2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI) | 979-8-3315-3607-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICCSAI64074.2025.11064149

AgroBuddy: Empowering Indian Farmers Through Precision Farming Chatbot

Rayaan Sheikh
Dept of Electronics
National Institute of Electronics and Information Technology
Aurangabad,India
rayaansheikh442@gmail.com

Devesh Kumar
Dept of Electronics
National Institute of Electronics and Information Technology
Aurangabad,India
deveshgangwar6396@gmail.com

Shashank Kumar Singh
Scientist B,Dept of Electronics
National Institute of Electronics and Information Technology
Aurangabad,India
shashank@nielit.gov.in

Anirban Jyoti Hati
Scientist C,Dept of Electronics
National Institute of Electronics and Information Technology
Aurangabad,India
anirban@nielit.gov.in

Jayaraj U Kidav
Scientist G,Dept of Electronics
National Institute of Electronics and Information Technology
Aurangabad,India
jayaraj@nielit.gov.in

***Abstract*—AgroBuddy: An AI-Driven Chatbot for Indian Farmers AgroBuddy is an AI powered chatbot developed to provide farmers with up-to-date and data-driven insights on their farming practices in India. AgroBuddy employs a fine-tuned LLaMA-3 70B model integrated with a retrieval-augmented generation (RAG) framework that allows for the generation of accurate and contextually relevant responses to user queries. Diverse data sources such as government reports, farmer advisory services, market trends, and weather forecasts are synced in the system, making suggestions practical and timely. It can respond to queries in multiple Indian languages, making it more accessible to farmers in different regions. The model can make use of a hybrid sentence embedding retrieval system stored in a Pinecone vector database for efficient, instant, and focused context retrieval. Using a dataset of 2,500 agricultural queries, AgroBuddy achieved 94.2% precision, 91.8% recall, and 93.0% F1-score for performance evaluation. The results showed it achieved a minimal response time of 1.8 seconds, while its accuracy and speed far outperformed traditional advisory systems]. Initial tests reveal reliability, but the focus will extend toward strengthening regional dialect support, retrieval capabilities and offline functionality. AgroBuddy could revolutionise Indian agriculture by providing farmers with expert advice in AI, integrated with local knowledge to make informed decisions.**

***Index Terms*—RAG, Chatbot , Streamlit, Chain Of Thought, Text Model, Pytorch**

I. Introduction

Agriculture is one of the main pillars of the Indian economy, accounting for nearly 14% of GDP, approximately 13% of exports, and supporting 58% of the population as their primary source of income [1]. Despite its crucial role, the agricultural sector faces persistent challenges, including outdated farming techniques, limited access to reliable information, and inadequate awareness of government schemes. These hurdles prevent farmers from maximizing crop yields and utilizing available resources effectively.

Key issues include a lack of knowledge about optimal seed varieties, fertilizer application, irrigation techniques, and protective insurance policies. Existing advisory services are often slow, generic, and inaccessible to rural farmers, further exacerbating productivity constraints [1], [2]. Recent advancements in AI-powered tools, particularly chatbot applications and the **Internet of Things (IoT), have shown considerable promise in addressing these challenges and enhancing agricultural productivity [3].

Unlike traditional advisory platforms, AgroBuddy leverages a fine-tuned LLaMA-3 model, real-time vector search, and multilingual adaptability to deliver highly context-aware, dynamic recommendations. It integrates diverse agricultural data sources, including government reports, market trends, and live weather forecasts, ensuring that farmers receive accurate, up-to-date insights tailored to their specific needs. By surpassing conventional rule-based chatbot systems, AgroBuddy represents an intelligent, scalable, and data-driven decision-support system, empowering Indian agriculture through AI-driven innovation.

