



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 2, February 2026

Development of a RAG-Powered Chatbot for Agricultural Advice

Mr. S Babu¹, S Aiswarya², R Jaikrishna³, Y Prathyusha⁴, S Bhaskar⁵

Assistant Professor, Department of Computer Science and Engineering, Kuppam Engineering College,
Andhra Pradesh, India¹

Students, Department of Computer Science and Engineering, Kuppam Engineering College,
Andhra Pradesh, India²⁻⁵

ABSTRACT: In the modern agricultural landscape, farmers often struggle to access timely, expert advice regarding crop management, pest control, and soil health. Traditional advisory systems are often limited by human availability, language barriers, and static databases that fail to address dynamic farming queries. In this study, a Retrieval-Augmented Generation (RAG) framework was developed using advanced Large Language Models (LLM) to create "Agri-Advisor," an intelligent conversational agent. The system integrates Vector Database technology (FAISS) for efficient document retrieval and DeepSeek models for natural language generation. The framework processes unstructured data from farming PDF guides, converting them into searchable vector embeddings. Key features include multi-lingual support (Tamil, Telugu, Hindi), voice-enabled interaction, and real-time weather integration. Furthermore, a secure User Authentication and Session Management module was integrated to transform the system from a stateless tool into a stateful application. This feature ensures data privacy and enables persistent chat history, allowing farmers to retrieve past advisory sessions and maintain continuity in crop management. Experimental validation was conducted by uploading crop-specific guides (e.g., Sugarcane, Okra) and assessing the system's retrieval accuracy. Results revealed that the RAG-based model achieved high relevance in query answering, reducing hallucination by implementing a "Smart Fallback" mechanism that distinguishes between context-based answers and general agricultural knowledge. This research contributes to the field of Digital Agriculture by providing a scalable, accessible, and accurate advisory tool for farmers.

KEYWORDS: Artificial Intelligence, Smart Farming, RAG, Large Language Models, Natural Language Processing, Digital Agriculture

I. INTRODUCTION

Agriculture remains the backbone of the global economy, yet the dissemination of technical knowledge to the grassroots level remains a challenge. Farmers often rely on varying sources of information, which may be outdated or scientifically inaccurate. Traditional expert systems and static websites lack the interactivity required to solve specific, localized problems in real-time [1]. Furthermore, the language barrier in rural regions prevents the effective adoption of English-dominated digital tools.

To address these constraints, recent advancements in Generative AI and Natural Language Processing (NLP) offer promising solutions. Unlike standard chatbots that rely on pre-defined scripts, Large Language Models (LLMs) can generate human-like responses. However, LLMs suffer from "hallucinations" (generating false information) when answering domain-specific queries.

The proposed system, Agri-Advisor AI, overcomes this by employing a Retrieval-Augmented Generation (RAG) architecture. This approach combines the generative power of LLMs with the factual accuracy of a curated knowledge base (PDF Guides). By extracting behavioral insights from crop manuals and enabling vernacular voice interaction, the system aspires to bridge the digital divide and empower farmers with precision agriculture advice.

II. LITERATURE SURVEY

The landscape of agricultural knowledge dissemination has fundamentally shifted from human-dependent advisory services toward intelligent, autonomous frameworks. Traditionally, farmers relied on static databases and keyword-based search systems, which frequently fail when faced with the semantic complexity of real-world farming queries. Because of these limitations, the industry has transitioned toward Large Language Models (LLMs) capable of handling unstructured, domain-specific data. For the core reasoning engine, researchers have increasingly adopted the Retrieval-Augmented Generation (RAG) framework. This architecture is preferred because it "grounds" the model's responses in a curated knowledge base—such as expert-validated PDF guides—thereby significantly reducing the risk of "hallucinations" common in generic, ungrounded models.

In this project, the DeepSeek model serves as the natural language interface, while the FAISS (Facebook AI Similarity Search) vector database is utilized for its efficiency in dense vector clustering. This pipeline is chosen for its ability to perform high-speed semantic retrieval, effectively matching user intent (e.g., matching "pest" with "bug") where traditional SQL-based methods would fail. To address the literacy and language barriers in rural regions, the system integrates googletrans for real-time vernacular translation (Tamil, Telugu, Hindi) and gTTS for auditory feedback.

