

Review

Intelligent Agents for Smart Agriculture: Architectures, Applications, and Future Challenges

Wenzheng Tao ¹, Qiwei Sang ¹, Cong Chen ² and Qirong Mao ^{1,3,4,*}

¹ School of Computer Science and Communication Engineering, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, China; 2212508023@stmail.ujs.edu.cn (W.T.); 2212508017@stmail.ujs.edu.cn (Q.S.)

² Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, NO.100 Liuying, Xuanwu District, Nanjing 210014, China; chencong@caas.cn

³ Jiangsu Engineering Research Center of Big Data Ubiquitous Perception and Intelligent Agricultural Applications, Zhenjiang 212013, China

⁴ Key Laboratory of Computational Intelligence and Low-Altitude Digital Agricultural New Technology of Jiangsu Universities, Zhenjiang 212013, China

* Correspondence: mao_qr@ujs.edu.cn; Tel.: +86-0511-8878-0371

Abstract

Intelligent agents are emerging as an important system-level paradigm for smart agriculture. This review focuses on modern agricultural intelligent agents driven by large language models and related multimodal foundation models and examines how this emerging field is reshaping the organization of intelligent agricultural systems. It first clarifies the conceptual boundaries of agricultural intelligent agents and distinguishes them from traditional multi-agent systems, agent-based modeling, agricultural foundation models, and static retrieval-augmented question-answering systems. It then synthesizes their architectural foundations, key capabilities, application scenarios, deployment challenges, and future research directions. The reviewed literature indicates that agricultural intelligent agents are moving beyond isolated perception, prediction, and response generation toward the goal-oriented coordination of agricultural knowledge, dynamic data, external tools, and decision-making processes across agricultural task chains. They are beginning to support more integrated forms of knowledge services, crop monitoring and diagnosis, decision support, and farm-level collaborative management. Nevertheless, their transition from prototype systems to dependable and deployable agricultural systems remains constrained by context-aware knowledge grounding, heterogeneous data and tool integration, long-horizon reliability, the stability of multi-agent collaboration, and system security. This review further introduces an assessment perspective based on evidence reported in the original studies, comparing representative agricultural intelligent agents in terms of task decomposition, agronomic evidence applicability, tool-use validity, workflow reliability, multi-agent coordination, and deployment-related evidence. By distinguishing demonstrated capabilities from unevaluated dimensions, this review provides a structured framework for understanding the current status of agricultural intelligent agents and for guiding their future development toward reliable, deployable, and domain-oriented intelligent systems for smart agriculture.

Keywords: agricultural intelligent agents; smart agriculture; agent technology; multi-agent collaboration; agricultural decision support



Academic Editor: Jianqiang Lu

Received: 20 July 2026

Revised: 17 August 2026

Accepted: 21 August 2026

Published: 23 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

With the rapid development of artificial intelligence, large language models, multimodal foundation models, and agent technologies, agricultural intelligent agents are gradually emerging as an important new direction in smart agriculture. Unlike conventional agricultural AI applications that are mainly designed for single tasks such as recognition, prediction, or question answering, agricultural intelligent agents provide a new system-level paradigm for agricultural knowledge understanding, complex task organization, and decision support. In recent years, research on agricultural foundation models has continuously expanded the application scope of foundation models in agricultural knowledge understanding, multimodal information processing, and complex decision support [1,2]. Meanwhile, rapid advances in LLM-based agents, particularly in task planning, tool use, memory management, and feedback-driven refinement, have provided an important methodological foundation for agricultural intelligent agents to evolve from “generative responses” toward “task-driven workflow orchestration” [3,4]. Driven by these advances, agricultural intelligent agents are expected to promote the transition of smart agriculture from isolated model applications toward task-oriented knowledge organization, tool orchestration, and intelligent decision-making.

The motivation for writing this paper stems from the urgent need for efficient, intelligent, and sustainable decision-support systems in agricultural production. With global population growth, the intensifying impacts of climate change, and increasingly severe constraints on agricultural resources and the environment, agricultural production systems are facing increasing demands in terms of production efficiency, resource utilization, environmental sustainability, and management complexity [5,6]. In this context, agricultural intelligent agents provide a new technical pathway for integrating agricultural domain knowledge, dynamic environmental information, external analytical tools, and management decisions, thereby offering new possibilities for improving the organizational capacity and intelligent service capacity of agricultural production. However, existing studies have mainly focused on agricultural foundation models, task-specific agricultural AI applications, or general-purpose LLM-based agents, while a systematic review of LLM-driven agricultural intelligent agents in smart agriculture remains limited. Accordingly, this paper reviews modern agricultural intelligent agents in the context of smart agriculture, with a particular focus on large language model-driven agricultural intelligent agent systems that are capable of knowledge integration, tool use, task organization, and collaborative decision-making.

The literature discussed in this review was organized around the research boundary of modern agricultural intelligent agents. Priority was given to studies that connect agricultural tasks with foundation-model-driven agent capabilities, such as task-oriented reasoning, external resource use, workflow organization, or collaborative decision-making. Related studies on traditional agents, agricultural AI models, foundation models, and retrieval-augmented systems were also considered when they helped clarify conceptual boundaries or provide methodological context.

Specifically, this paper first traces the evolution of the agent concept and defines the core characteristics and research boundaries of agricultural intelligent agents. It then summarizes the basic architectures and key capabilities of modern agents. Subsequently, using agricultural task chains as the main analytical thread, this paper reviews recent research progress on agricultural intelligent agents across representative agricultural scenarios. To avoid treating heterogeneous prototypes as equivalent examples of agentic systems, this review further develops an assessment matrix based on evidence reported in the original studies, distinguishing demonstrated capabilities from dimensions that remain unevaluated. Finally, the paper discusses the key challenges faced by agricultural intelligent agents

in real-world agricultural applications and outlines future research directions. This review aims to provide a clear conceptual foundation, technical trajectory, and evidence-based analytical framework for agricultural intelligent agent research, while offering guidance for future development and application in agriculture-related fields.

2. Overview of Agents

Since the mid-1990s, agents have been a prominent topic in computer science and artificial intelligence. The core concern of this field is how systems can exhibit autonomy, goal-oriented behavior, and interactive capabilities within a given environment [7]. Along this line, agent-based computing, software agents, multi-agent systems, and agent-based modeling have developed as related but distinct branches of traditional agent research [8]. In recent years, with the increasing capabilities of large language models in knowledge-based problem solving, complex reasoning, task planning, and tool use, agent research has been undergoing a new paradigm shift. The focus is gradually expanding from rule-driven mechanisms and interactions within closed environments to intelligent execution systems that use large language models as cognitive cores and are capable of multi-step planning, external tool invocation, process feedback, and dynamic refinement around specific goals [3].

For smart agriculture, this paradigm shift is of particular significance. Agricultural tasks are often characterized by the coupling of multi-source information, dynamic external environments, complex knowledge dependencies, and continuous decision-making processes. Standalone models, static knowledge-based question-answering systems, or closed rule-based systems are often insufficient to support complete task chains in real-world agricultural production. Therefore, before discussing agricultural intelligent agents, it is necessary to first review the concept and development of agents, and then further define the core characteristics and research boundaries of agricultural intelligent agents.

2.1. Concept and Development of Agents

Traditional agents do not refer to a single technical form, but rather represent a research paradigm that has gradually developed around autonomous behavior, environmental interaction, and coordination in complex systems. From the perspective of research focus, early agent studies can be broadly grouped into three representative directions. The first direction focuses on individual behavior and cognitive mechanisms, including reactive agents that emphasize stimulus–response behaviors in the environment, BDI agents based on the Belief–Desire–Intention architecture, and interface agents designed to assist users in specific tasks [9–11]. The second direction concerns Multi-Agent Systems (MAS), which focus on communication, negotiation, cooperation, and distributed decision-making among multiple autonomous entities in a shared environment [8]. The third direction is Agent-Based Modeling (ABM), which treats agents as simulation units with behavioral rules and analyzes the evolution and emergent phenomena of macroscopic systems by modeling individual decision-making and local interactions [12]. Therefore, the core value of traditional agents lies mainly in rule-based autonomy, structured decision-making, distributed coordination, and complex system modeling. Their implementation typically depends on manually designed state representations, behavioral logic, and interaction mechanisms.

In contrast, the recent emergence of LLM-based agents has promoted the evolution of agents from “autonomous units under predefined rules” to “foundation model-driven systems for open-ended task execution”. ReAct organizes reasoning and acting in an interleaved manner, enabling models to continuously advance tasks through a “thought–action–observation” loop and revise subsequent steps according to external feedback [13]. Toolformer further demonstrates that language models can learn when to call external

tools, how to pass parameters, and how to incorporate tool-returned results, thereby overcoming the limitations of completing tasks solely based on parametric knowledge [14]. Thus, the key difference between modern LLM-based agents and traditional agents is not merely the change of the underlying model, but a transformation in the way intelligence is organized. The former uses large language models or foundation models as cognitive cores and emphasizes open-ended semantic understanding, multi-step task planning, external tool use, and feedback-driven process refinement. The latter, by contrast, relies more heavily on explicit rules, predefined protocols, or simulation mechanisms to achieve autonomy, coordination, or modeling within relatively closed task boundaries. Therefore, the term “agent” discussed in the following sections of this paper mainly follows the meaning of modern LLM-based or foundation model-driven agentic AI.

The conceptual distinctions among these agent-related paradigms are summarized in Table 1.

Table 1. Conceptual distinctions among agent-related paradigms discussed in this review.

Paradigm	Main Research Focus	Typical Form of Intelligence Organization	Role in This Review
Traditional individual agents	Autonomous behavior, environmental interaction, and user assistance	Rule-based, reactive, BDI-based, or interface-oriented behavior mechanisms	Provides the historical foundation for the agent concept, but is not the core object of this review
Multi-Agent Systems (MAS)	Communication, negotiation, cooperation, and distributed decision-making among multiple agents	Coordination among multiple autonomous entities through predefined roles, protocols, or interaction rules	Used as an important conceptual background for understanding collaboration, but differs from modern LLM-driven agentic workflows
Agent-Based Modeling (ABM)	Simulation of individual behaviors and emergent system-level phenomena	Agents are modeled as rule-based simulation units to represent heterogeneous actors or decision entities	Used mainly as a boundary concept because it focuses on simulation rather than foundation-model-driven task execution
General LLM-based agents	Open-ended task execution based on reasoning, planning, tool use, memory, and feedback	Large language models act as cognitive cores and interact with external tools or environments through iterative loops	Provides the main methodological basis for defining modern agricultural intelligent agents
Agricultural intelligent agents	Agricultural task organization, decision support, and controlled execution in smart agriculture	Foundation-model-driven systems that connect agricultural objectives, domain knowledge, dynamic data, tools, and feedback	Constitutes the core object of this review

2.2. Core Characteristics and Research Boundaries of Agricultural Intelligent Agents

Agricultural intelligent agents in this review are defined in a relatively restrictive sense. They refer to foundation-model-driven systems that can organize agricultural objectives, domain knowledge, dynamic data, external tools, and feedback into a task-oriented process. This definition uses a narrower boundary than the loose use of the term “agent” in some agricultural AI studies. Providing a conversational interface, incorporating a large language model, or retrieving agricultural documents before answer generation is not sufficient for a system to be classified as an agricultural intelligent agent. What matters is whether the system can advance an agricultural task through planning, tool invocation, workflow organization, feedback, verification, or collaborative decision-making.