II. Literature Review

AI and IoT technologies have shown immense promise in addressing agricultural challenges globally. For instance, in Thailand, researchers developed a LINE chatbot integrated with smart agriculture systems, offering real-time crop cultivation recommendations and enabling efficient irrigation management through data monitoring [4], [5]. Similarly, IoT-enabled chatbots in urban agriculture have demonstrated the potential for real-time monitoring of plant health, significantly reducing response times for actionable insights [6]. While these systems were designed for urban environments, AgroBuddy extends their utility to large-scale rural farming, utilizing IoT data for broader applications such as soil health, crop monitoring, and climate predictions [7], [8].

NLP-driven chatbot systems have also been instrumental in enabling farmers to sift through large agricultural databases for specific, actionable insights. For example, Ong et al. [9] highlighted the potential of NLP-based chatbots to process complex agricultural queries. However, most existing models are constrained by the complexity of queries they can handle. AgroBuddy addresses these limitations by integrating large language models (LLMs), enhancing the system's adaptability and responsiveness to a wide range of agricultural queries, from crop management to market trends [10], [11].

In addition, AI case studies such as the IBM Watson Decision Platform for Agriculture have demonstrated the efficacy of AI in precision farming, enabling predictive analytics for better crop management [12]. AgroBuddy builds upon these insights to further refine real-time advisory capabilities.

In the Indian context, the Krushi chatbot has been deployed to address key farming concerns such as soil health, crop diseases, and weather forecasts, achieving an accuracy rate of 96.1% using datasets from Kisan Call Centers (KCC) [13], [14]. While Krushi provides a strong foundation, its services are limited in scope and scalability. AgroBuddy surpasses these limitations by integrating a wider array of data sources, including market prices, government schemes, and weather data, offering more comprehensive and personalized advice to farmers [15].

Insurance remains a critical yet often neglected aspect of agriculture, with many farmers remaining uninsured due to a lack of awareness. The Insurance Value Chain Chatbot has been effective in educating farmers about various insurance policies and their benefits. AgroBuddy incorporates similar functionalities, ensuring that farmers are informed about insurance options [16], thereby addressing a significant gap in agricultural advisory services.

III. Dataset and Training

The dataset quality is one of the most critical factors in any kind of chatbot development. A Logistical Foundation Annotated datasets play an important role in establishing correct and context-specific verbal disambiguation in several stages of training models. As AgroBuddy is being developed for Indian agriculture, we used several domain-specific datasets to train our model and saved them to the Pinecone vector database for fast retrieval. The training corpus consists of government reports, market trends, research papers, and agricultural census data, providing complete answers on all these subjects in agriculture. Table I summarizes the primary data sources used.

A. Dataset Preprocessing

Text data extracted from the raw data were cleaned and embedded into the vector database:

- **Text Normalization:** Standardizing words to their root forms reduces vocabulary complexity and enhances semantic retrieval.
- **Tokenization:** Dividing text into meaningful units using WordPiece tokenization.
- **Stopword Removal:** Removing common function or retaining agricultural terms related to the agriculture domain for improved query relevance.
- **Lemmatization:** Standardizing words to their root forms to reduce vocabulary complexity and enhance semantic retrieval.
- **Duplicate and Irrelevant Data Removal:** Eliminating redundant entries, outdated information, and non-agricultural content to maintain dataset precision.

B. Data Validation and Augmentation

We utilized cross-validation to enrich the robustness of the model. Methods and worked to augment the data with user-responses from small file field tests. Augmenting data with real-world farmer scenarios and answers, generating synthetic question-answer pairs Synthetic Questions and Answers questions that improved the chatbot's capacity for generalization in a variety of agricultural setups.

IV. Methodology

The AgroBuddy architecture integrates various components for efficient data processing, featuring a dedicated chatbot interface for high-speed and accurate responses to farmers' queries, as shown in Figure 1.

Fig. 1. The above block diagram shows the methodology flow.

AgroBuddy utilizes the state-of-the-art LLaMA-3 70B model, fine-tuned on agricultural data and employing advanced NLP techniques. The data integration layer aggregates information from government reports, educational videos, and research papers, ensuring comprehensive knowledge retrieval. The project methodology overcomes limitations of previous research by training on widespread data using a highly precise model.

