International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Domain-Specific vs. General Purpose Large Language Models in Agriculture: A Narrative Review of Retrieval Augmented Generation (RAG) for Evidence-Based Extension Services

Chris Adrian H. Uy, Ralph Laurence L. Vigo, Esparcia Jay Ar
Department of Computer Engineering
University of Southern Mindanao, Kabacan, Cotabato, Philippines

Abstract - The rapid development of Large Language Models (LLMs) has enabled the use of chat-based systems to support agricultural knowledge sharing, particularly for rural communities. However, general-purpose models often lack domain grounding, which can lead to inaccurate outputs and a localization gap where recommendations do not match local conditions and available resources. This study presents a narrative review guided by PRISMA methodology, analyzing literature published between 2018 and 2026 from platforms such as Google Scholar, ScienceDirect, IEEE Xplore, and MDPI. The review focuses on how LLMs are applied in agricultural extension services and examines their limitations in real-world use.

The study compares general-purpose models with domain-specific systems such as AgriGPT, AgroLLM, and AgriBERT. The findings indicate that domain-specific models perform better in terms of accuracy and relevance when trained on curated agricultural data, while general-purpose models are limited by lack of contextual grounding. Retrieval-Augmented Generation (RAG) is identified as an effective approach for improving reliability by linking outputs to verifiable sources. In addition, multimodal systems enhance performance in tasks such as crop disease detection. The reviewed literature suggests that integrating RAG-based systems may improve traceability, grounding and reliability in agriculture, particularly where evidence based extension support is critical.

Index Terms **- Data Integrity, Large Language Models, Retrieval Augmented Generation, Smart Agriculture**

I. INTRODUCTION[1]

The development of artificial intelligence (AI) has opened up many possibilities across various industries, including agriculture. One of the most significant advancements is the emergence of Large Language Models (LLMs), which can process vast amounts of data and provide useful insights through conversational interfaces. In the agricultural sector, these models have the potential to act as digital assistants, allowing farmers to access complex scientific knowledge using simple, everyday language. As digital platforms continue to evolve, AI is transforming from a simple tool into a collaborative partner that supports decision-making in farming practices..

Agricultural extension services, which traditionally rely on human experts visiting farms, have gradually integrated digital tools such as mobile applications, websites, and messaging systems. These tools help extend the reach of advisory services to more farmers, especially in remote areas. However, translating technical agricultural research into practical recommendations still requires significant manual effort and expertise.

Despite the advantages of AI, current general-purpose LLMs face major limitations when applied to agriculture. These models often lack domain-specific knowledge, which leads to inaccurate or unreliable recommendations. Because they are trained on broad datasets, they may generate responses that are not suitable for specific agricultural contexts.

Additionally, many existing systems are not locally aware, resulting in recommendations that do not align with the farmer's environment, available resources, or economic conditions. This can create serious risks, especially when AI is used to provide guidance on critical farming decisions such as pest control, fertilizer use, and crop management.

Two major gaps are evident in current AI applications for agriculture: the localization gap and the transparency gap. The localization gap refers to the inability of general-purpose models to adapt recommendations to local conditions. For

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

example, an AI system may suggest products or techniques that are unavailable or too expensive in certain regions.

The transparency gap, on the other hand, is related to the "black-box" nature of many AI systems. These models do not clearly show how they arrive at their recommendations, making it difficult for users to verify the accuracy and reliability of the information. This lack of traceability is especially problematic in agriculture, where incorrect advice can directly impact farmers' livelihoods and food security.

The primary purpose of this study is to evaluate and compare domain-specific and general-purpose Large Language Models in the context of agriculture. It focuses on how these models perform in delivering accurate, reliable, and locally relevant information for agricultural extension services. The study also aims to examine the role of Retrieval-Augmented Generation (RAG) as a solution to improve the reliability of AI systems by grounding their responses in verified technical sources.

To achieve this purpose, the study aims to:

1. Compare the domain grounding and scientific accuracy of domain-specific agricultural models and general-purpose LLMs.
2. Identify and analyze the localization and transparency gaps present in general-purpose systems.
3. Evaluate Retrieval-Augmented Generation (RAG) as a method for improving the traceability and reliability of AI-generated agricultural advice.

This study is significant because it addresses the critical need for reliable and safe AI systems in agriculture. Since AI-generated recommendations can directly affect farming decisions, ensuring their accuracy and credibility is essential. By identifying the limitations of general-purpose models and highlighting the advantages of domain-specific and RAG-based systems, this research provides guidance for developing more effective digital advisory tools.

Overall, the study contributes to improving agricultural extension services by promoting AI systems that are transparent, evidence-based, and tailored to local farming conditions, helping farmers make better-informed decisions.