This boundary is necessary because several adjacent paradigms are easily conflated in current agricultural AI research. Agricultural foundation models may provide domain knowledge understanding, text generation, or multimodal representation capabilities, but they do not automatically become agent systems if they lack explicit task planning, external resource use, or process-level interaction. Similarly, a RAG-based agricultural question-answering system can improve the evidential basis and domain specificity of responses, but it remains closer to a knowledge-enhanced advisory system if it only completes a “retrieval–generation” pipeline without forming an action chain toward a task objective [15,16]. Traditional agricultural expert systems, MAS, and ABM also differ from the modern agricultural intelligent agents discussed here, because their autonomy is usually grounded in predefined rules, interaction protocols, or simulation assumptions rather than foundation-model-driven reasoning and open-ended tool use.

Existing studies show that agricultural intelligent agents differ substantially in the depth of their agentic mechanisms. ShizishanGPT integrates large language models, search modules, agricultural knowledge graphs, RAG, and agricultural tools, and therefore represents a transition from ordinary agricultural question answering toward tool-augmented agricultural knowledge services [17]. CottonBot is more directly connected to field-level decision-making because it incorporates soil moisture sensors and weather forecast interfaces into an agentic workflow for cotton irrigation recommendation [18]. AgriWorld provides a stronger form of agenticity by allowing LLM agents to call geospatial queries, remote-sensing analysis, crop simulation, and code execution tools, and to revise reasoning on the basis of execution results [19]. These examples should not be treated as equivalent. Rather, they indicate a spectrum from tool-augmented advisory systems to verifiable task-execution systems.

Based on this boundary, agricultural intelligent agents generally involve four core characteristics. First, they are oriented toward explicit agricultural task objectives, such as advisory support, disease diagnosis, irrigation scheduling, phenotyping workflow construction, farm recordkeeping, or mission planning. Second, they use large language models, multimodal foundation models, or related foundation-model components as important cognitive, reasoning, or interaction cores. Third, they are able to access or coordinate external resources, including knowledge bases, domain-specific models, sensor streams, weather APIs, code execution environments, crop models, digital twins, farm databases, or equipment interfaces. Fourth, they possess some form of multi-step process organization, such as task decomposition, workflow orchestration, intermediate result interpretation, feedback-based refinement, or multi-agent collaboration.

It should be emphasized that these characteristics are not merely a checklist of technical components. The same component may indicate different levels of agenticity depending on how it is embedded in the system. For example, RAG can support ordinary knowledge retrieval, but it becomes more agent-relevant when retrieval is selectively triggered, compared with other information sources, and used to revise downstream reasoning. Likewise, external APIs or domain models do not necessarily make a system agentic unless the system can select, invoke, interpret, and adjust their use according to task requirements. Feedback and verification mechanisms represent a stronger level of agentic capability because they allow the system to reassess intermediate outputs rather than simply generate a final response.

Accordingly, the agricultural intelligent agents reviewed in the following sections mainly refer to systems that satisfy three minimum conditions: they are oriented toward specific agricultural tasks; they use foundation models as important reasoning, interaction, or orchestration components; and they connect knowledge, data, tools, workflows, feedback, or collaboration mechanisms in a way that supports task progression rather

than isolated response generation. This definition provides the basis for distinguishing agricultural intelligent agents from standalone agricultural foundation models, static RAG-based question-answering systems, conventional agricultural AI models, and traditional MAS/ABM approaches. It should be noted, however, that the boundary criteria in Table 2 define the inclusion scope of this review rather than the maturity level of individual systems. The evidence reported in the original studies and the deployment readiness of representative systems are therefore further examined after the application synthesis in Section 4.5.

Table 2. Research boundary and analytical criteria for agricultural intelligent agents in this review.

Analytical Criterion	Meaning in This Review	Boundary Implication
Explicit agricultural task objective	The system operates around a concrete agricultural problem rather than general-purpose dialogue.	Distinguishes agricultural intelligent agents from generic LLM applications.
Foundation-model-driven reasoning	Large language models or multimodal foundation models provide important reasoning, interaction, or orchestration support.	Distinguishes modern agricultural agents from traditional rule-based agricultural expert systems.
External resource or tool invocation	The system can access or call external knowledge sources, data streams, analytical models, APIs, or execution environments.	Distinguishes agentic systems from standalone agricultural foundation models or static prediction models.
Multi-step workflow organization	The system organizes agricultural tasks through sequential or iterative processing rather than one-step output generation.	Distinguishes agricultural intelligent agents from single-turn question-answering systems.
Feedback, verification, or constraint handling	The system uses intermediate results, external observations, human feedback, confidence assessment, or formal constraints to improve reliability.	Indicates a stronger transition from prototype assistance toward reliable task execution.
Decision support or controlled execution	The output is connected to agricultural management, resource allocation, farm operations, or controlled action planning.	Distinguishes task-chain-oriented agents from purely informational agricultural AI systems.

3. Basic Architectures and Key Capabilities of Agents

After clarifying the conceptual boundaries of agents and agricultural intelligent agents in Section 2, it is necessary to further discuss the internal structures and key mechanisms through which modern agents execute tasks. Unlike traditional single-turn generative models, modern large language model-driven agents are increasingly designed as composite systems consisting of multiple components, including goal setting, state and context perception, memory management, task planning, tool use, action execution, and feedback-based refinement. Existing studies generally suggest that memory, planning, and tool use constitute the core capability dimensions of LLM-based agents [3,4,20]. On this basis, this section reviews the key methodological foundations of modern agents from three perspectives: basic architectures, task planning and tool-based behavioral loops, and multi-agent collaboration and reliability evaluation.

3.1. Basic Architectures of Agents

Based on existing representative systems, the architectural design of modern LLM-based agents has gradually shifted from a single mode in which large language models directly respond to inputs toward composite systems supported by task objectives, contextual states, memory mechanisms, decision-making processes, and action interfaces. Generative Agents [21] enable agents to retain past experiences, derive higher-level reflections, and generate subsequent behaviors through a “memory stream–reflection–planning” mechanism. MemGPT [22] introduces an operating system-like hierarchical memory management approach, improving the ability of agents to maintain persistent states in long-range con-

versations and complex document-based tasks through main context, external storage, and control-flow scheduling. These studies indicate that the basic architecture of modern agents is not merely an external encapsulation of large language models, but an effort to construct stable internal states and behavioral organization mechanisms for continuous task execution.

The basic architecture of LLM-based agents can generally be summarized into five interrelated functional units. The first is the goal and task representation module, which receives user intentions, environmental objectives, or system constraints and transforms them into task descriptions for subsequent processing. The second unit concerns perception and context construction, drawing together user inputs, external observations, historical information, and tool-returned results as the current-state basis for reasoning. The third unit is memory management, which retains short-term task states, longer-term experiences, key intermediate results, and reusable knowledge. The fourth is the planning and decision-making module, which determines the next action based on the goal, context, and memory information, including whether to continue reasoning, whether to call tools, and when to terminate the task. The fifth is the action and feedback module, which executes text generation, API calls, code execution, or environmental interactions, and feeds the results back into the subsequent reasoning process. Although the implementation of these modules varies across systems, their common objective is to transform the generative capability of large language models into continuous task-execution capability for complex tasks [21].

Among these components, the memory mechanism is a key foundation for supporting long-term behavioral consistency and task continuity in agents. The memory mechanism in Generative Agents does not simply record historical interactions; instead, it transforms fragmented experiences into cognitive resources that can participate in subsequent decision-making through importance assessment, retrieval, and reflection mechanisms [21]. MemGPT [22] further demonstrates that when the task span exceeds the immediate context window of a model, agents need the ability to hierarchically store, selectively retrieve, and dynamically update information. Accordingly, memory in agents can be understood as a composite mechanism consisting of short-term working memory, long-term experiential memory, and task-state memory. Beyond information storage, memory helps maintain state stability and behavioral coherence across multi-turn tasks.

Beyond memory, modern agent architectures also show a growing emphasis on skill accumulation and experience reuse. Voyager [23] constructs an automatic curriculum, a skill library, and an iterative prompting mechanism in an open-world environment, enabling agents to consolidate successful behaviors into retrievable and composable programmatic skills and reuse previous experiences in new tasks. This mechanism suggests that agent architecture should not be limited to solving a single task, but should further support the gradual accumulation of capabilities through interaction. For vertical-domain applications, the integration of agricultural knowledge, retrieval modules, and domain-specific tools in ShizishanGPT [17], as well as the construction of external execution tools and code-execution environments in AgriWorld [19], can be regarded as domain-specific extensions of general agent architectures.

3.2. Task Planning, Tool Use, and Behavioral Loops

Task planning is a core capability through which agents transform high-level goals into executable processes. When facing complex tasks with multi-stage dependencies, agents need to identify the structure of the problem, decompose the overall objective into several actionable subtasks, and arrange the execution sequence accordingly. Plan-and-Solve Prompting [24] alleviates the problem of missing intermediate steps in multi-step reasoning through a “plan-first, solve-later” strategy. Tree of Thoughts [25] further

introduces branch search and path backtracking mechanisms, enabling models to compare and select among multiple candidate reasoning paths. Although these methods are not complete agent frameworks themselves, their ideas of task decomposition, path exploration, and intermediate decision-making provide an important methodological foundation for the planning modules of LLM-based agents.

In modern agents, task planning usually does not operate in isolation, but is closely coupled with tool use. ReAct organizes reasoning and acting in an interleaved manner, allowing models to call external resources and update subsequent judgments within a “thought–action–observation” loop [13]. Toolformer demonstrates that language models can learn when to call external APIs, how to pass parameters, and how to incorporate the returned results [14]. On this basis, Gorilla further addresses parameter-generation errors and tool hallucination in API calls, improving the accuracy and adaptability of tool use through retrieval-augmented mechanisms [26]. ToolLLM constructs tool-use datasets, training frameworks, and evaluation systems around large-scale real-world APIs, thereby promoting the study of tool use from isolated examples toward systematic capability evaluation [27]. In agricultural settings, such tool-use capability also depends on whether agent systems can interface with field environmental sensing and multi-sensor fusion resources, because real-time sensing provides the operational state information required for field decisions [28].

From the perspective of operational logic, tool use in agents is not a one-time “external query”, but a key component embedded in continuous behavioral loops. In this process, agents need not only to determine whether a tool should be called, but also to perform tool selection, parameter generation, interpretation of returned results, and subsequent strategy adjustment. The “reasoning–acting–observation” paradigm proposed by ReAct lays the foundation for this behavioral loop [13], while studies on Gorilla and ToolLLM further indicate that the core of tool-use capability is not merely whether a model can call tools, but whether it can call them in a stable, accurate, and adaptive manner [26,27]. Furthermore, whether behavioral loops can lead to continuous improvement depends on whether agents possess feedback-based refinement capabilities. Reflexion enables agents to summarize the causes of failure based on task feedback through linguistic reflection and episodic memory, and to incorporate these reflections into subsequent decision-making, thereby improving behavior without updating model parameters [29]. CRITIC introduces external tools into self-verification, allowing models to perform a “verification–critique–revision” loop around initial outputs and improve output quality in tasks such as question answering, code generation, and harmful content mitigation [30].

The overall architecture and behavioral loop of agricultural intelligent agents are illustrated in Figure 1.