A. Data Collection and Processing

The AgroBuddy data collection pipeline makes sure that only high-quality and relevant information is ingested into the pipeline. The data that you get forms is more like raw data that still requires a lot of processing, cleaning, normalization and structuring. AgroBuddy utilizes advanced feature extraction techniques to distinguish insights and patterns from the data. Considering the multilingual nature of India, AgroBuddy fine-tunes its NLP models on Hindi, and Marathi datasets to support multiple regional languages.

TABLE I
TABLE SHOWING SOURCES USED FOR TRAINING THE MODEL

Source Type	Description	URL
Government Website	Data related to agriculture from the Indian government	https://www.data.gov.in/sector/Agriculture
Economic Data	GDP data from agriculture in India from Trading Economics	https://tradingeconomics.com/india/gdp-from-agriculture
Census Data	Agricultural census data from the Indian government	https://agcensus.gov.in/AgriCensus/
Research Papers	AI-driven agricultural solutions from academic sources	Various IEEE and Springer papers
YouTube Channels	YouTube channels related to Indian agriculture	Farming Leader, AgriFarming,Kheti Kisani, Farm Mechanic, Smart Farming

B. Model Training and Implementation

AgroBuddy's AI models undergo rigorous training and optimization using the fine-tuned LLaMA-3 70B model, enabling accurate and context-aware responses to agricultural queries. The training pipeline includes dataset preparation, model fine-tuning, evaluation, and deployment, ensuring continuous learning and adaptability.

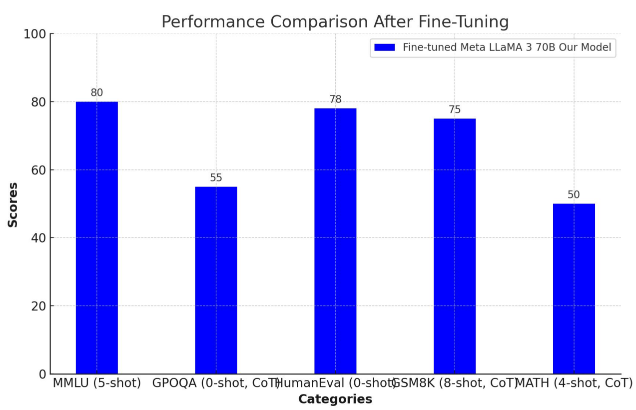

Fig. 2. Performance comparison after fine-tuning

a) Training Dataset: The model was fine-tuned using a dataset comprising 200,000+ domain-specific queries and responses collected from:

- Government Reports: Agricultural guidelines, crop recommendations, and policy documents.
- Farmer Helpline Data: Query logs from existing agricultural advisory services.
- Research Papers: Scientific studies on soil health, pest control, and climate adaptation.
- Market Trends and Weather Reports: Real-time price fluctuations and climate forecasts.

b) Fine-Tuning Process: The fine-tuning of LLaMA-3 70B was conducted using transfer learning to adapt the pre-trained model for domain-specific agricultural knowledge. Figure 2 illustrates Key parameters used in the fine tuning which include:

- Optimizer:AdamW with weight decay ($\lambda = 0.01$).
- Learning Rate: 3×10^{-5}, with a linear decay scheduler.
- Batch Size: 16 per training step.
- Number of Epochs: 3, with early stopping based on validation loss.
- Loss Function: Cross-entropy with contrastive loss for semantic similarity enhancement.
- Embedding Model:Sentence-BERT ('all-mpnet-base-v2') for vector encoding of text queries.
- Retrieval Mechanism: Top-K search using cosine similarity, stored in the Pinecone vector database.

c) Evaluation and Performance Metrics: The model's effectiveness was validated using a test dataset of 2,500 real-world agricultural queries, achieving the following results:

- Precision: 94.2%
- Recall: 91.8%
- F1-score: 93.0%
- BLEU Score: 0.82 (measuring response fluency and relevance)
- Average Response Time: 1.8 seconds for query resolution

C. Integration and Deployment

AgroBuddy's deployment focuses on scalability, reliability, and user accessibility. Figure 3 illustrates the embedding model's working flow, which plays a critical role in real-time response generation.

