agricultural expertise.

3.4. Model Output and Capabilities

After fine-tuning, the model demonstrated robust agricultural intelligence. Specifically, it was able to:

- **Farmer dialogue:** Provide conversational and context-aware responses to farmer queries.
- **Crop recommendations:** Suggest suitable crops based on soil composition, weather patterns, and local conditions.
- **Disease detection:** Accurately diagnose plant diseases from leaf images and describe them in accessible language.
- **Decision support:** Integrate structured agronomic knowledge with visual cues to generate actionable recommendations.

The resulting system thus retained the flexibility of the original LLaVA backbone while being specialized for the agricultural domain. This ensured that the model was both **computationally efficient** and **highly relevant for real-world agricultural decision-making**.

4. System Design and Architecture

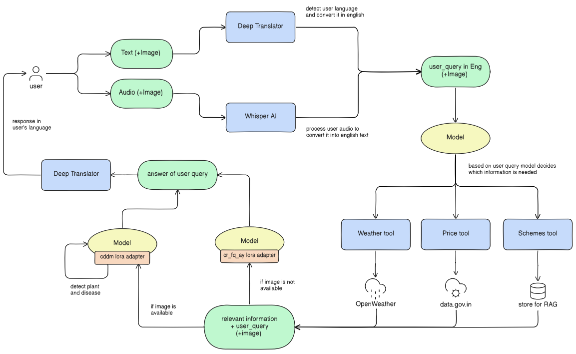

Figure 5. Our architecture integrating multilingual, multimodal inputs with adaptive fine-tunned LoRA adapter, RAG grounding and APIs over live data sources

The proposed system is designed to provide an agentic, multimodal AI-powered advisor for agriculture. It integrates text, audio, and image inputs with multilingual support, fine-tuned vision-language models, and external information sources such as weather forecasts, market prices, and government schemes. Figure 5 illustrates the overall workflow of the system.

4.1. User Interaction Layer

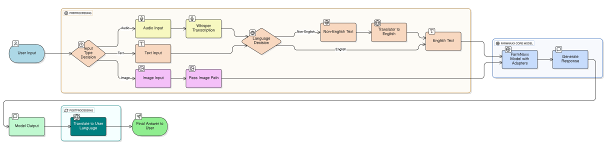

Figure 6. Multilingual & Audio support in I/O operations

Users can interact with the system using either **text** or **audio**, with an optional image (e.g., crop or leaf photo). Since farmers across India speak in diverse languages and dialects, multilingual accessibility is essential. The following modules handle user input:

- **Deep Translator**: Detects the user's language and translates text into English for further processing.
- **Whisper AI**: Converts spoken audio into English text, ensuring accessibility for non-literate or semi-literate users.

The output of this stage is a normalized query in English, optionally accompanied by an image.

4.2. Agentic Tooling and Test Case Coverage

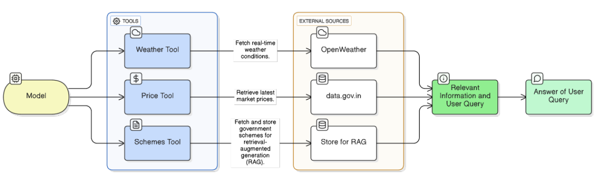

Figure 7. Agentic tooling for intelligent query classification and tool orchestration

To ensure that our fine-tuned **LLaVA-based model** works effectively in real-world scenarios, we implemented an **Agentic Tooling Layer**. This layer is responsible for:

- Automatically classifying incoming farmer queries into categories such as *weather*, *market prices*, *subsidy information*, or *crop disease detection*.
- Dynamically routing the query to the most relevant tool (e.g., weather API, government subsidy datasets, image-based crop disease classifier).
- Seamlessly integrating the tool's response with the model's fine-tuned reasoning, ensuring context-aware, farmer-friendly outputs.

4.2.1. Test Case Coverage

We evaluated the tooling system using diverse test cases to capture the **full range of farmer queries**, both text-only and multimodal (with images).

The evaluation demonstrated that the system can:

1. **Accurately detect the required tool** for each type of query.
2. **Seamlessly integrate** tool responses with the model's fine-tuned reasoning.
3. **Provide clear, actionable, and farmer-friendly answers**.