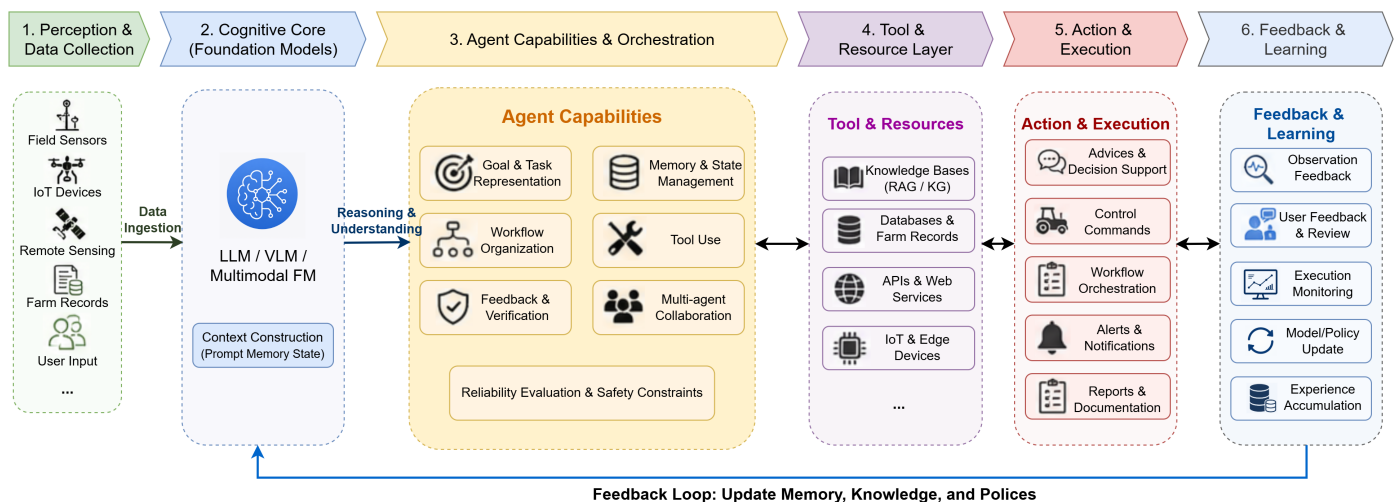


Figure 1. Overall architecture and behavioral loop of agricultural intelligent agents, showing the integration of cognitive cores, orchestration mechanisms, external tools, task outputs, and feedback loops.

3.3. Multi-Agent Collaboration and Reliability Evaluation

As task complexity increases, a single agent may be unable to simultaneously undertake multiple responsibilities, such as goal interpretation, solution design, context maintenance, result verification, and user interaction. Accordingly, multi-agent collaboration has gradually become an important direction in LLM-based agent research [31]. CAMEL [32] explores autonomous cooperation among multiple conversational agents through a role-playing mechanism. AutoGen [33] integrates multi-agent conversation, tool use, and human participation into a unified framework, providing a more composable infrastructure for constructing complex workflows. MetaGPT [34] further embeds standard operating procedures into multi-agent collaboration chains, reducing logical confusion and cascading errors in complex tasks through role division, intermediate artifact generation, and stage-wise verification.

From the perspective of collaboration objectives and organizational forms, existing studies on LLM-based multi-agent systems have developed several representative collaboration mechanisms. First, role division and task delegation assign complex tasks to different roles, such as planners, executors, evaluators, or domain experts, thereby improving the functional specialization of the system [32,35]. Workflow orchestration and stage-wise handoff distribute tasks among multiple agents through predefined or semi-structured workflows, with task transfer helping to preserve continuity between processing stages [33,36]. Peer evaluation, debate, and arbitration rely on mutual critique, output comparison, or final adjudication to identify potential errors [37,38]. In multi-agent debate, several model instances discuss candidate answers over multiple rounds, which has been shown to improve factuality and performance on complex reasoning tasks [37]. Rather than simply increasing the number of agents, these arrangements aim to improve complex task processing by structuring the exchange of responsibilities and intermediate results.

It should be noted that multi-agent systems do not inherently guarantee higher reliability. Without clear role boundaries, stable task-handoff mechanisms, and effective verification of intermediate results, multi-agent collaboration may lead to information redundancy, error propagation, and even ambiguity in responsibility. Existing studies further indicate that failures in multi-agent systems are often associated with insufficient system specification, mismatched coordination among agents, and incomplete verification and termination mechanisms [39]. In this sense, multi-agent collaboration is both a mechanism

for capability expansion and a new source of risk, and therefore needs to be co-designed with reliability evaluation systems.

Correspondingly, modern agent research is gradually developing specialized evaluation systems that differ from those used for traditional natural language processing tasks. AgentBench [40] constructs multiple interactive environments to evaluate agents in terms of long-horizon reasoning, decision-making, tool use, and instruction following, and shows that current models still exhibit clear limitations in the stability of long-range task execution. GAIA [41], oriented toward realistic assistant-like tasks, integrates reasoning, multimodal understanding, web browsing, and tool use into a unified evaluation setting, revealing that advanced models still show a significant gap from human-level performance on tasks closer to real-world applications. This suggests that agent evaluation cannot rely solely on static question-answering accuracy, but should further examine whether agents can actually complete tasks, use tools correctly, and maintain behavioral consistency in complex environments. Furthermore, reliability evaluation is expanding from assessing whether agents can complete tasks to assessing whether they can complete tasks stably and in compliance with external constraints. τ -bench [42] incorporates user interaction, tool use, and domain rules into evaluation, assessing agent reliability through the final database state and consistency across multiple runs. Experimental results show that even advanced function-calling agents still exhibit limited success rates and consistency in realistic domain-specific interactions. This indicates that agent reliability should be evaluated from at least four aspects: task completion, correctness of tool use, behavioral stability across multiple turns, and compliance with external rules or constraints.

Overall, multi-agent collaboration extends the organizational capability of agents for handling complex tasks, while reliability evaluation provides necessary constraints for their deployment in real-world applications. In agricultural scenarios, PhenoAssistant has already demonstrated the ability to coordinate multiple tools and orchestrate workflows [43], and AgroAskAI has also shown the collaborative potential of role-specialized agents in supporting smallholder farmers' enquiries [44]. However, their core methodological foundations still originate largely from general agent research on collaborative organization, process verification, and reliability evaluation. The capability dimensions summarized in Table 3 therefore provide not only a conceptual basis for understanding agricultural intelligent agents, but also the assessment concerns used later to examine representative systems. The following section further discusses how these capabilities are embodied in agricultural knowledge services, crop monitoring and diagnosis, resource management, and operational execution from the perspective of agricultural task chains.

Table 3. Key agent capabilities and assessment concerns for agricultural intelligent agents.

Agent Capability	Meaning in Agricultural Agents	Assessment Concerns
Goal and task representation	Interpreting farmer questions, research objectives, management goals, or mission requirements as agricultural task processes.	Goal decomposition, subtask ordering, task-state tracking, and premature termination.
Context construction and evidence grounding	Connecting user intent, crop type, region, growth stage, field condition, and external evidence into the current reasoning context.	Applicability of retrieved agronomic evidence, source credibility, and contextual mismatch.
Tool and resource use	Calling knowledge bases, APIs, databases, analytical models, code environments, sensors, crop models, or digital twins.	Tool selection, parameter validity, execution traces, and interpretation of returned results.
Memory and state maintenance	Preserving short-term task states, long-term user or farm context, reusable knowledge, and intermediate results.	Memory consistency, metadata continuity, retrieval consistency, and cross-stage information reuse.

Table 3. *Cont.*

Agent Capability	Meaning in Agricultural Agents	Assessment Concerns
Workflow reliability and verification	Organizing multi-step workflows and checking intermediate outputs under agronomic, operational, or safety constraints.	Contradiction, cascading errors, constraint satisfaction, repeated-run stability, uncertainty exposure, and verification.
Multi-agent coordination	Assigning subtasks to specialized agents and coordinating role interaction, information handoff, or result checking.	Handoff fidelity, deadlock prevention, correlated hallucination, and whether collaboration is compared with single-agent baselines.
Deployment and human oversight	Examining whether the system can operate under realistic agricultural, economic, and organizational conditions.	Latency, computational cost, human intervention frequency, scalability, and validation against agronomic or operational outcomes.

4. Task-Chain-Oriented Applications of Agricultural Intelligent Agents

The application value of agricultural intelligent agents should not be understood simply as the extension of large language models to agricultural question-answering scenarios. Their more distinctive contribution lies in their potential to reorganize agricultural tasks that were previously handled by separate models, databases, decision rules, or human experts. In real agricultural production and research, a practical problem is seldom completed by a single prediction or classification result. A disease-related inquiry may require symptom interpretation, image-based evidence, regional agronomic knowledge, and subsequent control suggestions. An irrigation decision may require soil-moisture observations, short-term weather forecasts, crop-stage information, and constraints on water use. A phenotyping task may involve data import, trait extraction, visualization, and model training. These examples indicate that agricultural intelligence is often process-oriented rather than output-oriented. It is in this process-oriented setting that agricultural intelligent agents begin to show system-level value beyond conventional agricultural AI models [45–47].

This task-chain perspective also helps clarify the relationship between agricultural intelligent agents and earlier agricultural modeling approaches. Crop system simulation models, represented by APSIM and DSSAT, have long provided important methodological foundations for yield prediction, management evaluation, climate-impact assessment, and agronomic scenario analysis by characterizing the relationships among weather, soil processes, crop growth, and management practices [48–51]. Their strength lies in mechanistic structure, agronomic interpretability, and scenario-based quantitative simulation. However, these models are not designed to interpret open-ended natural-language goals, select heterogeneous tools, or revise a task workflow through interaction. They provide important domain models and analytical tools for future agent systems, but they are not themselves open-ended task-execution agents.

Machine learning, deep learning, remote sensing, and agricultural Internet of Things technologies have further improved the automation of agricultural perception and prediction. They have been widely used in pest and disease recognition, crop classification, yield estimation, phenotypic analysis, precision monitoring, and data-driven agricultural management [52–57]. Nevertheless, many of these systems remain organized around relatively fixed input–output relationships. A model may identify a disease, estimate a yield value, detect a field object, or generate a local recommendation, but it usually does not determine how this output should be combined with user intent, regional constraints, external tools, or subsequent management actions. In this respect, conventional agricultural AI has substantially improved local automation, but it often lacks the cross-task organization and feedback mechanisms required for complete agricultural task chains. This comparison also indicates that agricultural intelligent agents should not be viewed as replacements for

existing agricultural models, decision-support systems, or automation technologies. Crop models, rule-based decision-support systems, machine-learning models, remote-sensing pipelines, and IoT-based control systems remain more suitable for tasks with stable inputs, well-defined rules, limited interaction requirements, and strict requirements for computational efficiency or interpretability. By contrast, agricultural intelligent agents are more appropriate when the task involves open-ended user goals, incomplete contextual information, heterogeneous tools, multi-step reasoning, or the need to coordinate several analytical components. Their practical value therefore lies less in replacing alternative approaches than in acting as an orchestration layer that connects models, data, tools, and human decisions within agricultural task chains.

Agricultural intelligent agents address this limitation only when they provide an organizational layer above individual models, knowledge resources, data streams, and decision-support tools. Their value should therefore be judged not merely by whether they use advanced models, but by whether they can structure agricultural problems into interpretable and controllable workflows. Some existing systems remain close to knowledge-enhanced advisory services; others have begun to connect dynamic data, external APIs, code execution, digital twins, or formal verification. This uneven development suggests that agricultural intelligent agents should be examined as a continuum. At one end are systems that improve knowledge access and contextual response generation; at the other end are systems that participate in verifiable reasoning, decision support, and controlled planning. Between these two ends are workflow-oriented systems that coordinate domain knowledge, multimodal evidence, and tool-based analysis.

Following this logic, this section organizes representative studies according to their positions in agricultural task chains rather than by technology labels alone. Four application directions are considered: agricultural knowledge services and advisory support; crop-state understanding, including monitoring, diagnosis, and phenotyping; resource management and decision support; and agricultural workflow orchestration and farm-level collaborative management. These categories are not intended to form rigid boundaries. Rather, they describe an increasing depth of agent involvement, from interpreting agricultural problems and retrieving applicable knowledge to organizing multimodal evidence, generating management schemes, coordinating tools, and supporting controlled execution.

The task-chain-oriented application framework described above is illustrated in Figure 2.