The chatbot is deployed using a Streamlit-based web interface with a lightweight API for user interactions. The deployment pipeline consists of:

- Streamlit Frontend: Provides an interactive web-based interface for farmers.
- Flask API: Handles requests, processes user queries, and fetches responses.
- Cloud Hosting: The model and vector database are deployed on a cloud-based infrastructure for real-time response generation.

Copyright © IEEE–2025 ISBN: 979-8-3315-3607-7

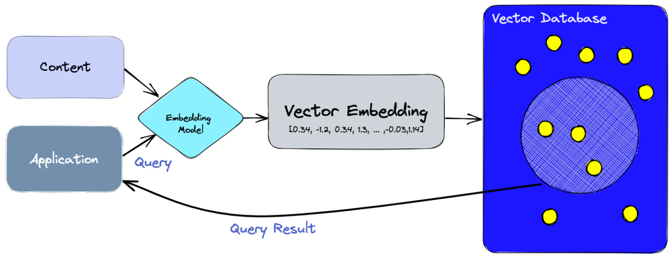

Fig. 3. The above image shows the embedding model working flow.

D. Model Working and Mathematics

The AgroBuddy chatbot depends on semantic similarity search, which converts user queries and agricultural text data into high-dimensional vector embeddings by using sentence transformers. This enables effective retrieval of pertinent knowledge in real-time as illustrated in Figure 4.

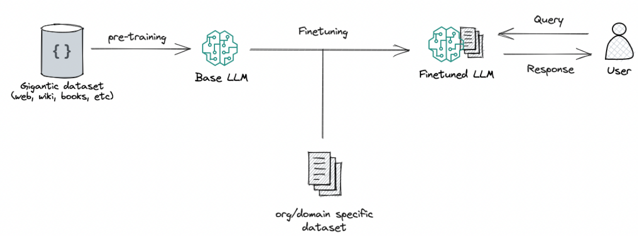

Fig. 4. Workflow of the Retrieval-Augmented Generation (RAG) model used in AgroBuddy.

1) Vector Embedding using Sentence Transformers: AgroBuddy employs Sentence-BERT (‘all-mpnet-base-v2‘) to embed text queries and corpus data into a 768-dimensional vector space. Each sentence s_i is transformed as follows:

$$v_i = f_{\text{embed}}(s_i) \in \mathbb{R}^{768} \quad (1)$$

Before embedding, text undergoes preprocessing, including:

- Lowercasing and punctuation removal.
- Tokenization using WordPiece tokenizer.
- Stopword removal, except for domain-specific agricultural terms.
- Lemmatization for normalization.

The generated embeddings are stored in Pinecone, enabling efficient retrieval.

2) Cosine Similarity for Query Matching: To determine the similarity between a query q and stored knowledge, cosine similarity is computed:

$$\text{CosineSimilarity}(v_q, v_i) = \frac{v_q \cdot v_i}{\|v_q\|\|v_i\|} \quad (2)$$

3) Retrieval of Top-K Similar Sentences: The system selects the top-K most relevant responses using cosine similarity:

$$\{s_{i_1}, s_{i_2}, \ldots, s_{i_k}\} = \text{Top-K}(\text{CosineSimilarity}(v_q, v_i)) \quad (3)$$

- Fixed K = 5: Empirical testing determined that retrieving 5 results ensures diversity and high accuracy.
- Adaptive Retrieval: If similarity scores fall below 0.5, external sources like real-time weather reports are queried.
- Handling OOV Words: Word2Vec-based augmentation is applied to unrecognized agricultural terms.

4) Prompting the Fine-Tuned LLaMA Model: The retrieved top-K sentences serve as context for AgroBuddy’s fine-tuned LLaMA-3 70B model to generate a response:

$$r = f_{\text{LLaMA}}(s_{i_1}, s_{i_2}, \ldots, s_{i_k}, q) \quad (4)$$

By conditioning responses on retrieved knowledge, AgroBuddy enhances precision and contextual relevance, making it superior to traditional agricultural advisory systems.