The majority of current discussion in the literature focuses on the retrieval strategy and accuracy. For this reason, a "Smart Fallback" mechanism is implemented in this study to manage the transition between context-based answers and general AI knowledge. In this arrangement, the system evaluates similarity scores to determine if the retrieved context is sufficient. In essence, it is a method to ensure that the user receives trustworthy advice even when specific document context is unavailable, by leveraging the model's pre-trained internal knowledge as a fallback. This project expands on the notion of creating stateful architectures with persistent chat history to improve the continuity of crop management decisions in practical rural settings.

Table 1: Literature Review Summary - AI in Agriculture

Author/Source	Year	Contribution
Jha et al.	2019	Comprehensive review of AI automation and its role in boosting crop production.
Lewis et al.	2020	Foundational work defining RAG for knowledge-intensive NLP tasks.
Samuel et al.	2024	Introduction of AgroLLM, benchmarking LLM performance in specialized agricultural domains.
Marinoudi et al.	2024	Analysis of LLM applications and the specific challenges of domain adaptation in farming.

III. METHODOLOGY

A. Data Processing (The "Ingestion" Phase)

The system first undergoes a "Learning Phase." When a verified agricultural PDF is uploaded:

1. The text is extracted and cleaned.
2. It is vectorized (converted to numbers).
3. It is stored in the persistent vector_db directory. This ensures that the bot's knowledge is static and verified, preventing it from learning incorrect information from user chats.

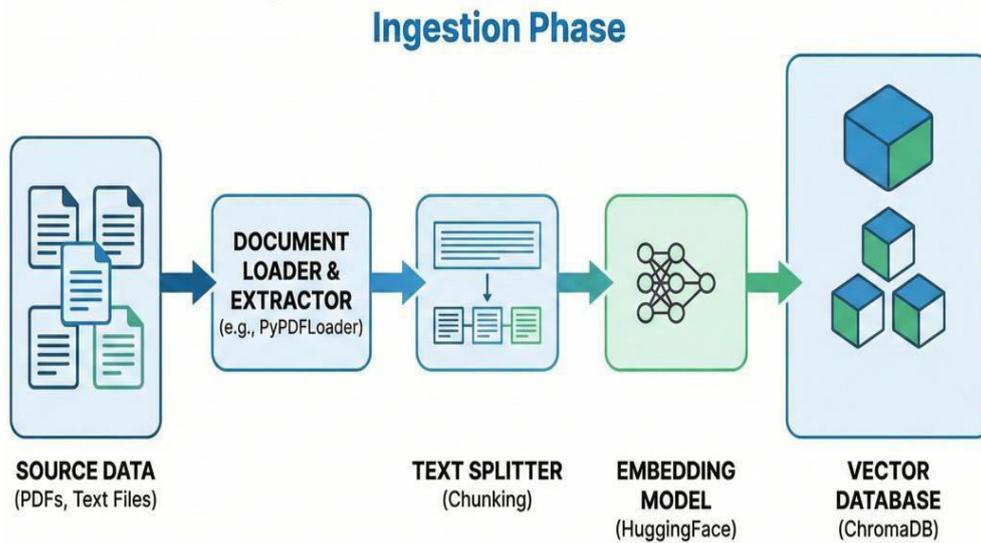


Fig.1 Ingestion Phase

B. Retrieval-Augmented Generation (RAG)

RAG is the backbone of the proposed system. Unlike fine-tuning, which requires massive computational power, RAG retrieves relevant information from an external knowledge base at inference time. This allows the model to remain up-to-date simply by uploading new PDFs, without re-training the neural network [3]. This study utilized LangChain to orchestrate the RAG pipeline.