II. REVIEW OF RELATED LITERATURE[2]

1. General Purpose Large Language Models (LLMs)

General-purpose Large Language Models (LLMs) are widely used AI systems trained on large and diverse datasets covering multiple domains. Because of this, they are able to perform a wide range of tasks, including answering questions, summarizing information, and generating human-like responses. In agriculture, these models are often explored as tools for improving access to knowledge, especially for farmers who may not have direct access to experts.

However, their broad training also introduces limitations. While they are effective in communication, they often lack domain grounding, which is necessary for making accurate agricultural recommendations (Chen et al., 2024). This can result in incorrect or misleading outputs, especially in high-risk areas such as crop disease management or chemical usage (Marinoudi et al., 2024). Because of this, general-purpose LLMs are better viewed as support tools for information access rather than fully reliable advisory systems.

To better understand these limitations, the following section analyzes major general-purpose LLMs based on their strengths, weaknesses, and relevance to agricultural use.

1.1 ChatGPT (OpenAI)

ChatGPT is one of the most widely used general-purpose LLMs and is often used as a baseline for comparison. It is built on a Transformer-based architecture and improved using reinforcement learning from human feedback (RLHF), which helps it generate helpful and conversational responses (OpenAI, 2023).

In agriculture, ChatGPT is useful for explaining technical concepts in simpler terms, making it helpful for knowledge dissemination and farmer education. Some efforts have also explored adapting it for agricultural advisory using RLHF-based approaches (Digital Green Academy, 2026). Because of its strong language ability, it performs well in translating research into understandable advice.

However, ChatGPT struggles in areas that require verified and precise information. As a closed system, it does not provide clear sources for its recommendations, which makes it difficult to validate its outputs. This becomes a major issue in agriculture, where incorrect advice can lead to financial loss or crop failure. Compared to other models, ChatGPT performs well in communication, but lacks traceability and domain reliability, making it less suitable for decision-critical tasks.

1.2 DeepSeek-R1 (DeepSeek)

DeepSeek-R1 represents a newer approach to LLMs by focusing more on reasoning ability rather than just language generation. It uses reinforcement learning techniques that allow the model to develop step-by-step thinking processes, such as self-correction and multi-step reasoning (DeepSeek-AI, 2025).

This makes it useful in agricultural scenarios that

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

require problem-solving, such as planning irrigation systems or analyzing farming strategies. Compared to ChatGPT, it shows improvement in handling complex queries that involve multiple steps.

Despite this, DeepSeek-R1 still depends heavily on its training data. Since its data may not include enough agriculture-specific or tropical farming information, its outputs may still lack relevance in real-world settings (DeepSeek-AI, 2025). This shows that even with better reasoning, the model can still produce incorrect or impractical recommendations.

Conceptually, DeepSeek-R1 improves how the model processes problems, but like other general-purpose models, it does not fully solve the issue of domain knowledge and localization.

1.3 Gemini (Google)

Gemini is designed as a multimodal model, meaning it can process both text and images. This is particularly useful in agriculture, where visual information—such as plant diseases or pest damage—is important (Gemini Team, 2023; Shen et al., 2024).

Because of this, Gemini can assist in tasks like crop diagnosis from images, which gives it an advantage over text-only models. Its large context window also allows it to process longer and more complex documents, which can be helpful for analyzing agricultural reports (Gemini Team, 2023).

However, similar to other general-purpose models, Gemini faces issues with localization. Its training data often reflects large-scale, industrial farming systems, which may not match the conditions of smallholder farmers (Marinoudi et al., 2024). As a result, it may recommend tools or methods that are not accessible or practical.

Compared to other models, Gemini improves input capability (multimodal), but still lacks context awareness, showing that more data types do not automatically lead to better real-world recommendations.

1.4 Llama (Meta)

Llama stands out because of its open-weight design, which allows developers to modify and fine-tune the model for specific use cases (Llama Team, 2024). This makes it more flexible compared to closed models like ChatGPT and Claude.

In agriculture, this flexibility is important because the model can be adapted using local datasets, making it more relevant to specific crops, regions, or farming practices. This has led to the development of systems like AgroLLM, which combine research knowledge with practical farming needs (Ravindran et al., 2026).

However, Llama itself does not automatically solve agricultural challenges. Its effectiveness depends on how well it is trained and customized. Without proper domain-specific data, it can still behave like a general-purpose model (De Clercq, 2024). Conceptually, Llama is different from other models because it provides a foundation for specialization, rather than acting as a complete solution on its own.

1.5 Claude (Anthropic)

Claude focuses on safety and reliability, using a framework called Constitutional AI to guide its responses and reduce harmful outputs (Bai et al., 2022). This makes it suitable for applications where incorrect information can have serious consequences.