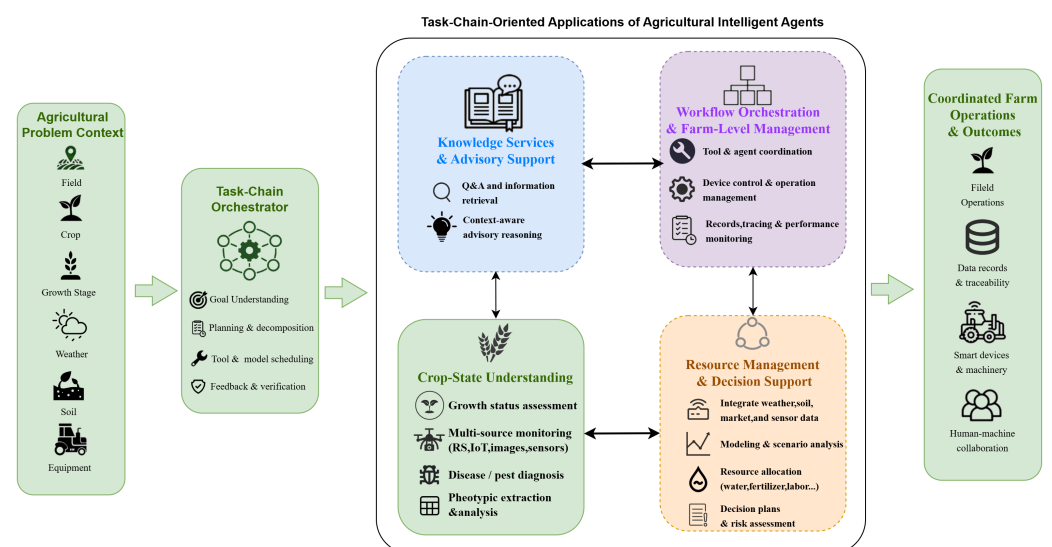


Figure 2. Overview of the task-chain-oriented application framework for agricultural intelligent agents.

4.1. Agricultural Knowledge Services, Question Answering, and Advisory Support

Agricultural knowledge services constitute the most immediate entry point through which intelligent agents participate in agricultural task chains. At the same time, this is also the application direction in which the boundary between ordinary LLM-based question answering and genuine agentic capability is most easily blurred. A system may retrieve agricultural documents, answer farmers' questions, or generate management suggestions, but these functions alone do not necessarily make it an agricultural intelligent agent. From the perspective of this review, the key issue is whether the system merely responds to an isolated query or constructs a context-aware problem representation that can support subsequent diagnosis, recommendation, or decision-making.

The difficulty of agricultural consultation lies in the strong contextuality of agricultural knowledge. Fertilization plans, crop-protection measures, sowing-date recommendations, and irrigation suggestions are rarely universally applicable. Their validity depends on crop species, growth stage, regional climate, soil condition, production mode, and management objective. A response may be factually plausible but still agronomically inappropriate if it is detached from its local production context. Therefore, agricultural knowledge-service agents should not be evaluated only by fluency, factual coverage, or retrieval relevance. A more important criterion is whether they can mediate among user intent, domain knowledge, regional context, and external evidence.

The agentic transformation of agricultural knowledge services is first reflected in the shift from relying on the parametric knowledge of large language models to coordinating multiple knowledge sources and information channels. Agricultural knowledge is distributed across structured knowledge bases, extension documents, local experience, real-time environmental information, and outputs from domain-specific analytical models. Knowledge graphs can represent agricultural entities, attributes, and relationships in a structured form, thereby supporting semantic retrieval and knowledge association [58]. RAG can further improve the evidential basis of generated responses by introducing external retrieval before response generation [16,59]. However, retrieval itself is not sufficient. It becomes more relevant to agentic agricultural systems only when it is embedded in a broader process of problem interpretation, evidence selection, response synthesis, and possible downstream decision support.

From this perspective, tool-augmented agricultural question-answering systems represent an intermediate stage between ordinary knowledge services and stronger agricultural agents. A system that integrates search engines, agricultural knowledge graphs, RAG modules, and domain-specific agricultural tools can move beyond a single-turn question-answer pattern by forming a lightweight workflow for interpreting questions, selecting resources, supplementing evidence, and synthesizing responses [17]. Its contribution should be understood not simply as improving answer quality, but as introducing an organizational mechanism that connects agricultural questions with external knowledge and tools. Nevertheless, such systems remain closer to knowledge-service agents unless their outputs are further connected to task planning, feedback-based revision, or verifiable decision processes.

Whether technical capability can be translated into usable advisory support is another consideration. In practice, agricultural consultation rarely begins with a complete and well-structured task description. Farmers may use local expressions to describe symptoms, omit relevant contextual information, or combine production experience with an immediate management concern. Advisory-oriented agents therefore need to infer the purpose of a query, identify information that is still missing, and adapt their responses to language, service accessibility, and trust-related requirements. Large-scale advisory services for small-holder farmers illustrate that agricultural knowledge services depend not only on model

reasoning or retrieval accuracy, but also on interaction design, localization, and practical usability [15].

Furthermore, the reliability of agricultural knowledge services depends not only on factual correctness, but also on contextual appropriateness. Agricultural recommendations are subject to strong regional constraints. A sowing-date arrangement, fertilizer application scheme, or pest management strategy that is appropriate in one region may be unsuitable or even misleading when transferred to another agroecological context. To address this issue, AgriRegion proposes region-aware retrieval, which improves the regional consistency and applicability of agricultural advice by incorporating geographic metadata, region-prioritized retrieval, and constraints from local agricultural extension knowledge [59]. Agricultural knowledge service agents therefore need to move from retrieving relevant knowledge to retrieving contextually applicable knowledge. If the front-end knowledge service layer suffers from regional mismatch, subsequent diagnosis and decision-making, regardless of their technical complexity, may still be grounded in an inappropriate production context.

Beyond knowledge organization, user adaptation, and regional constraints, agricultural knowledge services are also evolving from information response toward pre-decision support. For climate adaptation, disaster risk management, and complex agronomic decision-making, farmers often require not isolated pieces of information, but integrated judgments involving risk identification, comparison of management options, and feasible action pathways. AgroAskAI employs role-specialized agents, a chain-of-responsibility coordination mechanism, real-time tool access, and internal feedback to extend smallholder advisory services into a multi-step reasoning and contextual recommendation process [44]. The role of agricultural knowledge service agents is gradually shifting from information provision to preliminary reasoning, providing structured premises for downstream decision-making within agricultural task chains.

Overall, agricultural knowledge service and question-answering applications show a clear progression from knowledge accessibility to evidence grounding, contextual adaptation, and preliminary decision support. ShizishanGPT, Farmer.Chat, AgriRegion, and AgroAskAI illustrate this evolution from different perspectives: they respectively emphasize tool-augmented knowledge organization, low-barrier advisory interaction, region-aware retrieval, and multi-agent reasoning for smallholder support. In this sense, agricultural knowledge services should not be viewed as a low-level auxiliary component of agricultural intelligent agents, but as a foundational layer that affects the reliability of subsequent state analysis, resource allocation, and field-level execution.

Nevertheless, several limitations remain in this direction. First, most existing systems still focus on single-turn consultation or limited multi-turn interaction, while long-term memory, continuous user modeling, and cross-stage task tracking remain underdeveloped. Second, the evaluation of agricultural knowledge services largely relies on answer quality, retrieval relevance, or user satisfaction, and has not yet established sufficient closed-loop validation with real agricultural production outcomes. Third, the structured organization of region-specific knowledge, local agronomic experience, and minor-crop scenarios remains insufficient, limiting the adaptability of agents in non-mainstream crops and complex local production systems. Future agricultural knowledge service agents should therefore develop traceable, updatable, and context-aware agricultural knowledge foundations to support the stable progression of downstream agricultural task chains.

4.2. Crop-State Understanding: Monitoring, Diagnosis, and Phenotyping

Crop monitoring, diagnosis, and phenotyping provide information that is later used in agricultural diagnosis and management, rather than serving as final outputs by them-

selves. Machine learning, deep learning, and computer vision have substantially improved the extraction of crop-state information from images, UAV data, and sensor observations [52–54,60–62]. UAV imagery and robot-oriented visual detection have been applied to weed detection, fruit target localization, and other crop-state perception tasks, providing visual evidence for subsequent field management or robotic operations [63,64]. Even so, most perception systems are designed for fixed input–output tasks. They may detect objects, quantify traits, or estimate yield-related indicators [65], but their outputs still need to be interpreted with reference to agronomic conditions, user needs, and later management decisions. The role of an agent is therefore not to replace perception models, but to place their outputs within a diagnostic or analytical process that can draw on other relevant evidence.

Plant phenotyping illustrates why this broader workflow perspective is needed. Visual sensors and three-dimensional reconstruction have already been used to measure plant canopy structure and extract structural traits [66]. In practice, phenotyping rarely ends with a single image-processing operation, since data import, image analysis, feature extraction, result visualization, model training, and experimental interpretation are often connected within the same analytical workflow. PhenoAssistant uses large language models to orchestrate tools for phenotypic extraction, data visualization, and model training, enabling researchers to drive complex analytical workflows through natural language interaction [43]. Rather than replacing a particular phenotyping algorithm, this system transforms workflows that previously required substantial computational expertise into interactive, composable, and repeatable task processes. In this sense, the role of agricultural intelligent agents in phenotyping is closer to that of a research workflow coordinator. They connect user intent, analytical tools, and intermediate outputs, thereby lowering the operational barrier of complex agricultural research tasks.

In crop disease diagnosis, the key agentic transformation is to extend the diagnostic process from disease-name recognition to evidence organization, coordinated reasoning, and explainable judgment generation. Conventional disease recognition models mainly answer which disease may be present in an image, whereas real agricultural diagnosis also requires symptom interpretation, assessment of evidence sufficiency, task-type differentiation, and linkage to subsequent management actions. Recent studies have begun to reconstruct this process through multi-agent collaboration and multimodal reasoning. Chat Demeter organizes plant disease diagnosis into a collaborative workflow involving task planning, reasoning evaluation, and result presentation [67]. AgriDoctor further integrates modules such as a router, classifier, detector, knowledge retriever, and LLM, allowing the system to select different processing pathways according to input forms and user intent, and to incorporate image understanding, knowledge retrieval, and language interaction into a unified framework [68]. These studies collectively indicate that agricultural diagnostic systems are shifting from a model that produces a classification result to an agentic system that organizes analytical pathways around diagnostic tasks.

From the perspective of agricultural task chains, an important change is that these systems explicitly organize diagnostic objects, evidence sources, and processing procedures. A user query can first be interpreted as an image recognition task, a knowledge consultation task, or a multimodal diagnostic task, after which the relevant modules can be selected and domain knowledge added where needed. This is more suitable than traditional end-to-end classification models when agricultural inputs are not standardized, for example, when images are blurred, descriptions are expressed in natural language, or contextual information is incomplete. In such cases, agricultural intelligent agents do more than strengthen recognition. They coordinate the diagnostic workflow by deciding which capabilities to use, what evidence remains missing, and how the diagnostic judgment should be reached.

As diagnostic chains become more complex, reliability becomes increasingly important. If agricultural perception systems only pursue final conclusions without identifying uncertainty or rechecking results, they may cause inappropriate pesticide application, delayed intervention, or increased production costs in real-world use. Agentic AI with Orchestrator-Agent Trust separates visual perception agents from a non-visual reasoning orchestrator, and combines confidence calibration, image-RAG, and iterative reassessment to re-evaluate visual classification results [69]. This work is important because it brings trust assessment into the agent's internal logic, allowing the system to consider when perception results are reliable and when further verification is needed. Accordingly, agricultural diagnostic agents are beginning to move from being able to recognize crop problems toward being able to judge whether their own recognition results are sufficiently reliable.