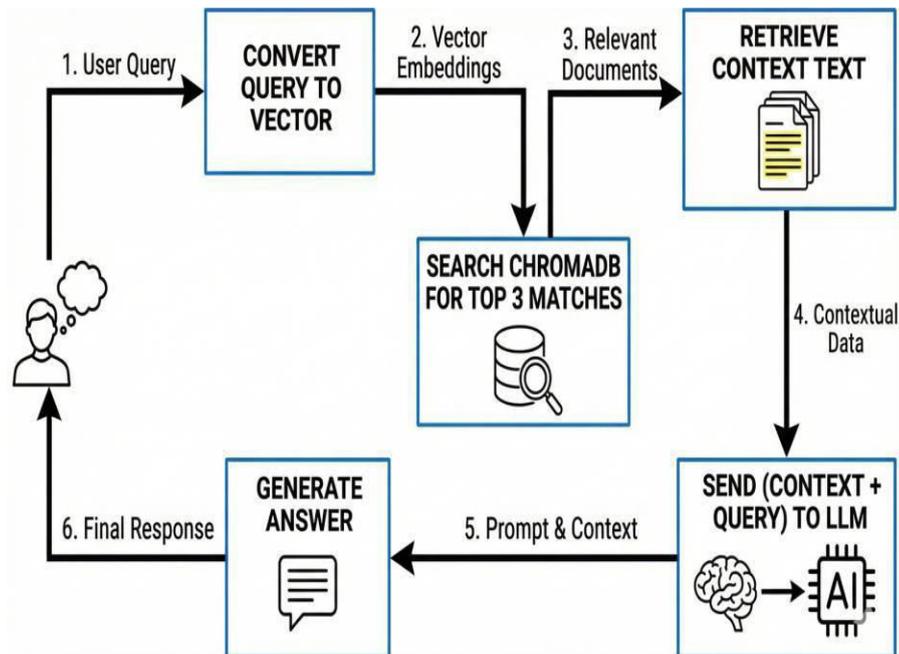


Fig.2 RAG pipeline

C. Vector Embeddings (FAISS)

To perform semantic search, the system utilizes Vector Embeddings. Words with similar meanings are mapped closer together in vector space. FAISS (Facebook AI Similarity Search) was employed for dense vector clustering.

This method outperforms keyword-based search (like SQL LIKE) because it understands the intent of the query (e.g., matching "Bug" with "Pest") [4].

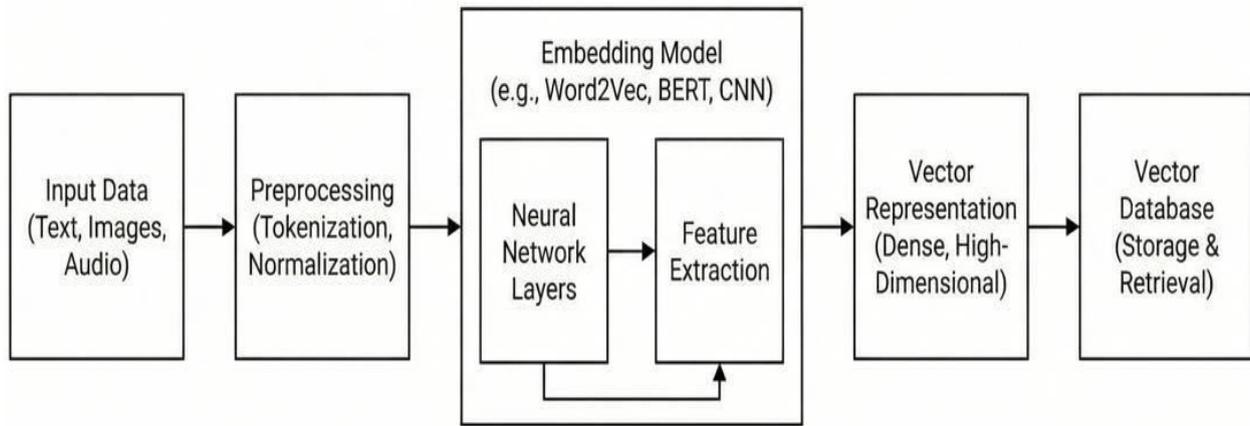


Fig.3 Facebook similarity search

D. Large Language Model (DeepSeek)

The Generative component utilizes a pre-trained Transformer model. The model acts as a natural language interface, synthesizing the retrieved technical data into a friendly, readable answer. The temperature parameter was set to 0.3 to reduce randomness and ensure factual consistency [5].

E. Multi-Lingual & Voice Processing

To make the system accessible to illiterate or semi-literate farmers, SpeechRecognition and gTTS (Google Text-to-Speech) libraries were integrated. The system supports real-time translation between English and Indian regional languages, effectively breaking the language barrier in digital agriculture extension services [6].