In agriculture, this safety feature is useful because it reduces the risk of giving dangerous recommendations, especially in areas like chemical usage or pest control. Compared to other models, Claude is more cautious in its responses.

However, like ChatGPT, Claude is still a closed system, meaning it does not provide clear sources or explanations for its outputs. This limits its usefulness in professional agricultural settings, where users need evidence-based recommendations (Sawant et al., 2026).

In comparison, Claude improves safety, but does not address transparency, which is equally important in agricultural decision-making.

1.6 Synthesis and Comparison

When comparing these general-purpose LLMs, it becomes clear that each model improves a specific aspect of AI:

ChatGPT → strong in communication
DeepSeek-R1 → improved reasoning
Gemini → multimodal capability
Llama → adaptability and customization
Claude → safety and alignment

Despite these improvements, all models share common limitations when applied to agriculture. These include:

- Lack of domain-specific knowledge, leading to inaccurate recommendations (Chen et al., 2024)
- Localization issues, where outputs do not match local farming conditions (Marinoudi et al., 2024)
- Lack of transparency, making it difficult to verify information (Sawant et al., 2026)

These limitations show that general-purpose LLMs are not fully suitable for evidence-based agricultural advisory systems. While they can support communication and basic guidance, they require additional systems to improve accuracy and reliability.

Because of this, there is a need to move toward domain-specific models and approaches such as Retrieval-Augmented Generation (RAG), which can connect AI outputs

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

to verified sources and local data (Sawant et al., 2026; Yang et al., 2025).

2. Domain Specific LLMS

Domain-specific Large Language Models were developed to address a major weakness of general-purpose systems: their tendency to produce broad but weakly grounded outputs when applied to specialized fields. In agriculture, this limitation becomes more serious because correct advice depends on technical accuracy, local conditions, and the practical constraints faced by farmers. Unlike general-purpose LLMs, domain-specific models are designed to learn from curated agricultural sources, making them more capable of producing outputs that reflect field-based knowledge rather than general internet patterns (Awais et al., 2025; Bang et al., 2023).

The main advantage of domain-specific LLMs is not simply that they are trained on more relevant data, but that their training process is aligned with a narrower objective. This allows them to perform better in tasks that require agricultural terminology, crop-related reasoning, and context-sensitive recommendations. However, this specialization also creates a trade-off. While domain-specific models are often more accurate within their target field, they may be less flexible outside that field and may still struggle when the available agricultural data are limited, outdated, or unevenly distributed across regions (Ling et al., 2025; Marinoudi et al., 2024).

For agriculture, this is an important distinction. A model may appear useful because it understands technical terms, but its value depends on whether it can generate advice that is both scientifically grounded and practically usable in local farming conditions. This is why domain specialization is not just a technical improvement but a necessary response to the limits of general-purpose LLMs in evidence-based advisory systems.

2.1 AgroLLM

AgroLLM represents one of the clearest examples of how domain specialization can improve agricultural knowledge transfer. Rather than depending on broad pretraining alone, it integrates structured agricultural knowledge with language model capabilities to support more field-relevant responses (Ravindran et al., 2026). This is important because agriculture is not a single uniform domain. It includes crop science, soil management, pest control, climate variability, and farm economics, all of which require context-aware interpretation rather than generic language generation.

The value of AgroLLM lies in how it narrows the gap between academic knowledge and practical use. General-purpose models may explain agricultural concepts well, but they often fail when asked to translate these concepts into actionable advice. AgroLLM attempts to solve this by using curated sources and domain framing, which improves relevance and reduces the chance of unsupported outputs. In this sense, the model is not only a language tool but also a knowledge alignment system.

At the same time, AgroLLM also shows that domain specialization is not a complete solution. Its performance still depends on the quality, coverage, and freshness of the agricultural sources used during construction. If the underlying corpus is incomplete or biased toward certain crops, regions, or practices, the model may still generate narrow or uneven advice. This means that domain specificity improves reliability, but only when the training data are sufficiently diverse and well maintained (Ravindran et al., 2026; Ling et al., 2025).

2.1.1 Source Selection and Filtering

The source selection process used in AgroLLM is important because it directly affects the quality of the model's outputs. By prioritizing textbook materials, recent references, and content that supports agronomic reasoning, the developers increased the likelihood that the model would learn from authoritative and instructional sources rather than from noisy or unverified text (Ravindran et al., 2026). This is a meaningful improvement over general-purpose training because it reduces the chance that the model will reproduce irrelevant or low-quality patterns.