This shift also expands the evaluation criteria for agricultural monitoring agents. Traditionally, disease recognition or phenotyping systems have often been evaluated mainly by accuracy, recall, or error metrics. Under an agentic framework, however, whether a system can explain its evidence, identify uncertainty, and trigger re-retrieval or reassessment when necessary should also become important evaluation dimensions. This is particularly important in agricultural production environments, where image quality is unstable, anomalous samples are scarce, and long-tail diseases are difficult to cover. Static test sets alone are often insufficient to reflect real deployment capability. To address this issue, Enhancing the Evaluation of Fault Detection Models in Smart Agriculture Using LLM Agents for Rule-Based Anomaly Generation explores the use of LLM agents to generate rule-driven anomaly samples, thereby improving the test coverage of fault detection models in smart agriculture [70]. Agricultural intelligent agents are beginning to support both diagnosis itself and the testing and validation of agricultural monitoring systems.

Overall, agricultural intelligent agents for crop monitoring, diagnosis, and phenotyping are evolving along three interconnected pathways. First, agents are beginning to act as organizers of analytical workflows, enabling complex phenotyping and research tasks to be structured and executed with lower operational barriers. Second, systems are forming multimodal diagnostic chains that integrate images, text, domain knowledge, and user intent into crop-state judgment. Third, increasing attention is being paid to trustworthy perception and evaluation mechanisms, including confidence calibration, re-verification, and anomaly-based testing. PhenoAssistant, Chat Demeter, AgriDoctor, Orchestrator-Agent Trust, and the anomaly generation evaluation framework reflect this development from different perspectives [43,67–70]. Therefore, this research direction should not be understood as a simple combination of agricultural vision models and large language models, but rather as the evolution of agricultural perception systems toward a continuous task chain of state understanding, diagnostic organization, and trustworthy verification.

At present, this direction still faces two major limitations. First, many systems are still mainly validated using static images or offline datasets, and remain distant from continuous monitoring, temporal state tracking, and long-term disease progression analysis in real field environments. Second, the evaluation systems for monitoring- and diagnosis-oriented agricultural intelligent agents remain fragmented. Evaluation remains fragmented: some studies report recognition accuracy, whereas others focus on tool selection, retrieval performance, or confidence calibration. As a result, good performance in one component does not necessarily show whether the full diagnostic process is dependable. A common framework that follows the chain from input interpretation to diagnostic output is still missing. Further work should link multimodal perception with long-term state memory and diagnostic reliability, while examining whether these components improve subsequent management decisions rather than only problem detection.

4.3. Resource Management and Decision Support

Resource management and decision support are the points at which agricultural intelligent agents are most directly exposed to production consequences. Unlike knowledge services, which mainly organize information, or crop-state understanding, which mainly interprets observed conditions, resource-management agents must connect evidence with decisions on irrigation, fertilization, crop selection, pest intervention, and input allocation. These decisions are rarely determined by a single variable. They involve trade-offs among crop demand, soil status, weather variability, sensor observations, production objectives, economic cost, and environmental constraints. For this reason, the value of an agricultural agent in resource management should be assessed by how it constructs a decision process under constraints, not by whether it can generate a plausible agronomic recommendation in natural language.

Irrigation management provides a useful example of this distinction. Smart irrigation research has long relied on sensor observations, crop water-demand estimation, and scheduling methods to support water management decisions [71]. A language model may explain cotton water demand in general terms, but this is not sufficient for deciding whether a particular field should be irrigated under current soil and weather conditions. The significance of LLM-RAG and agentic-tool integration in cotton irrigation is that soil moisture data and weather forecast interfaces are brought into the same advisory workflow, allowing the recommendation to be grounded in field-level and time-sensitive evidence [18]. Earlier reinforcement-learning-based irrigation systems had already shown the value of sequential decision strategies in cotton water management [72,73]; the agentic contribution is different. It lies in connecting natural-language interaction, retrieval, environmental data access, and recommendation generation into a more interpretable decision-support process. Thus, irrigation-oriented agents should be understood as a bridge between agronomic knowledge services and data-conditioned field management, rather than as chatbots for irrigation advice.

For resource-management-oriented agents, domain knowledge must be combined with task-specific integration of multi-source information under explicit constraints. Agricultural resource allocation commonly involves multi-variable coupling, and a single data source is rarely sufficient for comprehensive decision-making. For example, fertilizer recommendation depends not only on soil properties, but also on crop type, climatic conditions, and expected yield targets. Crop selection is influenced by soil and weather conditions, as well as regional suitability and production objectives. Soil-state analysis can also involve spatially explicit anomaly detection; for example, machine learning combined with spatial statistics has been used to identify abnormal patterns in soil total nitrogen, providing a more detailed evidence layer for nutrient-related decision support [74]. The agentic AI-driven autonomous decision-support system for smart agriculture organizes soil image recognition, soil parameter analysis, crop recommendation, and fertilizer suggestion into a continuous decision-making workflow [75]. Such work goes beyond stacking multiple models together. Rather, it attempts to integrate originally separate agricultural analysis modules into a decision chain with sequential dependencies: environmental state recognition provides inputs for subsequent recommendations, and recommendation results further support resource input planning and management scheme formulation. In this view, resource management is no longer merely the output of a single prediction model, but is reconceptualized as scheme generation based on multi-stage evidence integration.

On this basis, agricultural intelligent agents further highlight the need for executable analysis and verifiable reasoning. Agricultural resource decisions often involve structured data, spatial information, remote-sensing time series, and crop simulation outputs [76]. If such decisions rely entirely on language-model reasoning in textual space, they may easily

produce pseudo-precise conclusions without sufficient analytical grounding. AgriWorld addresses this issue through a World–Tools–Protocol framework, in which geospatial queries, remote-sensing analysis, crop simulation, and task predictors are encapsulated as callable tools, while an LLM agent performs an “execute–observe–revise” reasoning loop within a code execution environment [19]. In this setting, resource decisions are not based only on linguistic explanation. For yield prediction, anomaly analysis, or counterfactual reasoning, the agent can obtain intermediate evidence through the relevant tools and adjust the analysis when the returned results require further examination. The resulting judgment is therefore linked to the computational steps and execution evidence used during the task, rather than resting on generated text alone.

From a broader precision agriculture perspective, resource-management-oriented agents also have the potential to shift agricultural management from local optimization toward multi-objective coordination. Modern agricultural resource management focuses not only on yield improvement, but also on water saving, fertilizer reduction, environmental load control, and sustainable production [77]. Agentic AI for smart and sustainable precision agriculture connects agentic AI with intelligent sensing, distributed collaboration, and continuous optimization, highlighting its potential in precision resource allocation and sustainable agricultural management [78]. Agentic AI with Large Language Models for Precision Farming: Advancing Sustainable Resource Optimization in Smart Agritech further discusses the role of LLM-driven agents in improving resource-use efficiency from the perspectives of irrigation, fertilization, and agronomic input optimization [79]. These studies indicate that the task boundary of agricultural intelligent agents in resource management has expanded from producing one-time management advice to coordinating production efficiency, input intensity, and environmental constraints around long-term objectives. This shift makes agricultural intelligent agents closer to the integrated decision-making problems faced by real farm managers than traditional single-task prediction systems.

However, agricultural resource decision-making cannot focus only on recommendation generation; it must also address decision safety and outcome constraints. When agent-generated schemes involve fertilizer reduction, irrigation adjustment, or changes in crop protection intensity, the system needs to further determine whether the proposed scheme satisfies agronomic constraints, whether it may introduce new risks, and whether it remains valid under expected environmental conditions. Digital twins have increasingly been regarded as important infrastructures for representing agricultural production processes, integrating virtual–physical information, and supporting simulation-based decision-making [80]. LLM-Orchestrated Digital Twins for Safe, Human-Centered Decision Support in Precision Agriculture directly addresses this issue. In this framework, a tool-using LLM agent operates on top of crop–soil dynamic twins and farm-operation twins. Farmers express high-level goals in natural language, after which the system generates candidate management schemes and verifies their feasibility and safety through digital twins before recommendations are presented or instructions are issued [81]. This framework avoids treating agricultural intelligent agents as isolated decision-makers that directly output an optimal answer. Instead, it embeds them into a controlled decision-support system composed of models, constraints, tools, and verification mechanisms. For high-risk agricultural management tasks, such a “generation–simulation–verification” mechanism is closer to the form required for deployable systems than merely improving the reasoning ability of language models.

Resource-management and decision-support agents are therefore better understood as organizers of data, models, tools, and constraints rather than as language-based wrappers for conventional decision-support systems. Their contributions can be separated into three levels. At the first level, they connect recommendations with dynamic production contexts

by incorporating sensor data, weather information, and field-level observations [18]. At the second level, they coordinate multiple analytical components, such as soil analysis, crop recommendation, fertilizer suggestion, remote-sensing tools, crop simulation, and code execution, so that recommendations are formed through an inspectable chain of evidence rather than through text generation alone [19,75]. At the third level, they introduce simulation or verification mechanisms, such as digital twins, to examine whether candidate management schemes satisfy agronomic, operational, and safety constraints before being presented or executed [80,81]. These levels differ substantially in deployment readiness and should not be treated as equivalent examples of resource-management agents.

At present, this direction remains closer to constrained decision-support prototyping than to stable farm-scale deployment. Many systems are still evaluated in limited scenarios or task-specific settings, and their robustness across crops, regions, seasons, and production systems remains insufficiently demonstrated. Tool and data dependencies also create practical difficulties: sensors, APIs, crop models, remote-sensing services, and farm databases often differ in resolution, reliability, and interface design. More importantly, the quality of a resource-management decision may only become visible after a production cycle, whereas many current evaluations focus on short-term recommendation quality or local task performance. Future resource-management agents should therefore strengthen dynamic data access, coupling with agronomic models, constraint-aware reasoning, and long-term outcome evaluation. Without these elements, such systems may remain useful advisory prototypes but fall short of dependable agricultural decision-support systems.

4.4. Agricultural Workflow Orchestration and Farm-Level Collaborative Management

Agricultural workflow orchestration and farm-level collaborative management represent the stage at which agricultural intelligent agents move closest to operational systems. The preceding sections mainly concern problem interpretation, state understanding, and recommendation generation. In farm-level settings, however, the central issue is not only what decision should be recommended, but how that decision is translated into coordinated workflows involving field conditions, equipment interfaces, operation records, human managers, and safety constraints. Recent progress in agricultural equipment has already embedded machine vision, intelligent sensing, decision algorithms, and control technologies into machinery and production systems [82]. The agentic question is different: whether these heterogeneous components can be organized around agricultural objectives in a traceable and controllable manner. This section therefore does not treat agricultural agents as another form of machinery automation [83], but as a potential orchestration layer for connecting task goals, farm data, device resources, human oversight, and controlled execution.

At the level of agricultural operations, agents first demonstrate their value in coordinating multi-source states and interconnected farming processes. Traditional agricultural automation systems, including IoT-based monitoring and control systems in controlled-environment agriculture, are often designed around a single device, action, or monitoring object [84,85], whereas farming operations usually require coordinated consideration of soil conditions, weather, crop status, visual observations, and management objectives. The Agentic AI Framework to Automate Traditional Farming for Smart Agriculture organizes functional units such as soil sensing, meteorological analysis, visual perception, and supervisory coordination in a multi-agent manner, aiming to construct a more interconnected automation framework for traditional farming processes [86]. This line of research reframes agricultural operations as workflow problems involving multi-state perception, multi-stage linkage, and multi-actor coordination, rather than as the simple addition of task modules. For agricultural intelligent agents, operational execution begins not with

the immediate triggering of physical actions, but with the unified organization of task objectives, environmental states, and executable conditions.