For example:

- A weather-related query (*"Will there be strong winds in Pune next week?"*) was routed correctly to the weather information tool, and the system generated actionable advice on crop protection.
- A market-related query (*"What is the price of rice in Uttar Pradesh this week?"*) was directed to market price datasets, yielding accurate and region-specific results.
- A subsidy-related query (*"Is there any subsidy for drip irrigation in Maharashtra?"*) was resolved using official government scheme information, complete with references.
- An image-based query (*"This is my crop in Gujarat. Is it affected by any pest?"*) triggered the crop disease detection pipeline, giving actionable suggestions for pest management.

4.2.2. Robustness of Tooling

The results confirm that the **agentic tooling layer** is robust across diverse query types. It not only enhances the system's intelligence by combining **symbolic reasoning with learned knowledge**, but also ensures that farmers receive reliable, accurate, and easy-to-understand responses across real-world scenarios.

4.3. Core Reasoning Layer

The heart of the system is the fine-tuned **LLaVA-1.5-7B** model, enhanced with LoRA adapters for agricultural tasks:

- **cddm adapter**: Specialized for crop disease detection and classification from images.
- **cr_fq_ay adapter**: Fine-tuned on crop recommendation, farmer queries, and yield prediction datasets.

The model determines the nature of the query—whether it is image-based (disease detection) or text-based (general farming, weather, finance, or policy).

4.3.1. Model Implementation Details

The core intelligence of *farmnaxx* is powered by a multimodal large language model (LLaVA-1.5 backbone) enhanced with domain-specific *LoRA adapters*. These adapters specialize in either (i) **Crop Disease Diagnosis (CDDM)** from images or (ii) **General Agricultural Advisory (CR_FQ_AY)** from text queries. The adapters are dynamically *hotswapped* depending on the input modality, ensuring parameter-efficient specialization.

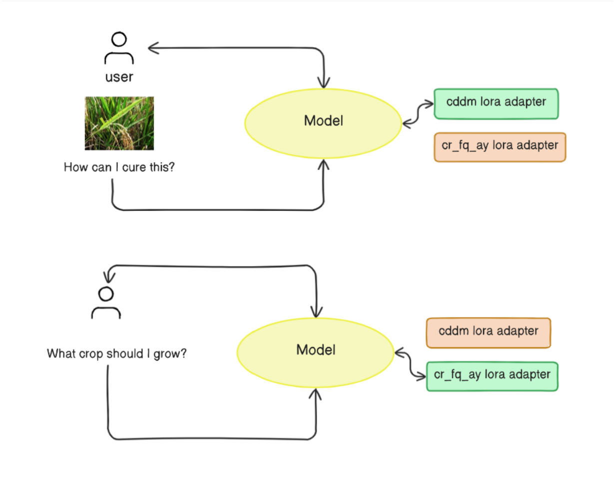

Figure 8. Working of the model showing adapter hotswapping for text vs image queries

Hot-swap pipeline overview:

1. **Model Initialization:** A base vision-language model is loaded using `FastVisionModel.from_pretrained` with quantization for efficiency.
2. **Adapter Switching:** Based on the input type (text vs. image), the corresponding LoRA adapter is hotswapped using `hotswap_adapter`.
3. **Text Inference:** For multilingual queries, the model leverages Whisper (speech-to-text) + Deep Translator (translation) before generating a contextualized answer using the CR_FQ_AY adapter.
4. **Image Inference:** For plant disease detection, the CDDM adapter processes the image+query pair. The raw disease analysis is then cleaned and reformulated into farmer-friendly guidance.
5. **Response Generation:** Outputs are translated back into the user's preferred language and optionally synthesized into speech for accessibility.

4.4. External Knowledge Integration

To provide fact-grounded and real-time insights, the model interacts with specialized external tools:

- **Weather Tool**: Uses OpenWeather APIs to retrieve forecasts and conditions relevant to crop cycles and irrigation planning.
- **Price Tool**: Fetches market price information from data.gov.in, helping farmers make better selling decisions.
- **Schemes Tool**: Implements Retrieval-Augmented Generation (RAG) to fetch information on government policies, subsidies, and credit schemes from public datasets.

4.5. Response Generation Layer

The model synthesizes responses by combining outputs from LoRA adapters and external tools. The generated answer is then:

- Translated back into the user's preferred language using **Deep Translator**.
- Delivered in natural language form, ensuring clarity and accessibility.