However, the filtering process also reveals a limitation common to many domain-specific systems: quality control depends heavily on manual curation. While removing front matter, indices, duplicates, and irrelevant passages improves dataset quality, it also requires sustained expert effort. This means that the effectiveness of a domain-specific model is tied not only to model design but also to the availability of domain experts who can maintain the dataset. In practice, this raises the cost of scaling such systems to new crops, languages, or regions.

2.2 AgroGpt

AgroGPT highlights a different direction in domain specialization by focusing on multimodal agricultural understanding. Its design reflects the reality that many agricultural problems are visual as well as textual. Farmers often need help identifying weeds, pests, diseases, or damaged crops from images, and text-only systems are limited in this type of task (Awais et al., 2025). AgroGPT addresses this gap by using an image-based instruction-tuning approach, which gives it more relevance in real-world diagnosis tasks.

This is significant because it shows that domain-specific LLMs are not only about subject matter, but also about matching the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

type of input used in the domain. Agriculture frequently depends on visual inspection, so a model that can interpret images and generate useful explanations has a stronger practical role than a text-only assistant. In other words, AgroGPT demonstrates that domain specialization can improve not just accuracy, but also usability.

Still, the model also reflects a broader challenge in agricultural AI: performance depends on the availability of paired data. Since image-text datasets in agriculture are limited, the system must rely on synthetic instruction generation and carefully designed training pipelines. This improves capability, but it also introduces dependence on the quality of the generated instructions and the representativeness of the visual dataset. As a result, AgroGPT is promising, but its reliability in diverse field conditions still depends on continued validation (Awais et al., 2025).

2.3 AgriGpt

AgriGPT strengthens the argument for domain-specific systems by showing how structured agricultural data and benchmarking can improve model performance. Its development process uses a data engine and a Tri-RAG framework to build a large instruction dataset that covers multiple agricultural tasks (Yang et al., 2025). This is important because one of the major weaknesses of general-purpose LLMs is their inability to consistently handle specialized tasks with domain precision.

Compared with simpler adaptation methods, AgriGPT's ecosystem approach is more complete. It does not only train a model; it also builds the surrounding data and evaluation structure needed for agricultural deployment. This makes the system more suitable for research and applied use because it can be tested against actual agricultural tasks rather than only general language benchmarks. That difference matters because a model can score well on general NLP measures while still failing in farm-specific decision making.

Even so, the model still faces the same issue found in other specialized systems: agricultural knowledge changes across location, season, and crop type. A model trained on one set of practices may not transfer perfectly to another region with different climate or farming methods. This suggests that domain-specific LLMs improve task fit, but they still need localization mechanisms and frequent updates to stay practically useful (Yang et al., 2025; Marinoudi et al., 2024).

2.4 Yunnan Arabica Coffee cultivation

The Yunnan Arabica coffee study is useful because it demonstrates how retrieval-based grounding can improve the trustworthiness of agricultural LLM systems. Instead of depending only on pretrained knowledge, the system retrieves supporting evidence and uses reranking to improve answer quality (Chen et al., 2025). This matters because agricultural advice must often be traceable. Users need to know not only what the model recommends, but also where the recommendation came from.

This approach addresses one of the central weaknesses of general-purpose LLMs: unsupported confidence. A general model may give a polished answer even when it lacks direct evidence, but a retrieval-augmented system can anchor the response to a source base. In agriculture, that difference is critical because wrong advice can affect crop health, productivity, and input costs. The Yunnan system therefore shows that retrieval does not merely improve technical performance, but also increases auditability and practical trust.

At the same time, the study shows that retrieval quality depends on vocabulary, indexing, and reranking. This means that RAG is not automatically reliable. If the retrieval base is weak or poorly localized, the generated answer may still be incomplete or misleading. The study therefore supports a more balanced conclusion: RAG improves domain-specific LLMs, but its value depends on the quality of the underlying agricultural knowledge base (Chen et al., 2025).

2.5 Synthesis of Domain-Specific LLMs

Taken together, the reviewed studies show that domain-specific LLMs are more effective than general-purpose models when the task requires technical accuracy, agricultural context, and evidence-based reasoning. Their main strength is not broader language ability, but tighter alignment between model behavior and domain needs. AgroLLM improves knowledge grounding, AgroGPT expands visual understanding, AgriGPT strengthens task coverage, and the Yunnan system demonstrates the value of retrieval-based evidence support.

However, the review also shows that domain specificity does not eliminate all weaknesses. These models still face problems related to data quality, localization, maintenance, and scalability. A model may be highly accurate in one agricultural context but less effective in another if local crop types, practices, or resource levels are not represented in the training or retrieval data. This suggests that the real advantage of domain-specific LLMs is not perfection, but a better starting point for reliable agricultural AI.