Farm-level collaborative management also depends on the continuous accumulation of accurate, reusable, and context-rich information. Agricultural production is highly temporal and process-dependent. Sowing, fertilization, spraying, or field inspection can support subsequent traceability analysis, farm management, and intelligent decision-making only when properly recorded and supplemented with contextual metadata. Meta Ag 2.0 introduces a retrieval-augmented agentic-AI framework into agricultural recordkeeping and metadata completion. By interacting with a large language model and a private database, the system dynamically generates follow-up questions based on existing spatiotemporal information, validates user inputs, and improves the completeness of operational records [87]. This direction extends agricultural intelligent agents from short-term interactive tools to organizers of long-term farm information flows. If agricultural activities can be transformed by agents into structured, traceable, and reusable data assets as they occur, subsequent disease traceback, input analysis, operation evaluation, and farm optimization can be built on a more reliable informational foundation.

Closely related to record governance is the issue of interoperability among heterogeneous field devices and data systems. Sensors, agricultural machinery terminals, meteorological platforms, imaging devices, and farm management information systems in smart agriculture often adopt different data formats and interface protocols, which makes it difficult for information to flow smoothly across systems. A study on leveraging agent-based experts on large language models explores data sharing among heterogeneous agricultural IoT devices [88]. In practical terms, farm-level collaborative management depends on both the reasoning capability of a single agent and its ability to establish semantic connections among devices, protocols, and data sources. In other words, if agricultural intelligent agents are expected to assume farm-level coordination responsibilities, they must be able to connect equipment, organize data, and promote cross-system collaboration. Otherwise, so-called intelligent execution may remain limited to local automation rather than forming integrated farm workflow coordination.

When agents move closer to embodied operations and robotic execution, task planning, execution mapping, and safety verification become central issues. Agricultural robots and intelligent agricultural machinery have been increasingly studied for field monitoring, spraying, harvesting, automatic steering, and other operation-oriented tasks, providing an important embodied foundation for agent-based mission planning in agriculture [89,90]. Robotic scouting, sampling, spraying, and multi-equipment coordination in precision agriculture require systems not only to understand natural-language goals, but also to translate them into task plans that satisfy spatial, sequential, resource, and safety constraints. Against this background, *As You Wish: Mission Planning with Formal Verification using LLMs in Precision Agriculture* emphasizes that mission plans generated by LLM agents should be checked using formal methods before deployment, so as to avoid infeasible or unsafe plans in real operational environments [91]. This perspective is particularly important because agricultural operations have direct real-world consequences: planning errors may lead to repeated operations, resource waste, equipment conflicts, or production risks. For this reason, embodied agricultural agents should not pursue unconstrained direct execution, but should instead establish a closed-loop structure of goal understanding, plan generation, constraint verification, and controlled execution.

The application scope of agricultural operation execution and collaborative management should also not be limited to crop production. Dairy science and broader livestock production also involve continuous tasks such as knowledge retrieval, record management, database interaction, operational analysis, and daily decision-making. Agents are all

you need: Pioneering the use of agentic artificial intelligence to embrace large language models into dairy science extends the discussion of agentic AI into the context of dairy science [92]. Related studies further suggest that, under constraints related to on-farm computing resources and data privacy, locally deployable small language models and multi-task agent frameworks can support literature retrieval, web search, database interaction, and visualization of predictive results in agricultural decision-support tasks [46]. From this perspective, farm-level collaborative management involves not only the scheduling of field equipment, but also the organization of knowledge, data, and management activities across operational processes. At this level, the value of agricultural intelligent agents lies in helping production entities integrate fragmented information resources and management tasks into unified workflows.

At present, this direction remains at an early exploratory stage. Many studies are still dominated by framework design, prototype validation, or demonstrations in limited scenarios. This is understandable, because execution-level agents face requirements that are substantially higher than those of question-answering or advisory systems. They must handle equipment interfaces, task specifications, exception states, human authorization, data privacy, and operational safety at the same time. The risk is that an agent may appear capable at the level of workflow description while remaining unable to operate reliably across devices, farms, and production cycles. Future work should therefore give priority to standardized interfaces, auditable task states, long-term memory for farm operations, human-agent authority allocation, formal constraints, and rollback mechanisms. Only with these safeguards can agricultural agents move from conceptual workflow orchestration toward production-ready farm management systems.

As shown in Table 4, the four application directions differ not only in agricultural scenario, but also in the depth to which agents intervene in the task chain. Knowledge-service systems are mostly located at the front end of the chain, where the main concern is whether agricultural advice can be grounded in appropriate evidence, user context, and regional constraints. Monitoring and diagnosis-oriented systems move one step further by treating visual or multimodal outputs as diagnostic evidence rather than as final results. Resource-management agents are held to a stricter standard because their recommendations need to be checked against dynamic field data and analytical results; some systems also incorporate simulation or verification. Farm-level collaborative systems are more demanding still, since they involve operational records, device interoperability, human-machine authority allocation, and safety-sensitive planning. The number of agricultural scenarios covered is therefore a poor proxy for system maturity. A more meaningful criterion is whether the system deepens task-chain control, strengthens external-resource coupling, and makes the decision or execution process more traceable and constrained. At the same time, this comparison remains partly qualitative unless the evidence supporting each capability is made explicit. A further assessment is therefore needed to distinguish demonstrated capabilities, partially supported claims, and dimensions that have not been evaluated.

Table 4. Synthesis of application directions of agricultural intelligent agents.

Application Direction	Agentic Contribution	Key Limitations
Agricultural knowledge services and advisory support	Organizes user intent, retrieved knowledge, regional context, and domain information into evidence-grounded advisory processes.	Limited long-term user modeling, production-level validation, and coverage of local knowledge or minor-crop scenarios.
Crop-state understanding: monitoring, diagnosis, and phenotyping	Connects visual or multimodal perception outputs with diagnostic reasoning, phenotypic analysis, workflow coordination, and reliability assessment.	Limited continuous monitoring, temporal state tracking, uncertainty handling, and full-process diagnostic evaluation.
Resource management and decision support	Integrates dynamic field data, analytical tools, crop models, code execution, or digital twins for contextualized management decisions.	Limited cross-region robustness, tool-chain validation, long-term agronomic evaluation, and safety-aware recommendation testing.
Agricultural workflow orchestration and farm-level collaborative management	Extends agents from information support to record governance, device interoperability, mission planning, and controlled workflow organization.	Limited evidence on device interoperability, human–machine authority allocation, privacy protection, deployment feasibility, and operational safety.

4.5. Comparative Assessment of Agentic Capabilities and Deployment Readiness

The preceding subsections show that agricultural intelligent agents have been explored in advisory services, crop-state interpretation, resource management, and farm-level workflow organization. These applications differ not only in agricultural scenario, but also in the extent to which they demonstrate agentic capabilities. Some systems mainly improve knowledge access and response grounding, whereas others organize perception results, invoke external tools, coordinate multiple agents, or verify candidate decisions through execution, simulation, or formal constraints. Therefore, the application synthesis above should be complemented by a more explicit comparison of what each representative system actually demonstrates.

This comparison is based on evidence reported in the original studies rather than on new benchmark experiments conducted in this review. The reviewed systems were developed for different crops, tasks, data sources, tools, and deployment assumptions. In addition, many original studies do not consistently report quantitative measures such as tool-selection validity rates, parameter-validity rates, repeated-run stability, human intervention frequency, latency, computational cost, or validation against agronomic and operational outcomes. Therefore, the purpose of the following matrix is not to rank systems by a universal performance score, but to clarify the evidence status of key agentic capabilities across representative studies.

To avoid unsupported numerical scoring, this review uses a qualitative evidence-status coding scheme. A dimension is marked as “D” when the original study explicitly reports the corresponding system component together with evaluation evidence, execution traces, validation procedures, or empirical results. A dimension is marked as “P” when the study describes the capability at the design or workflow level, or provides indirect evidence, but does not separately evaluate it. A dimension is marked as “NE” when the study does not report sufficient evidence for assessment. A dimension is marked as “NA” when it is not applicable to the system type. In particular, NE should not be interpreted as poor performance; it only indicates that the published study does not provide sufficient evidence for that dimension. The coding was assigned by checking whether the corresponding capability or validation evidence was explicitly reported in the system descriptions, experiments, evaluations, or case demonstrations of the reviewed studies.

Table 5 shows that systems with similar application labels may differ substantially in the type and strength of evidence they provide. Knowledge-service systems often report evidence grounding or regional adaptation, but they usually provide limited evidence on

tool-parameter validity, memory consistency, workflow stability, or practical validation. Resource-management and digital-twin-oriented systems more often report tool use, execution results, simulation, or constraint checking, but most still lack long-term field validation and cost-related evidence. These gaps should not be interpreted as confirmed weaknesses of individual systems. Rather, they show that current studies still report agent capabilities in uneven ways.

Table 5. Comparative capability profile of representative agricultural intelligent agents.

Representative System	GT	EG	TV	MS	WR	MA	PV
ShizishanGPT [17]	P	D	P	NE	NE	NA	NE
Farmer.Chat [15]	P	P	NE	NE	NE	NA	P
AgriRegion [59]	NE	D	NE	NA	P	NA	NE
AgroAskAI [44]	D	D	P	P	P	D	P
PhenoAssistant [43]	D	P	D	P	P	D	P
Chat Demeter [67]	P	P	P	NE	P	D	NE
AgriDoctor [68]	P	D	P	NE	P	NA	NE
Orchestrator-Agent Trust [69]	P	P	P	NE	D	P	NE
CottonBot [18]	P	D	D	NE	P	NA	P
Agentic AI-driven DSS [75]	D	P	P	NE	NE	NA	NE
AgriWorld [19]	D	D	D	NE	D	NA	P
LLM-orchestrated digital twins [81]	D	D	D	NE	D	NA	P
Meta Ag 2.0 [87]	P	D	D	D	P	NA	P
As You Wish [91]	D	P	D	NE	D	NA	P

Notes: GT = goal and task representation; EG = context construction and agronomic evidence grounding; TV = tool and parameter validity; MS = memory and state consistency; WR = workflow reliability and constraint satisfaction; MA = multi-agent coordination; PV = practical validation, including field or user validation, deployment-related evidence, latency or cost reporting, human intervention, and validation against agronomic or operational outcomes. D = demonstrated with reported evidence; P = partially demonstrated or indirectly supported; NE = not evaluated or not reported; NA = not applicable. NE does not indicate poor performance. DSS = decision-support system.

The MA column should be interpreted cautiously. It records whether the original study reports multi-agent coordination, role specialization, or collaborative workflow organization; it does not indicate that the multi-agent design has been proven superior to a comparable single-agent design. Among the reviewed systems, AgroAskAI provides comparative evidence against ChatGPT (GPT-5.1) and CROPWAT 8.0 and includes an evaluation of its Reviewer Agent, but this still does not amount to a complete cost-adjusted comparison with a matched single-agent architecture. For PhenoAssistant, Chat Demeter, and Orchestrator-Agent Trust, the available evidence mainly supports the feasibility of role specialization, workflow organization, or trust-aware orchestration, while matched single-agent baselines, additional model calls, communication overhead, latency, and correlated hallucinations are rarely quantified in a systematic manner. Therefore, current evidence is insufficient to claim that multi-agent systems are generally superior to single-agent systems after cost and reliability trade-offs are considered.

Overall, the comparative assessment suggests that current agricultural intelligent agents remain uneven in agentic maturity and deployment readiness. Some systems mainly strengthen knowledge access and contextual response generation; some organize multi-modal diagnosis or phenotyping workflows; and a smaller number incorporate executable tools, simulation, digital twins, or formal verification. The development of agricultural intelligent agents should therefore be evaluated less by the presence of an LLM or the number of covered scenarios, and more by evidence reported in the original studies regarding task decomposition, agronomic evidence applicability, tool validity, workflow reliability,

memory consistency, multi-agent coordination, practical validation, and validation against agricultural outcomes.