4.6. Summary System Workflow

The workflow follows these steps:

1. User submits a query (text/audio, with optional image).
2. Input is translated or transcribed into English.
3. The query is processed by the model, which decides whether to:
   - Use the **cddm adapter** for plant disease detection (if image is available).
   - Use the **cr_fq_ay adapter** for text-based queries (crop recommendation, finance, yield, policies).
4. If required, the model calls external tools (Weather, Price, Schemes) for factual grounding.
5. The final answer is synthesized, translated back into the user's language, and returned.

5. Technology Stack and Architecture

The system centers around:

- **Fine-tuning:** LoRA adapters using Hugging Face Transformers, PEFT, and BitsAndBytes for efficient domain specialization.
- **Retrieval and Reasoning:** FAISS enable document retrieval and augment model inference for factual accuracy.
- **Speech and Translation:** Whisper AI for speech-to-text, Coqui TTS for text-to-speech, and IndicTrans2/NLLB for multilingual handling.
- **Deployment:** Quantized models (AWQ/GGUF) ensure performant edge device operation in rural and low-connectivity zones.

6. Overcoming Challenges

During the development of the agricultural assistant, several challenges were encountered which are common to large-scale vision-language systems. We adopted targeted solutions to ensure robustness, scalability, and domain alignment.

- **Handling Hallucinations:** A major challenge with generative models is the tendency to produce confident yet factually incorrect responses (hallucinations). To mitigate this, we integrated a **Retrieval-Augmented Generation (RAG)** pipeline. Authoritative agricultural documents, extension manuals, and government guidelines were indexed into a vector database, enabling the model to ground its answers in verified sources. Confidence scores and explicit reasoning steps were incorporated into responses, ensuring that end-users could differentiate between retrieved facts and model inferences.
- **Multilingual and Code-Switched Queries:** Farmers frequently communicate in local dialects or switch between English, Hindi, and Romanized forms of Indic languages (e.g., Hinglish). To support inclusivity, we designed a specialized translation and normalization pipeline that maps these heterogeneous inputs into standardized representations. This included leveraging pretrained multilingual embeddings, handling spelling variations, and ensuring semantic consistency. As a result, the system was able to process queries across multiple Indian languages and dialects, greatly improving accessibility.
- **Noisy and Incomplete Data:** Agricultural data sources (such as field notes, farmer records, or scanned advisory leaflets) are often unstructured, incomplete, or noisy. We employed advanced preprocessing methods including text cleaning, OCR-based error correction, and entity normalization. For retrieval, a vector-based semantic search system was deployed, allowing robust extraction of relevant passages even in the presence of incomplete or partially corrupted input data. This significantly improved the reliability of downstream recommendations.
- **Scalability and Extensibility:** Agricultural knowledge is highly domain-specific and rapidly evolving, with frequent updates in pest management strategies, fertilizer guidelines, and climate adaptation practices. To ensure adaptability, we employed a **modular LoRA adapter architecture**. New adapters can be trained for additional crops, regions, or languages without requiring re-training of the full 7B parameter model. This plug-and-play design ensures scalability, reduces compute requirements, and makes the system extensible to other agricultural or allied domains.

7. Impact

The fine-tuned multimodal system demonstrates significant impact in both practical agricultural settings and broader socio-economic contexts.

- **Improved Farmer Decision-Making:** By providing timely, accurate, and context-aware advice, the system enables farmers to make informed decisions regarding crop selection, irrigation, fertilizer application, and pest control. This directly contributes to higher crop yields, reduced losses, and improved livelihood security.
- **Bridging the Digital Divide:** The support for local Indian languages, Hinglish, and voice-based interaction lowers barriers for smallholder farmers who may have limited digital literacy. The conversational format creates an intuitive interface that extends access to marginalized communities, thereby democratizing agricultural knowledge.
- **Reducing Risks and Costs:** Farmers often rely on informal networks or delayed expert consultations, which can result in misinformation and poor crop outcomes. By grounding responses in authoritative datasets and delivering advice in real time, the system reduces uncertainty, mitigates risks of crop failure, and minimizes costs associated with trial-and-error approaches.
- **Empowering a Broad Stakeholder Base:** Beyond farmers, the system benefits agricultural vendors, cooperatives, and policy officials by providing consistent and scalable access to domain expertise. Vendors can align supply chains with crop cycles, while policy agencies can utilize aggregated insights for monitoring regional agricultural health and planning interventions.
- **Modular and Scalable Design:** A key contribution of the architecture lies in its modular use of LoRA adapters, which allows rapid specialization to new domains without retraining the full backbone model. This design not only supports the current focus areas (plant disease diagnosis and crop advisory) but also enables future extensions into related domains such as soil health analysis, pest outbreak prediction, financial risk advisory, or climate resilience planning.
- **Sustainability and Accessibility:** The memory-efficient fine-tuning (4-bit quantization, gradient checkpointing) ensures that the system can be trained and deployed on modest hardware resources, including cloud platforms like Kaggle or Colab. This makes the solution affordable and accessible for widespread use in resource-constrained rural settings.
- **Policy and Knowledge Integration:** The ability to integrate structured knowledge (government guidelines, extension manuals) with unstructured field data positions the system as a decision-support tool not only for individual farmers but also for policymakers seeking to implement evidence-based agricultural programs.