From a broader perspective, the literature indicates that the most effective agricultural systems are likely to combine specialization with retrieval, validation, and localization. Domain-specific LLMs improve relevance, but their long-term

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

value depends on whether they can remain updated, transparent, and adaptable to changing farm conditions. This is why they should be viewed not as final solutions, but as more suitable foundations for evidence-based agricultural advisory services than general-purpose LLMs.

III. METHODOLOGY

This study adopts a narrative review guided by the PRISMA framework to ensure transparency, reproducibility, and methodological rigor in the selection and analysis of literature. While PRISMA is commonly used for systematic reviews, its structured approach is applied in this study to strengthen the credibility of the narrative synthesis.

Unlike a purely descriptive review, this study applies comparative analytical criteria to evaluate the performance of general-purpose and domain-specific Large Language Models in agricultural contexts, with emphasis on accuracy, localization, and transparency.

A. Identification of Literature

The procedure of identification of literature began with gathering academic materials obtained from digital libraries and technical archives. The academic materials are then identified through filtering via queries such as : "Large Language Models, Retrieval Augmented Generation, and Smart and Precise Agriculture". The search terms are used to select high quality papers from reputable databases including MDPI, PAGEPress, Google Scholar, Science Direct- Elsevier, ResearchGate, and IEEEXplore. The databases were selected to capture peer-reviewed and technically relevant literature across AI and agricultural technology .Table 1 shows the different sources where the academic materials are obtained.

TABLE 1 Digital Databases and Search Platforms

PLATFORM	TYPE	URL
MDPI	Publisher Platform	https://www.mdpi.com (accessed on 4 February, 2026)
PAGEPress	Publisher Platform	https://www.pagepress.com (accessed on 5 February, 2026)
Google Scholar	Search Engine	https://www.scholar.google.com (accessed on 3 February, 2026)
Science Direct- Elsevier	Digital Library	https://www.sciencedirect.com (accessed on 4 February, 2026)
IEEEXplore	Digital Library	https://ieeexplore.ieee.org (accessed on 23 February, 2026)

B. Screening and Eligibility Criteria

The literature search was executed through a structured application of the boolean terms and search queries detailed in Table 2. These specific queries were formulated to capture the intersection of advanced computational models and agricultural sciences, ensuring a comprehensive retrieval of relevant academic materials. The selection of these terms was guided by the primary research objective: to identify literature that distinguishes between general-purpose and domain-specific applications within the agricultural sector.

By combining broad technical descriptors like "Large Language Models" and "Artificial Intelligence" with domain-specific keywords such as "cultivation," "farming," and "ecosystem," the search strategy was designed to bridge the gap between high-level AI research and practical agronomic needs. This methodology ensures that the included papers provide the technical configurations and model designs necessary to evaluate the safe deployment of AI in evidence-based extension services. The use of digital libraries, including MDPI, IEEE Xplore, and ScienceDirect, allowed for the rigorous application of these queries across diverse technical archives.

TABLE 2 Search Queries Used for the Paper

	Search Queries (SQ)
SQ1	"LLMs AND agriculture OR ecosystem"
SQ2	"LLMs AND agriculture OR cultivation"
SQ3	"LLMs AND AI AND machine learning OR deep learning"
SQ4	LLMs OR AI AND farming AND machine learning OR deep learning"

C. Inclusion and Exclusion Criteria

In order to obtain the final research papers, the PRISMA protocol and principles were followed in order to form a set of Inclusion Criteria as well as the Exclusion Criteria. The inclusion criteria (IC) were defined to ensure that the selected literature provides sufficient technical depth and relevance to current developments in agricultural AI. The review focused on studies that explicitly examine or compare general-purpose and domain-specific Large Language Models within agricultural contexts. Priority was given to research that includes model design, system architecture, or practical implementation, particularly those involving Retrieval-Augmented Generation (RAG). This requirement ensures that the analysis is based on verifiable technical foundations rather than purely descriptive discussions. By selecting studies with clear methodological detail, the review enables a more precise evaluation of how model structure and training approach

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

influence the accuracy and reliability of agricultural knowledge delivery.

The exclusion criteria (EC) were applied to maintain the reliability and practical relevance of the findings. In addition to removing duplicate, outdated, or inaccessible studies, the review excluded papers that lack agricultural context, geographical awareness, or any form of validation. Studies that present generalized AI outputs without consideration of local farming conditions were also omitted. This is important because agricultural recommendations are highly context-dependent, and unsupported or non-localized outputs may lead to incorrect or impractical guidance. By filtering out studies that do not address these factors, the review ensures that the final dataset reflects systems capable of supporting evidence-based and context-aware agricultural extension services. The criteria shown on Table 3 collectively strengthen the analytical foundation of the study and support a more meaningful comparison between general-purpose and domain-specific LLMs.