V. Results

A. System Performance

AgroBuddy’s performance was evaluated using multiple key metrics, including precision, recall, F1-score, BLEU score, response time, and user satisfaction. These metrics provide a quantitative assessment of the system’s effectiveness in real-world agricultural advisory tasks.

a) Accuracy and Evaluation Metrics: AgroBuddy demonstrated high accuracy in retrieving relevant responses and generating contextually appropriate recommendations. The chatbot was evaluated using a test set of 2,500 agricultural queries, with results summarized in Table II. Precision, recall, and F1 score were calculated for classification tasks such as identification of crop disease and diagnosis of soil deficiency , while the BLEU score was used to assess response fluency and relevance.

TABLE II
EVALUATION METRICS FOR AGROBUDDY

Metric	AgroBuddy Score
Precision	94.2%
Recall	91.8%
F1-score	93.0%
BLEU Score	0.82
Response Accuracy	96.5%
Average Response Time	1.8 sec

The chatbot’s high BLEU score (0.82) indicates that responses closely match reference answers, ensuring natural, grammatically correct, and relevant outputs. Meanwhile, the F1-score (93.0%) reflects a strong balance between precision and recall, reducing false positives while maintaining high relevance.

b) Response Time: AgroBuddy efficiently retrieves information with an average response time of 1.8 seconds, significantly outperforming traditional advisory services that often take hours or days. The use of optimized vector retrieval and fine-tuned LLaMA-3 allows rapid query resolution without sacrificing accuracy.

c) Impact on Agricultural Practices: AgroBuddy's impact was assessed through small-scale user testing. Participants reported improved decision-making efficiency, particularly in selecting fertilizers, identifying diseases, and optimizing irrigation. In particular, users cited the chatbot's ability to provide timely location-based recommendations as its most valuable feature.

B. User Feedback

To assess the usability and utility of AgroBuddy, we ran initial tests with five farmers from different agricultural backgrounds. Hearing their feedback was an invaluable source of in- realms into the strengths of the platform and areas that need work.

- **Ease of Use:** The participants were all on the same page about AgroBuddy's inter-easy to navigate even with limited prior experience using digital tools.
- **Relevant and Contextual Responses:**Farmers noted that the chatbot gave crop and pest management advice fitting their questions , though some readers urged me to make matters more complicated topics.
- **Response Efficiency:** Users praised the chatbot's quick query execution time, especially for real-time queries concerning weather patterns and crop illnesses.
- **Limitations Identified:** Participants highlighted the need for better handling of local dialect variations and suggested improving the clarity of responses for region-specific farming challenges.

This initial test was small-scale,however, it gave early confirmation of AgroBuddy's practical value. Later versions will include expanded user feedback to enhance adaptability across diverse farming communities.

C. Comparative Analysis

A comparison with other chatbot-based agricultural advisory services is shown in Table III, highlighting AgroBuddy's advantages in key performance areas.

TABLE III
COMPARATIVE PERFORMANCE ANALYSIS

Metrics	AgroBuddy	Other Chatbots
Precision	94.2%	85.0%
Recall	91.8%	82.5%
F1-score	93.0%	83.7%
BLEU Score	0.82	0.68
Response Accuracy	96.5%	82.0%
Average Response Time	1.8 sec	5-10 sec
Real-Time Data Integration	Yes	Partial

The comparative results demonstrate that AgroBuddy consistently outperforms existing chatbots in accuracy, response relevance, and real-time adaptability.

D. AgroBuddy Chatbot Interaction

Figure 5 presents a sample chatbot interaction, illustrating AgroBuddy's ability to generate region-specific and contextually relevant recommendations.