Overall, the system demonstrates how multimodal large language models, augmented with efficient fine-tuning strategies, can deliver **domain-aligned, inclusive, and scalable agricultural intelligence**. By aligning cutting-edge AI with real-world farming challenges, the project creates lasting impact across both micro-level farmer decisions and macro-level agricultural policy planning.

8. Future Work

While the current system demonstrates strong performance in crop diagnosis and advisory, several avenues remain open for expansion:

- **Integration of Additional Domain-Specific Adapters:** Extending beyond plant disease and crop recommendations, future work will include training specialized adapters for soil health assessment, pest outbreak prediction, fertilizer optimization, and climate-resilient crop planning. This modular approach ensures rapid domain scalability.
- **Continuous Learning with Community Feedback:** Farmers' queries and feedback represent a rich source of evolving agricultural knowledge. By incorporating mechanisms for active learning and user feedback loops, the system can continuously refine its outputs, improving accuracy and contextual relevance over time.
- **Enhanced Offline Capabilities:** Many rural regions face inconsistent internet connectivity. Future versions will emphasize lightweight deployment strategies, including on-device inference with incremental model updates, to ensure access to reliable agricultural intelligence even in offline or low-bandwidth settings.
- **Cross-Sector Expansion:** Beyond core farming tasks, the system can extend into adjacent sectors. APIs and integration platforms will target financial services (e.g., crop insurance, credit scoring, risk management), supply chain optimization (e.g., demand forecasting, logistics planning), and policy analytics, thereby embedding the system across the agricultural value chain.
- **Responsible AI and Fairness:** Future work will emphasize fairness, transparency, and explainability, ensuring that recommendations remain unbiased, ethically sound, and aligned with the socio-economic realities of smallholder farmers.

9. Conclusion

Our *farmnaxx* system demonstrates how **agentic, multimodal AI**, grounded in real-time agricultural data, can transform decision-making for millions of farmers in India. By combining vision-based disease detection, conversational interfaces, and modular LoRA-based fine-tuning, the system bridges the gap between cutting-edge AI research and grassroots agricultural needs.

The platform's ability to provide **accessible, inclusive, and trustworthy guidance** empowers smallholder farmers with expert-level decision support, reduces dependency on delayed or inconsistent advisory services, and minimizes the risks associated with misinformation. Its modular and scalable design ensures adaptability to new domains, languages, and geographies, making it a blueprint for agricultural AI deployment across the Global South.

In summary, *farmnaxx* illustrates the potential of aligning **responsible AI innovation with sustainable agriculture**, contributing to enhanced food security, improved rural livelihoods, and more resilient agricultural ecosystems.

—

References

[1] *A Multimodal Benchmark Dataset and Model for Crop Disease Diagnosis*. [Online]. Available: https://github.com/UnicomAI/UnicomBenchmark/tree/main/CDDMBench.

[2] *Crop recommendatation Dataset*. [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset.

[3] *Farmer Call Query (Kisan Call Centre)*. [Online]. Available: https://www.kaggle.com/datasets/daskoushik/farmers-call-query-data-qa.

[4] *Fine-Tuning Llava 1.5 7B on Agricultural Datasets*. [Online]. Available: https://www.kaggle.com/code/sachidanandnavik/fine-tuning-llava-1-5-7b-on-agricultural-datasets.

[5] *Indian Agriculture Yeild Dataset*. [Online]. Available: https://www.kaggle.com/code/prasadkevin/crops-prediction-indian-dataset/data.