TABLE 3 Inclusion Criteria & Exclusion Criteria

	List of Inclusion and Exclusion Criteria
Inclusion Criteria (IC)	
IC1	Should contain at least one of the keywords
IC2	Published within the last 8 years (2018 – 2026)
IC3	Research that is being examined should have matching title, abstract, and full text
Exclusion Criteria (EC)	
EC1	Redundant items
EC2	Whole text of paper cannot be taken
EC3	Non-english documents
EC4	Purpose of the paper is not related to LLMS

D. Prisma Diagram

Figure 1 illustrates the PRISMA flow diagram, which serves as the structural roadmap for the literature selection process. This systematic approach was employed to ensure that the transition from a broad initial search to a final focused dataset was transparent, reproducible, and free from selection bias.

The identification phase yielded a total of 152 records across the selected digital databases. During the initial screening stage, duplicates and studies with titles or abstracts unrelated to Large Language Models in agriculture were removed, leaving 97 records for intensive evaluation. This stage isolated research that specifically addressed the technical intersection of AI and agronomy, rather than general computer science or broad environmental topics.

The remaining 97 papers underwent a full-text eligibility assessment based on the predefined inclusion and exclusion criteria. This stage ensures each study provides sufficient technical configurations and results in 50 studies selected for the final review. The use of the PRISMA-guided process strengthens the reliability of the findings: it ensures that the synthesis is built upon a foundation of localized, safe, and technically sound literature suitable for evidence-based extension services.

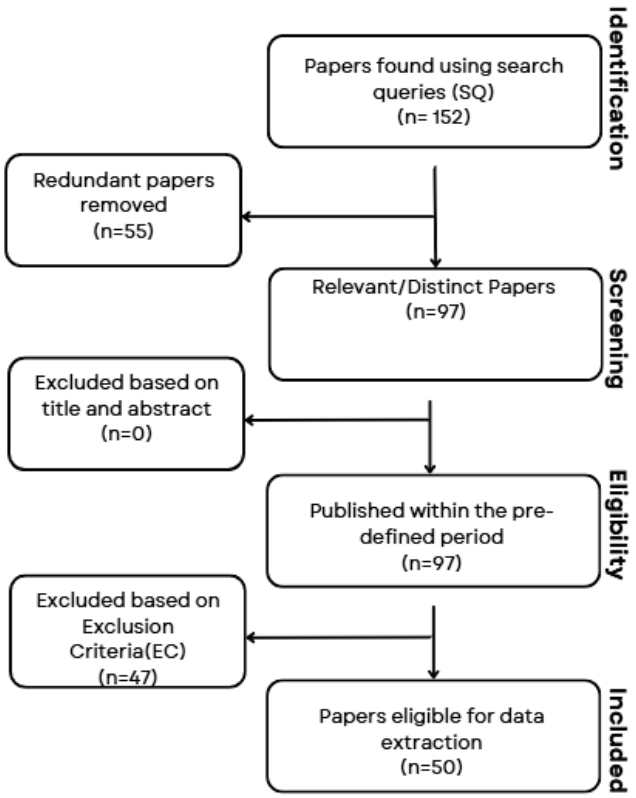

Figure 1. Flow diagram of review using PRISMA

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

E. Performance Gaps for LLMs

Table 4 presents the comparative framework used to evaluate general-purpose and domain-specific Large Language Models (LLMs) in agricultural applications. The table organizes the selected models according to key attributes, including model category, primary strengths, limitations, and relevance to agricultural use cases.

This structured comparison was used to identify performance differences across model types based on predefined evaluation dimensions, such as domain grounding, localization, transparency, and practical applicability. General-purpose models were assessed in terms of their reasoning and language capabilities, while domain-specific models were evaluated based on their alignment with agricultural knowledge and task-specific performance.

The inclusion of both model categories allows for a consistent basis of comparison, supporting the identification of strengths and limitations across different system designs. This framework provides a structured foundation for analyzing performance gaps and informs the discussion of how domain specialization and retrieval-based approaches contribute to improving reliability in agricultural advisory systems.