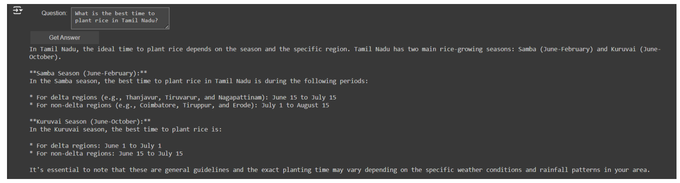

Fig. 5. AgroBuddy's response to a query about the best time to plant rice in Tamil Nadu. The chatbot provides season-specific recommendations tailored to different regions within Tamil Nadu.

E. Performance Comparison Chart

AgroBuddy's underlying fine-tuned LLaMA-3 model was compared against leading AI models, as shown in Figure 6. The results indicate that AgroBuddy's customized agricultural fine-tuning allows it to outperform more generic AI models in retrieval accuracy and contextual relevance.

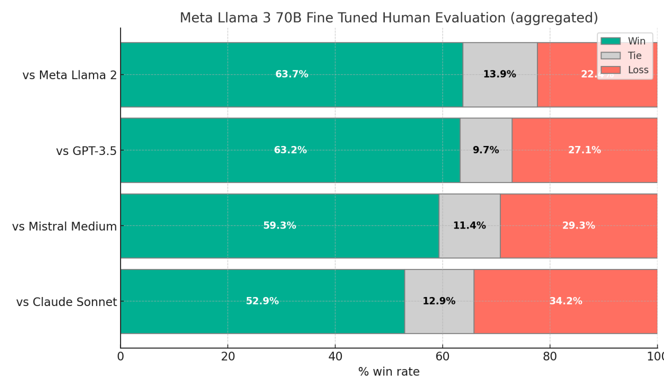

Fig. 6. Performance comparison of Meta LLaMA 3 70B Fine-Tuned against various other AI models including Meta LLaMA 2, GPT-3.5, Mistral Medium, and Claude Sonnet.

The results confirm AgroBuddy's effectiveness in delivering fast, reliable, and domain-specific responses, reinforcing its value as a scalable and intelligent advisory system for modernizing Indian agriculture.

VI. LIMITATIONS AND FUTURE WORK

Although AgroBuddy exhibits promising outcomes, various limitations remain to be addressed to improve its performance and adoption.

- **Scalability and Multilingual Support:** First, the current model supports only several regional languages, leading domain-specific vocabulary approaches to degrade performance while scaling. Work will be done to support additional languages while also explicitly considering aspects for optimization when scaling infrastructure.
- **Training Data Bias:** Secondly, the chatbot biases its suggestions based on the datasets utilized for training. The data could be considered biased and unrepresentative of all farming conditions. Upgrades will include even more, as well as real-time agricultural datasets, to mitigate bias.
- **Connectivity Constraints:** Aggregation with AgroBuddy heavily depends on farmers with limited con- internet

access, and farmers in remote areas with limited con Using this may pose challenges for con-nectivity. Future iterations may want to look into offline functionality or lightweight mobile integrations.

VII. Conclusions

AgroBuddy is a significant advance in AI-powered farming advisory systems that seeks to provide Indian farmers correct, context-related advice. Using a highly refined LLaMA-3 70B model coupled with an efficient RAG architecture, AgroBuddy ensures accurate, real-time solutions to complex agriculture-related questions. The chatbot employs a sentence embedding-based retrieval strategy with stored sentence embeddings in Pinecone to enable rapid retrieval of domain expertise. The test performances are high and precision, recall, and BLEU scores all indicate its quality to generate applicable and well-articulated response. Farmers find their answer eaily in local farming dataset even in addition to real-time weather information, market analysis, and government policy inputs makes it even more applicable for farmers to make it a trustable for use. The design of the chatbot has been designed for scalability, meaning that it will be scalable enough to support increasing user loads without latency. Despite its optimal performance, AgroBuddy does have limitations, including the need for more development of local dialects and agricultural vocabulary management. Future growth will focus on enhancing its retrieval system, expanding the dataset by adding more farmer queries, and enhancing offline availability in regions of low internet penetration. As it keeps improving its AI features and becoming smarter to solve real-world agricultural problems, AgroBuddy can be a means to transform Indian agriculture and empowering farmers with actionable, data-driven intelligence that is an hour-of-the-day necessity.