5. Challenges of Agricultural Intelligent Agents in Agricultural Applications

Agricultural intelligent agents are pushing smart agriculture beyond isolated model applications toward system-level frameworks that integrate knowledge organization, tool coordination, task planning, and process-oriented decision-making. However, most existing agricultural intelligent agents remain at the stages of methodological exploration, prototype development, or validation in limited scenarios, and a substantial gap still exists before they can achieve long-term, stable, trustworthy, and scalable application in real agricultural environments. Compared with general-purpose agents, agricultural intelligent agents must cope not only with inherent risks of large language models, such as hallucination, tool failure, and multi-agent coordination errors, but also with domain-specific challenges, including highly contextualized agricultural knowledge, dynamic production environments, real-world consequences of physical operations, and heterogeneous deployment conditions. Against this background, this section discusses six key challenges: agricultural knowledge trustworthiness and contextual adaptation; coupling of multi-source data, toolchains, and physical environments; reliability and verifiability of long task chains; multi-agent collaboration failures and human-agent coordination; security risks in agricultural agent systems; and economic feasibility, scalability, and integration with existing agricultural information systems.

The six major deployment challenges and their effects across agricultural task-chain stages are summarized in Table 6.

Table 6. Major deployment challenges of agricultural intelligent agents.

Challenge Dimension	Core Problem	Typical Manifestation in Agriculture	Affected Task-Chain Stage
Knowledge trustworthiness and contextual adaptation	Context-dependent agronomic knowledge across regions, crops, growth stages, and production systems	Contextual hallucination, outdated knowledge, regionally unsuitable advice, and weak minor-crop coverage	Knowledge services, advisory support, diagnosis, and decision support
Multi-source data, toolchains, and physical-environment coupling	Heterogeneous data, tools, interfaces, devices, and field conditions	Inconsistent sensor data, mismatched remote-sensing inputs, unstable API calls, weak interoperability, and infeasible plans	Data access, tool use, resource management, and operational planning
Reliability and verifiability of long task chains	Error accumulation across retrieval, tool use, reasoning, recommendation, and feedback	Plausible but unreliable outputs, subgoal deviation, unstable long-horizon execution, and weak process evaluation	Multi-step reasoning, diagnosis chains, decision workflows, and controlled execution
Multi-agent collaboration and human-agent coordination	Unclear role boundaries, shared states, handoffs, and human oversight	Error propagation, redundant or inconsistent outputs, unclear responsibility, and weak human review	Multi-agent advisory systems, diagnostic cooperation, workflow orchestration, and farm-level management
Security risks in agricultural agent systems	Expanded attack surface from tools, retrieval, memory, external content, and cyber-physical links	Prompt injection, memory poisoning, tool hijacking, unauthorized access, contaminated records, and cyber-physical risks	Retrieval, memory, tool use, farm databases, external platforms, and equipment-related workflows
Economic feasibility, scalability, and information-system integration	Cost, scalability, and compatibility constraints in real deployment	High inference, sensor, maintenance, expert-review, training, and integration costs	Advisory services, decision support, farm recordkeeping, farm management, and large-scale deployment

5.1. Agricultural Knowledge Trustworthiness and Insufficient Contextual Adaptation

The first challenge concerns whether the knowledge foundations of agricultural intelligent agents are trustworthy, applicable, and continuously updated. Agricultural knowledge is not a unified set of rules that can be directly transferred across regions and crops. Rather, it is highly dependent on climatic conditions, soil types, cropping systems, growth stages, input levels, and local management experience. A sowing recommendation, fertilization scheme, or pest and disease control measure that is valid in one region may become unsuitable or even misleading when transferred to another. AgriRegion has pointed out the problem of “contextual hallucination” in agricultural advisory services, which differs from general factual hallucination: a response may appear scientifically reasonable on the surface, but lose practical validity because it ignores geographical and production contexts [59]. Agricultural intelligent agents require both knowledge retrieval capabilities and the ability to judge the applicability boundaries of agricultural knowledge.

This issue is also prominent in multimodal agricultural agents. The study of AgriDoctor shows that, even with the introduction of large language models and vision-language models, agricultural diagnosis still faces limitations such as insufficient domain-specific datasets, inadequate domain adaptation, and the difficulty of deeply incorporating agricultural knowledge into the reasoning process [68]. When images, text, and domain knowledge jointly participate in diagnosis, the system may generate judgments that appear coherent but lack sufficient agronomic grounding if it cannot distinguish decisive evidence from auxiliary information, or identify the regional limitations of certain knowledge. Therefore, grounding in agricultural intelligent agents should not stop at knowledge retrieval. It should further involve source discrimination, recognition of applicable contexts, and hierarchical organization of evidence.

More broadly, the limitation is not confined to language-based agricultural knowledge. Current agricultural intelligent agents also inherit weaknesses from the agricultural data and perception systems on which they depend. Agricultural benchmarks indicate that existing large language models still have clear limitations in domain knowledge, understanding, reasoning, and generation tasks [93]. At the same time, agricultural remote sensing and mapping remain affected by limited annotated data, regional distribution shifts, and insufficient generalization across production environments [94]. These limitations are not separate technical problems. Within agent systems, weak domain reasoning, unstable perception, and poor cross-regional transfer can compound across the task chain. An incomplete knowledge base may skew problem interpretation, bias in a perception model may affect crop-state assessment, and limited regional generalization can carry through to later recommendations. Stronger language models alone will not address this issue; agricultural intelligent agents also require dependable domain data, robust perception models, and adaptation mechanisms for regional variation.

The construction of trustworthy agricultural knowledge involves at least three key difficulties. First, it is necessary to reduce factual errors and contextual mismatches produced by general-purpose large language models in agricultural domains. Second, structured, region-specific, and temporally updated agricultural knowledge resources need to be developed. Third, the reasoning process of agents should explicitly represent the applicability conditions, uncertainty, and source credibility of knowledge. Existing studies on agent hallucination have shown that hallucinations may occur not only at the stage of final answer generation, but also during task planning, tool invocation, and intermediate state judgment [95]. For agricultural intelligent agents, this means that once knowledge reliability problems occur, they may be amplified along the subsequent task chain and ultimately affect diagnosis, decision-making, and even execution outcomes.

Adaptation to different agro-climatic conditions and crop types therefore requires explicit mechanisms rather than only larger models or broader retrieval corpora. Future systems should combine region-aware retrieval, crop-specific profiles, growth-stage constraints, soil and climate filters, local extension knowledge, and historical farm records to determine whether a recommendation is applicable to a particular production context. They should also be able to ask for missing contextual information, such as crop variety, sowing date, irrigation condition, soil type, recent weather, and disease history, before generating high-responsibility advice. For cross-regional deployment, feedback from local agronomists and farmers should be used to update knowledge bases, calibrate decision rules, and identify cases where a recommendation should be restricted, revised, or escalated to human experts.

5.2. Difficulties in Coupling Multi-Source Data, Toolchains, and Physical Environments

For agricultural intelligent agents oriented toward complex decision-making, resource management, and operational coordination, stable collaboration with external resources such as sensors, remote-sensing platforms, meteorological services, crop models, agricultural machinery, and digital twin systems is a critical prerequisite for real-world deployment. Studies reviewed in Section 4, including CottonBot, AgriWorld, digital twin-based decision-support frameworks, and farm-level collaborative management systems, all show that the capability of agricultural intelligent agents depends heavily on access to external resources. However, current agricultural information systems still differ substantially in data formats, interface protocols, semantic representations, and system boundaries, making it difficult for agents to establish stable and reusable coordination mechanisms across heterogeneous resources.

The first layer of difficulty comes from the semantic fragmentation of agricultural data. Field sensors, UAV or satellite images, weather records, machinery operation logs, and farm management records are usually generated by different systems and at different spatial and temporal scales. Their meanings are not always directly compatible. A soil-moisture reading, a remote-sensing index, a weather forecast, and a management record may all describe the same field, but they may refer to different time windows, spatial units, measurement depths, or management objects. In such cases, data access alone does not guarantee usable context. Agricultural agents need to determine whether heterogeneous observations are comparable, whether their temporal and spatial resolutions are suitable for the task, and whether the field or crop unit referred to by one data source is consistent with that used by another. Research on semantic interoperability in smart agriculture has similarly emphasized the need for metadata, semantic annotation, and standardized descriptions to support cross-platform data sharing and machine reasoning [96]. Without such a semantic layer, an agent may appear to reason over multiple data sources while in fact combining mismatched evidence.

Tool use creates another source of uncertainty for agricultural intelligent agents. In a typical workflow, an agent has to choose an appropriate tool, prepare the required parameters, interpret the returned results, and decide whether a further action is needed. Each step depends on details that are easy to overlook. A weather API, for example, must be queried with an appropriate location and time window; a crop model depends on suitable soil, cultivar, management, and climate parameters; remote-sensing analysis requires a consistent field boundary and sensing date; and a digital twin is only meaningful when its assumptions fit the management scenario being considered. Tool-integrated LLM agents may fail at any of these points through misunderstanding tool descriptions, constructing unsuitable parameters, calling tools in an inappropriate order, or misreading returned outputs [97]. In agriculture, the consequences are not limited to a poor response, because such

outputs may later be used as evidence for irrigation, fertilization, disease control, or operation planning. What matters is therefore not simply whether an agent can access a tool, but whether it can use that tool within its practical conditions and interpret the result in the relevant agricultural setting.

The difficulty becomes more pronounced once an agent is linked to the physical environment. Agricultural fields are open systems, and observations can become outdated quickly. Crop growth, soil moisture, pest pressure, equipment availability, and weather may differ by the time a recommendation or operation plan is produced from the conditions recorded earlier. A plan may appear coherent in language while remaining impractical in the field because the observations are incomplete, the field state has changed, or assumptions about terrain and equipment are wrong. Work on embodied agents has similarly shown that language-driven planning can fail when task instructions are not sufficiently aligned with actual environmental states [98]. This concern applies directly to UAV scouting, machinery scheduling, robotic spraying, and field inspection. For agricultural robotic systems, this coupling difficulty is also reflected in perception and navigation under open-field or orchard conditions. Orchard localization, crop-row detection, and obstacle classification require robust adaptation to dynamic visual scenes, field structures, and misleading environmental cues [99–101]. Connecting devices is not enough: agricultural agents also need to relate natural-language objectives to sensed field conditions, operational constraints, and actions that can actually be carried out.

Transferability presents a related problem. Many agricultural agent prototypes are developed around a particular database, API, toolset, experimental field, or crop scenario. They may work satisfactorily within that setting, yet become difficult to reuse after changes of crop type, region, data infrastructure, or farm equipment. This is more than a matter of engineering convenience, since it determines whether a local prototype can become a reusable deployment framework. When an agent relies on a customized combination of data formats, tool interfaces, and task assumptions, its workflow may not remain valid outside the original environment. Standardized task interfaces, explicit context models, tool descriptions, and environmental feedback mechanisms are therefore needed alongside stronger model reasoning capability [96].

5.3. Reliability, Verifiability, and Evaluation of Long Task Chains

Agricultural intelligent agents derive much of their value from advancing tasks across multiple stages rather than completing one-off knowledge answering or single-point prediction. Agricultural advisory services may involve “problem understanding–knowledge retrieval–recommendation generation”; resource management may involve “state perception–tool invocation–scheme formulation”; and operational execution may involve “goal interpretation–plan generation–constraint checking–feedback refinement”. In such long task chains, errors at any stage may propagate through subsequent steps and affect the final judgment. Recent studies on long-horizon agents show that strong performance on short tasks does not naturally translate into stable reliability in long-duration tasks, and that success rate, consistency, and robustness may decline as task length and step dependency increase [102,103]. HORIZON further shows that long-horizon agentic systems are prone to subgoal deviation, persistence along incorrect trajectories, and instability across steps [103]. Reliability assessment should examine not only whether the final textual output appears reasonable, but also whether the agent maintains goal consistency, sufficient evidence, and traceable reasoning throughout the task process.