TABLE 4 Comparison of General Purpose and Domain Specific LLMs

Model	Category	Strength	Limitation	Gap in Agriculture	Relevance to Agriculture
ChatGPT	General-purpose	Strong conversational ability	Limited traceability	Lacks verifiable outputs	Education and basic explanations
DeepSeek-R1	General-purpose	Strong reasoning	Weak domain grounding	Poor agricultural accuracy	Planning and structured thinking
Gemini	General-purpose (multimodal)	Text and image processing	Localization issues	Weak adaptation to local farming	Visual and text-based support
Claude	General-purpose	Safety-focused responses	Low source transparency	Limited evidence report	General guidance
Llama	Open-weight general model	Customizable	Depends on customization quality	Inconsistent without domain data	Domain adaptation
AgroLLM	Domain-specific	Strong agricultural grounding	Depends on curated data quality	Limited by dataset scope	Knowledge transfer
AgroGPT	Domain-specific (multimodal)	Handles image based agricultural tasks	Limited dataset coverage	Incomplete field representation	Crop and pest diagnosis
AgriGPT	Domain-specific	Strong task coverage and benchmarking	Scalability challenges	Needs continuous up	Agricultural decision support

IV. RESULTS AND DISCUSSION

The results of this review indicate a clear performance difference between general-purpose and domain-specific Large Language Models when applied to agriculture. General-purpose models such as DeepSeek-R1 and GPT-4 demonstrate strong capabilities in reasoning, language processing, and handling multilingual inputs. However, these strengths are not directly associated with accuracy in agricultural applications. The findings suggest that their limitations are mainly due to the lack of domain-specific data, which may lead to outputs that are not aligned with actual farming conditions. For example, recommendations related to crop management, pest control, or chemical usage may be too general or not applicable to local environments, which reduces their reliability in real-world use.

This limitation is closely associated with the previously identified localization and transparency gaps. Because general-purpose models are trained on broad datasets, they may produce responses that appear correct but are not grounded in verified agricultural knowledge. This increases the risk of hallucinations, where incorrect or unsupported information is presented as factual. As a result, while these models are useful for general communication and knowledge access, they may not be suitable for decision-critical agricultural tasks.

In contrast, domain-specific models such as AgriGPT, AgroLLM, and AgriBERT show improved performance because they are trained on curated agricultural datasets. The results indicate that these systems are better at handling technical agricultural concepts and generating more relevant recommendations. This suggests that domain specialization plays a key role in improving both accuracy and usability. However, it is also observed that domain-specific systems may still face limitations in scalability and data availability, especially when applied to different regions or crops.

Another important finding of this study is the effectiveness of Retrieval-Augmented Generation (RAG). The results suggest that integrating RAG into LLM systems may significantly improve reliability by linking outputs to verified technical sources. Instead of relying only on pre-trained knowledge, RAG allows models to retrieve updated and context-specific information, which improves both transparency and traceability. This is particularly important in agriculture, where decisions must be supported by evidence.

In addition, multimodal systems such as AgroGPT indicate improvements in tasks that require visual understanding, such as crop disease detection. The ability to process both text and images is associated with higher accuracy in identifying field-level problems. This suggests that combining multimodal capabilities with domain-specific grounding and RAG may further enhance the performance of AI systems in agriculture.

Overall, the findings indicate that while general-purpose LLMs provide a strong foundation, their effectiveness in agriculture is limited without additional mechanisms. The

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

integration of domain-specific data and retrieval-based approaches is strongly associated with improved accuracy, reliability, and practical usability in agricultural extension services.

V. CONCLUSION

The findings of this study suggest that the application of Large Language Models in agriculture requires more than general language and reasoning capabilities. While general-purpose models such as GPT-4 and DeepSeek-R1 are effective in communication and problem-solving, they are limited by their lack of domain grounding and transparency. This indicates that relying solely on these models for agricultural advisory may lead to unreliable or impractical recommendations.

The study further indicates that domain-specific models may improve performance by incorporating curated agricultural knowledge, allowing for more accurate and relevant outputs. In addition, the integration of Retrieval-Augmented Generation (RAG) is associated with increased reliability, as it enables models to link responses to verifiable sources. This approach may also improve transparency, which is important for building trust in AI-based advisory systems.

Overall, the results suggest that the most effective approach is not to replace general-purpose models, but to enhance them through domain specialization and retrieval-based methods. This combination may improve the quality of agricultural extension services by providing recommendations that are both accurate and evidence-based. Future work may focus on expanding localized datasets and improving system integration to ensure that these technologies are accessible and practical for real-world farming communities