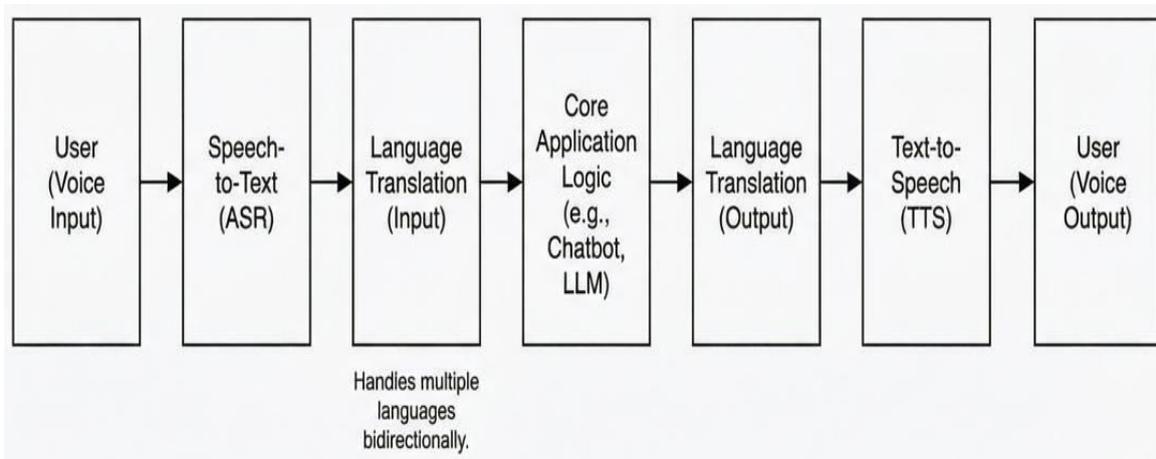


Fig.4 Language Translation

F. Database & State Management Implementation To facilitate long-term data retention without relying on external cloud dependencies, a lightweight SQLite database was integrated via SQLAlchemy ORM. The database consists of two primary entities:

User Entity: Stores credentials and profile metadata.

History Entity: Stores the timestamped query-response pairs, linked via a Foreign Key to the User Entity. This implementation ensures that when a farmer logs in, the system queries the history table to reconstruct their entire conversation timeline, providing a seamless and personalized user experience (UX).