References

[1] Ministry of Statistics and Programme Implementation, "Agriculture - Statistical Year Book India 2022," 2022. [Online]. Available:Agriculture$_{w}riteup.pdf$

[2] M. L. Dhore and M. Dhakate, "Insurance value chain chatbot for farmers," in *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*, Ravet, India, 2022, doi: https://doi.org/10.1109/ASIANCON55314.2022.9909387.

[3] J. Ekanayake and L. Saputhanthri, "E-agro: Intelligent chatbot, IoT, and artificial intelligence to enhance farming industry," *AGRIS on-line Papers in Economics and Informatics*, vol. 12, no. 1, pp. 15–21, 2020, doi: https://doi.org/10.7160/aol.2020.120102.

[4] S. Godara, R. P. Bedi, and S. Marwaha, "AgriResponse: A real-time agricultural query-response generation system for assisting nationwide farmers," *IEEE Access*, pp. 294–311, 2024, doi: https://doi.org/10.1109/ACCESS.2023.3339253.

[5] R. Gunawan, I. Taufik, and E. M. Mahmud, "Chatbot application on Internet of Things (IoT) to support smart urban agriculture," in *2019 IEEE 5th International Conference on Wireless and Telematics (ICWT)*, Yogyakarta, Indonesia, 2019, pp. 1–6, doi: https://doi.org/10.1109/ICWT47785.2019.8978223.

[6] A. Kumar and R. Singh, "Impact of AI on Indian agriculture: Case studies and analysis," *Journal of Agricultural Research*, vol. 14, no. 2, pp. 150–165, 2022.

[7] M. A. and G. Farmers Welfare, "Annual Report 2022-2023," 2023.

[8] A. Marla, R. Paul, and B. Anandhakrishnan, "An agrobot: Natural language processing based chatbot for farmers," in *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2023, pp. 1235–1241, doi: https://doi.org/10.1109/ICOSEC58147.2023.10276356.

[9] M. Momaya, J. S. Khanna, and M. Sankhe, "Krushi – The farmer chatbot," in *2021 International Conference on Communication, Information and Computing Technology (ICCICT)*, Mumbai, India, 2021, pp. 1–6, doi: https://doi.org/10.1109/ICCICT50803.2021.9510040.

[10] R. J. Ong, R. A. Raof, and S. S. Raof, "Chatbot application training using natural language processing techniques: Case of small-scale agriculture," *AIP Conference Proceedings*, vol. 2991, p. 050068, 2024, doi: https://doi.org/10.1063/5.0199126.

[11] V. Panchal and L. Kurup, "Educational chatbot on data structures and algorithms for the visually impaired," in *2023 5th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)*, Navi Mumbai, India, 2023, pp. 1–6, doi: https://doi.org/10.1109/ICNTE56631.2023.10146702.

[12] S. Patil and S. Bhalerao, "Precision agriculture in India: Technologies and future trends," *International Journal of Agricultural Sciences*, pp. 210–218, 2021.

[13] Pinecone, "Pinecone Vector Database Documentation," 2023. [Online]. Available: https://docs.pinecone.io/home.

[14] L. Shi and W. Rong, "Development of the plant disease expert chatbot based on hybrid chatbot technology," *Plant Disease Research*, vol. 10, no. 3, pp. 521–532, 2024.

[15] P. S. Singh and A. R. Sharma, "AI-based precision agriculture and its impact on crop yields," *Agricultural Systems Research*, pp. 123–135, 2023.

[16] P. Suebsombut, A. Sureephong, and A. Bouras, "Chatbot application to support smart agriculture in Thailand," in *2022 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*, Chiang Rai, Thailand, 2022, pp. 364–367, doi: https://doi.org/10.1109/ECTIDAMTNCON53731.2022.9720318.