REFERENCES

[1] M. Awais et al., "AgroGPT: Efficient agricultural vision-language model with expert tuning," arXiv preprint arXiv:2410.08405 [Online]. Available: https://arxiv.org/pdf/2410.08405
[2] Y. Bai et al., "Constitutional AI: Harmlessness from AI Feedback," arXiv preprint arXiv:2212.08073 [Online]. Available: https://arxiv.org/abs/2212.08073
[3] Y. Bang et al., "A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity," 2023 [Online]. Available: https://aclanthology.org/2023.ijcnlp-main.45.pdf
[4] Z. Chen et al., "A RAG-augmented LLM for Yunnan Arabica coffee cultivation," Agriculture, vol. 15, no. 22, 2025 [Online]. Available: https://www.mdpi.com/2077-0472/15/22/2381
[5] Z. Z. Chen et al., "A survey on large language models for critical societal domains," arXiv preprint arXiv:2405.01769 [Online]. Available: https://arxiv.org/pdf/2405.01769
[6] M. T. Chiu et al., "Agriculture-vision: A large-scale agriculture strategy dataset for semantic segmentation of high-resolution aerial imagery," arXiv preprint arXiv:1909.07606 [Online]. Available: https://arxiv.org/pdf/1909.07606
[7] D. De Clercq, "Large language models can help boost food production," Frontiers in Artificial Intelligence, vol. 7, p. 1326153, 2024 [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2024.1326153/full
[8] DeepSeek-AI, "DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning," arXiv preprint arXiv:2501.12948 [Online]. Available: https://arxiv.org/abs/2501.12948
[9] Digital Green Academy, "Application of reinforcement learning from human feedback for agricultural advisory," CGIAR, 2026 [Online]. Available: https://cgspace.cgiar.org/items/6b1d14c3-ec96-40e0-8c3b-dad0dae3991c
[10] Gemini Team, Google, "Gemini: A Family of Highly Capable Multimodal Models," arXiv preprint arXiv:2312.11805 [Online]. Available: https://arxiv.org/abs/2312.11805
[11] X. He, "Enhancing agriculture QA models using large language models (LLMs)," BIO Web of Conferences, vol. 61, p. 01005, 2024 [Online]. Available: https://www.bio-conferences.org/articles/bioconf/pdf/2024/61/bioconf_isaeb2024_01005.pdf
[12] B. Kariyanna and M. Sowjanya, "Unravelling the use of artificial intelligence in management of insect pests," Smart Agricultural Technology, vol. 8, p. 100517, 2024 [Online]. Available: https://doi.org/10.1016/j.atech.2024.100517
[13] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine Learning in Agriculture: A Review," Sensors, vol. 18, no. 8, p. 2674, 2018 [Online]. Available: https://doi.org/10.3390/s18082674
[14] C. Ling et al., "Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey," arXiv preprint arXiv:2305.18703, 2025 [Online]. Available: https://arxiv.org/html/2305.18703v7#S1
[15] W. Liu et al., "K-BERT: Enabling Language Representation with Knowledge Graph," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 3, 2020, pp. 2901–2908 [Online]. Available: https://doi.org/10.1609/aaai.v34i03.5681
[16] Llama Team, Meta, "The Llama 3 Herd of Models," arXiv preprint arXiv:2407.21783 [Online]. Available: https://arxiv.org/abs/2407.21783
[17] V. Marinoudi et al., "Large language models impact on agricultural workforce dynamics: Opportunity or risk?" Smart Agricultural Technology, vol. 9, p. 100677, 2024 [Online]. Available: https://doi.org/10.1016/j.atech.2024.100677
[18] OpenAI, "GPT-4 Technical Report," arXiv preprint arXiv:2303.08774, 2023 [Online]. Available: https://arxiv.org/abs/2303.08774
[19] D. J. S. Ravindran et al., "AGROLLM: Connecting farmers and agricultural practices through large language models for enhanced knowledge transfer and practical application," AgriEngineering, vol. 8, no. 1, p. 38, 2026 [Online]. Available: https://doi.org/10.3390/agriengineering8010038
[20] S. Rezayi et al., "AgrIBERT: Knowledge-Infused Agricultural Language Models for Matching Food and Nutrition," in Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, 2022, pp. 5150–5156 [Online]. Available: https://doi.org/10.24963/ijcai.2022/715
[21] O. Rozenstein et al., "Data-driven agriculture and sustainable farming: friends or foes?" Precision Agriculture, vol. 25, no. 1, pp. 520–531, 2023 [Online]. Available: https://doi.org/10.1007/s11119-023-10061-5
[22] S. Sawant et al., "Empowering Farmers With Artificial Intelligence: A Retrieval-Augmented Generation Based Large Language Model Advisory Framework," Journal of Agricultural Engineering, 2026 [Online]. Available: https://doi.org/10.4081/jae.2026.1908
[23] S. Sengupta et al., "MAG-V: A multi-agent framework for synthetic data generation and verification," arXiv preprint arXiv:2412.04494, 2025 [Online]. Available: https://doi.org/10.48550/arXiv.2412.04494
[24] Y. Shen et al., "Harnessing large vision and language models in agriculture," Frontiers in Artificial Intelligence, vol. 7, p. 1243642, 2024 [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12436425
[25] B. Yang et al., "AgriGPT: A large language model ecosystem for agriculture," arXiv preprint arXiv:2508.08632, 2025 [Online]. Available: https://arxiv.org/pdf/2508.08632