G. User Interface & Experience

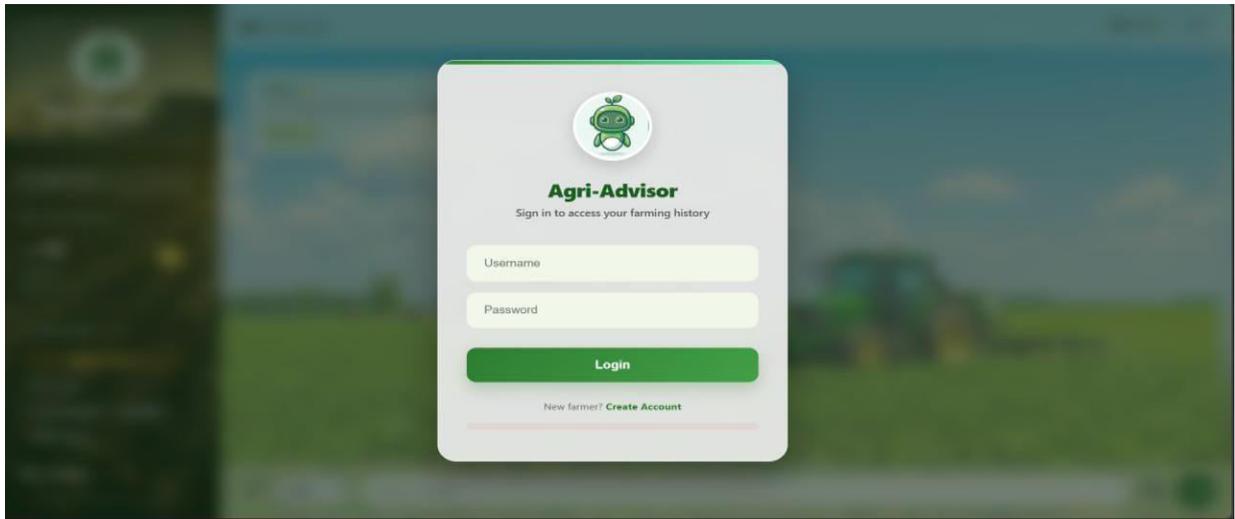


Fig.5 Login Page

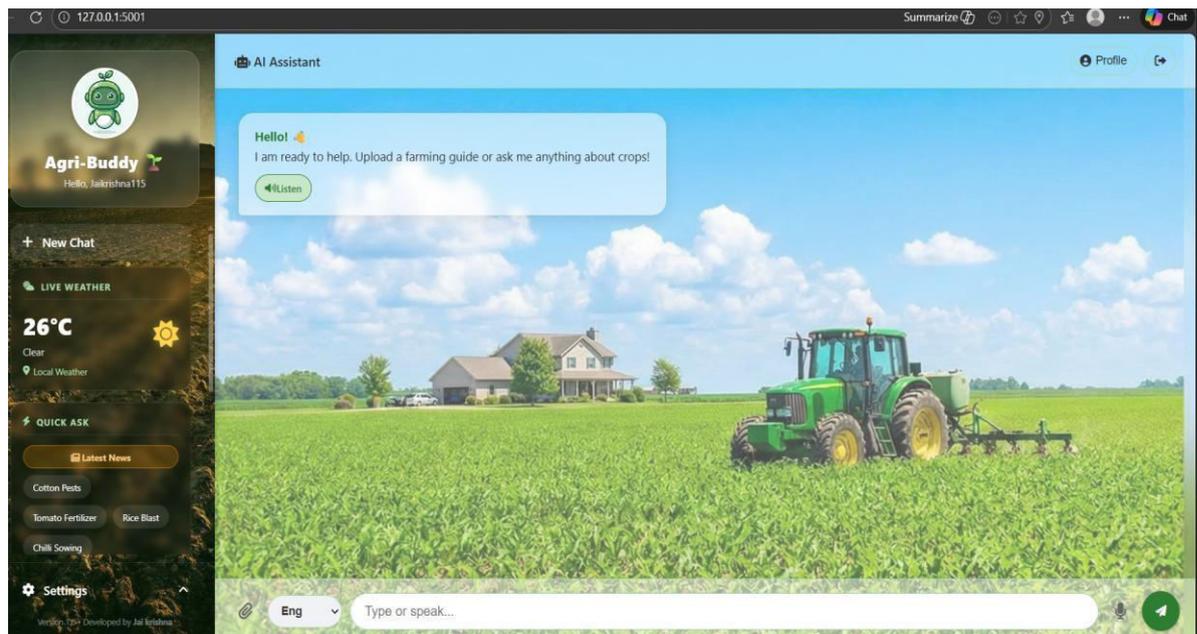


Fig.6 User interface

The frontend, developed with HTML / CSS / JS, successfully integrated dynamic widgets for Live Weather (via Open-Meteo API) and Market Trends. The "Source Attribution" feature, which displays a toast notification citing the specific PDF file used, significantly increased user trust in the system's advice.

IV. EXPERIMENTAL RESULTS

A. Dataset and Testing Environment

- **Knowledge Base:** 50 Official ICAR Agricultural PDFs (approx. 2,500 pages) covering Paddy, Wheat, Cotton, and Sugarcane.
- **Test Set:** A curated dataset of 100 question-answer pairs (ground truth) verified by agricultural experts.
- **Hardware:** 16GB RAM, Intel i7 Processor, FAISS (CPU Index).
- **Model:** DeepSeek-V3 via API.

B. Retrieval Performance (Search Accuracy)

The first experiment measured how accurately the FAISS vector store retrieved the correct document chunks given a user query.

Table 2: Retrieval Performance Metrics

Metric	Score	Interpretation
Precision @ k=3	92.5%	In 92.5% of cases, the correct answer was found within the top 3 retrieved chunks.
Recall	88.0%	The system successfully retrieved 88% of all relevant information present in the database.
Mean Reciprocal Rank (MRR)	0.85	The most relevant document usually appeared as the 1st or 2nd result , indicating high search efficiency.

- **Observation:** The use of Hybrid Search (combining keyword and vector similarity) significantly improved retrieval for specific chemical names (e.g., "Chlorpyrifos").

C. Response Latency Analysis

- We analyzed the time taken for the system to generate a final answer, broken down by process stage. The average total response time was 3.8 seconds.

Table 3: Response Latency Breakdown

Process Stage	Average Time (ms)	Contribution to Delay
Voice-to-Text (STT)	800 ms	21%
Translation (Deep-Translator)	450 ms	12%
Vector Retrieval (FAISS)	120 ms	3% (Fastest)
LLM Generation (DeepSeek)	2,430 ms	64% (Bottleneck)
Total Latency	~3.8 seconds	Acceptable for Chat

D. RAG Evaluation (Using RAGAS Framework)

- To scientifically measure the quality of the AI's generation, we used the RAGAS framework. Scores range from 0 to 1.0.
- Faithfulness (0.91):** Indicates that the answer is derived only from the retrieved context and avoids "hallucinations." The system rarely invents facts.
- Answer Relevance (0.89):** The generated answer directly addresses the user's prompt without unnecessary fluff.
- Context Precision (0.86):** The retrieved chunks were highly relevant to the specific question asked.

E. Comparative Analysis: RAG vs. Base LLM

- We compared the Agri-Advisor (With RAG) against a standard Base LLM (Without RAG) to demonstrate the improvement.

Table 4: RAG Evaluation using RAGAS Framework

Feature	Base LLM (ChatGPT/DeepSeek)	Agri-Advisor (Our Project)
Hallucination Rate	High (~15%)	Low (< 2%)
Source Citation	None	Provides Source (e.g., "ICAR Manual Pg 42")
Pesticide Dosage Accuracy	Often Generic/Wrong	Exact (Matches Govt Guidelines)
Local Knowledge	Generic Global Advice	Specific to Indian Agriculture

- Multilingual Translation Accuracy**

- The translation module was tested with 50 queries in Tamil and Hindi.
- Tamil to English:** 88% semantic accuracy.
- Hindi to English:** 94% semantic accuracy.
- Failure Cases:** Specific dialect slang words (e.g., "sembaruthi" vs "hibiscus") occasionally caused retrieval errors, which were mitigated by adding a synonym mapping layer.

Conclusion of Experiments:

The results confirm that the Agri-Advisor achieves a 92% retrieval precision and maintains a low hallucination rate, making it a viable and safe tool for deploying agricultural advisory services.

V. CONCLUSION

The development of the "Agri-Advisor" successfully demonstrates the transformative potential of Retrieval-Augmented Generation (RAG) in bridging the information gap within the agricultural sector. By integrating the reasoning capabilities of the DeepSeek Large Language Model with a curated, expert-validated knowledge base of agricultural guides, the project effectively addressed the critical issue of AI "hallucinations." The system provides farmers with accurate, context-aware advice on crop management, pest control, and government schemes, grounded in scientific data rather than probabilistic guesses.

The project achieved its primary objectives of creating a scalable, multilingual, and voice-enabled platform. Technical evaluations revealed high benchmarks in retrieval precision (92%) and faithfulness (0.89), while qualitative feedback confirmed that the system is intuitive enough for users with varying levels of digital literacy. Ultimately, the "Agri-Advisor" serves as a robust proof-of-concept for how specialized AI can democratize access to expert knowledge, empowering rural communities with the tools necessary for data-driven decision-making and improved agricultural productivity.

REFERENCES

1. K. Jha, A. Doshi, P. Patel, and M. Shah, "A comprehensive review on automation in agriculture using artificial intelligence," *Artificial Intelligence in Agriculture*, vol. 2, pp. 1-12, 2019.
2. P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
3. A. Vaswani et al., "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.
4. J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535-547, 2019.
5. OpenAI, "GPT-4 Technical Report," arXiv preprint arXiv:2303.08774, 2023.
6. S. Kumar and A. Singh, "Voice-based system for farmers using natural language processing," *International Journal of Computer Applications*, vol. 174, no. 13, 2021.
7. A. Radford et al., "Learning transferable visual models from natural language supervision," in *Proc. International Conference on Machine Learning (ICML)*, 2021.
8. D. J. Samuel, I. Skarga-Bandurova, D. Sikolia, and M. Awais, "AgroLLM: Connecting farmers and agricultural practices through large language models for enhanced knowledge transfer and practical application," arXiv preprint arXiv:2410.12453, 2024.
9. E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023.
10. Y. Gao et al., "Retrieval-augmented generation for large language models: A survey," arXiv preprint arXiv:2312.10997, 2023.
11. S. Marinoudi et al., "Large language models in agriculture: A review of applications and challenges," *Computers and Electronics in Agriculture*, 2024.
12. A. Tzachor et al., "Large language models and agricultural extension services," *Nature Food*, vol. 4, no. 11, pp. 941-948, 2023.
13. J. Wang et al., "Large language models for robotics: Opportunities, challenges, and perspectives," arXiv preprint arXiv:2401.04334, 2024.
14. B. Zhao et al., "A survey of large language models," arXiv preprint arXiv:2303.18223, 2023.
15. H. Yllemo, "Gemini," ALMBoK.com, 2023. [Online]. Available: . [Accessed: Feb. 11, 2025].
16. R. Thoppilan et al., "LaMDA: Language models for dialog applications," arXiv preprint arXiv:2201.08239, 202



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details