In parallel with long-chain reliability problems, current agricultural evaluation systems still exhibit clear gaps. Existing benchmarks have made progress in evaluating the local capabilities of agricultural large models and multimodal models. AgriEval evalu-

ates agricultural knowledge capabilities in terms of memory, understanding, reasoning, and generation [93]; AgReason assesses contextualized agricultural reasoning [104]; and AgMMU focuses on agricultural multimodal understanding and reasoning [105]. However, these benchmarks mainly evaluate knowledge question answering, agricultural reasoning, or multimodal understanding, rather than the process-level reliability of complete agent workflows. In other words, current evaluations have begun to examine whether models possess agricultural knowledge and reasoning ability, but remain insufficient for determining whether agents can stably complete full task chains involving retrieval, tool use, scheme generation, and feedback-based refinement.

Consistent with the capability and evidence gaps identified in Section 4.5, agricultural intelligent agents cannot be evaluated by simply extending conventional model benchmarks. A benchmark that measures whether a model answers agricultural questions correctly is useful, but it does not reveal whether an agent can select the right tool, interpret tool outputs correctly, maintain a consistent task state, or stop when evidence is insufficient. Recent reviews of LLM-agent evaluation have argued that planning, tool use, memory, robustness, safety, and cost should be assessed through more fine-grained and process-oriented methods [106]. In agricultural scenarios, these dimensions require further domain adaptation. Evaluation should ask whether the system identifies the agricultural objective accurately, whether the evidence it retrieves is applicable to the target crop and region, whether tool calls satisfy agronomic assumptions, whether intermediate conclusions remain consistent, and whether the system exposes uncertainty before producing a management suggestion. The evaluation object should therefore include task trajectories, evidence chains, tool behaviors, constraint compliance, and failure-handling processes, rather than only final textual outputs.

In agriculture, the consequences of a recommendation often emerge later and are influenced by conditions that cannot be fully controlled. An irrigation recommendation may appear appropriate when issued, but its value must eventually be judged against subsequent weather, soil response, and crop water status. A fertilization plan likewise needs to be considered in relation to yield response, nutrient loss, cost, and environmental constraints, while disease-control advice may become excessive or inadequate when infection pressure changes. For high-risk tasks, agent outputs should not move directly from generation to recommendation or execution. The digital twin-based decision-support framework and formal mission-planning studies discussed in Section 4 point to simulation, rule checking, or formal constraints as ways to examine candidate schemes before practical use [81,91]. The relevant question is not simply whether an agent returns a complete answer, but whether its reasoning path, tool use, assumptions, and constraints can be inspected before a decision affects production. Otherwise, an agricultural intelligent agent may appear convincing at the interface while remaining unreliable for decision support.

5.4. Multi-Agent Collaboration and Human-in-the-Loop Coordination

Multi-agent collaboration is widely regarded as an important pathway for improving the ability of agents to handle complex tasks, but it does not inherently guarantee higher reliability. Compared with a single-agent system, a multi-agent system must additionally address role assignment, information transmission, state sharing, result verification, and task termination. If these mechanisms are insufficiently designed, collaboration itself may become a new source of errors. A systematic analysis of mainstream multi-agent frameworks shows that common failure modes are mainly associated with insufficient system specification and role definition, misalignment in goal understanding or information transfer among agents, and the absence of effective verification and termination mechanisms [39].

Performance improvement in multi-agent systems depends less on increasing the number of agents than on clearly constrained and effectively verified collaboration mechanisms.

This problem is amplified in agricultural intelligent agents because agricultural tasks often depend on sequential handoffs among knowledge retrieval, diagnosis, resource decision-making, and action recommendation. A failure at one stage may not appear as an obvious error; it may instead enter the next stage as an apparently valid premise. For instance, contextually unsuitable agronomic knowledge may be accepted by a diagnostic agent, converted into a management suggestion by a decision agent, and only become visible as a problem when the recommendation is applied under real field conditions. The risk therefore lies not only in whether each agent performs its assigned role, but also in whether assumptions, uncertainty, evidence sources, and task states are preserved during inter-agent transfer.

For this reason, agricultural multi-agent systems require more than role specialization. They need explicit handoff protocols, shared state representations, confidence or evidence records, and escalation rules for cases in which evidence is incomplete or consequences are high. Human participation should also be treated as part of the system design rather than as a post hoc safety measure. Agronomists, farm managers, and farmers may need to intervene when recommendations involve costly inputs, pesticide use, machinery operations, or uncertain field observations. In such cases, the key design question is not how to maximize agent autonomy, but where human judgment should enter the workflow, what information should be exposed for review, and how responsibility is assigned when an agent-generated recommendation is accepted, revised, or rejected.

5.5. Security Risks in Agricultural Agent Systems

Security should be treated as a separate deployment challenge for agricultural intelligent agents, rather than as a minor extension of hallucination or model accuracy. The reason is that agent systems do not only generate textual advice. Once they are connected to retrieval systems, farm databases, sensor interfaces, weather services, crop models, machinery platforms, or digital twins, they begin to operate across multiple trust boundaries. A misleading answer may affect a user's judgment, but a compromised tool call, poisoned memory, or unauthorized database action may alter the information basis on which later agricultural decisions are made. This difference is important: security risks in agricultural agents arise from the combination of generative reasoning, external content, tool permissions, persistent memory, and agricultural digital infrastructure.

One important source of risk is external content. Tool-augmented agents often read web pages, retrieved documents, policy materials, pest and disease knowledge bases, database outputs, or responses returned by third-party tools. These materials may look like ordinary task evidence, but they may also contain instructions or misleading information that the agent should not follow. InjecAgent shows that indirect prompt injection can hide malicious instructions in external content read by tool-integrated agents, leading them to take actions that deviate from the original user intent [107]. AgentDojo further shows that tool-use agents remain vulnerable when task-relevant data and untrusted content are mixed in the same environment [108]. In agriculture, this risk is not limited to open-web retrieval. It may also appear when agents read extension documents, pesticide-use information, farm service outputs, input-market records, or locally maintained knowledge bases. The security problem, therefore, is not simply whether retrieved information is factually correct, but whether the agent can distinguish evidence to be used from instructions that should be ignored.

Tool use introduces a second trust boundary. Agricultural agents for irrigation, fertilization, disease diagnosis, recordkeeping, or operation planning may need to call weather

APIs, crop models, remote-sensing tools, farm databases, or equipment-related services. Even without malicious attacks, tool-integrated agents can be unstable in tool selection, parameter construction, invocation order, and interpretation of returned results [97]. Under unclear permissions or partly trusted environments, these routine tool-use failures can become security concerns rather than isolated execution mistakes. A weather API queried with an unsuitable time window may distort irrigation advice; an inappropriate crop-model configuration may affect yield-related reasoning; and a remote-sensing tool applied to the wrong field boundary may return misleading evidence about field conditions. When such errors occur alongside contaminated external content, excessive permissions, or weak parameter validation, their effects can extend beyond the local task. Agricultural agents therefore require bounded tool permissions, parameter checks, validation of returned outputs, and mechanisms that prevent a failed call from affecting the rest of the workflow.

Persistent memory and RAG-based knowledge repositories introduce a related risk because incorrect or malicious information may remain available after the original interaction has ended. This is relevant to agents that support cross-stage consultation, farm record tracking, personalized advisory services, or the reuse of historical management experience. AgentPoison shows that long-term memory or knowledge bases can be poisoned to trigger malicious or biased agent behavior without necessarily reducing performance on ordinary tasks [109]. Agent Security Bench also evaluates prompt injection, memory poisoning, tool attacks, and backdoor attacks across system prompts, user inputs, tool use, and memory retrieval [110]. In agricultural settings, farm records, disease-control cases, local agronomic knowledge, and historical management decisions may be consulted repeatedly. If these sources are corrupted, the same distorted information can reappear in later irrigation, fertilization, crop-protection, or farm-management analyses, sometimes without being obvious in a single interaction.

Integration with IoT-enabled precision agriculture systems and cyber-physical infrastructures raises the stakes further. Precision agriculture relies on sensor networks, remote-control devices, machinery terminals, cloud platforms, and farm management databases, while threat modeling for IoT-enabled precision agriculture has identified spoofing, data tampering, information disclosure, denial of service, and privilege escalation as relevant risks [111]. When an agricultural agent coordinates these resources, manipulated sensor values or poorly controlled access privileges may affect resource allocation, operation planning, and management records, rather than only the wording of a recommendation. Contaminated soil-moisture readings may bias irrigation decisions, altered disease-monitoring records may affect crop-protection strategies, and unauthorized access to farm databases or equipment interfaces may expose operational information. Agricultural agent security is therefore tied to both data governance and cyber-physical operation, not only to LLM safety.

Security failures may also spread along the task chain. A compromised retrieval result can influence later tool selection, poisoned memory may remain inactive until a later crop stage, and an unnecessary permission can allow a local error to reach several databases or services. Looking only at the final answer is unlikely to reveal such failures. Source credibility, retrieval filtering, tool-call boundaries, memory-writing rules, access permissions, and audit trails need to be considered together at the workflow level. Existing agent-security studies also suggest that stronger defenses can reduce task utility or workflow flexibility [110]. This trade-off is particularly relevant in agriculture, where systems still need to remain usable for farmers, agronomists, and farm managers under time-sensitive conditions.

Security controls should therefore be built into agricultural agent systems rather than added after deployment. External information needs to be handled separately from exe-

cutable instructions, with its provenance retained where possible. Tool calls should operate under explicit permission scopes and validated parameters, while long-term memory and RAG repositories require controlled updates and periodic checking. Human confirmation and rollback mechanisms are especially important when an agent may influence pesticide use, irrigation adjustment, machinery operation, or farm-data modification. Security evaluation should also follow complete task trajectories rather than isolated outputs. Reliable deployment will depend on whether retrieval, memory, tool use, permissions, and human oversight are controlled together within the agricultural workflow.

5.6. Economic Feasibility, Scalability, and Integration with Agricultural Information Systems

Economic feasibility is a central condition for moving agricultural intelligent agents from prototypes to real production. The cost of implementation is not limited to model inference or API calls. It also includes sensors, network connectivity, data cleaning, knowledge-base maintenance, tool-interface development, model updating, farmer training, expert review, and long-term technical support. These costs may be acceptable for high-value crops, research farms, or large commercial farms, but may become prohibitive for smallholders or regions with weak digital infrastructure. Therefore, the economic value of agricultural intelligent agents should be evaluated not only by whether they improve response quality or automate a workflow, but also by whether they reduce decision delay, lower labor requirements, improve input efficiency, or support better production outcomes over a complete agricultural cycle.

Existing studies provide some practical-facing examples, but the evidence remains uneven. Farmer.Chat represents one of the clearest examples of real-world deployment in agricultural advisory services, showing how generative AI can be used to scale localized support for smallholder farmers. CottonBot demonstrates the potential of connecting LLM-RAG and agentic tools with soil-moisture sensors and weather APIs for field-specific irrigation advice, but it should still be regarded as a field-data-connected prototype rather than evidence of full on-farm deployment. Meta Ag 2.0 illustrates how agentic AI can support farm recordkeeping by connecting user interaction, private databases, contextual questioning, input validation, and record writing. These examples indicate practical potential, but they also show that successful deployment requires more than technical feasibility. Evidence on cost-effectiveness, long-term user retention, maintenance burden, agronomic outcome improvement, and scalability across regions and farm types remains limited.

Scalability should also be considered at multiple levels. At the user level, agents must support farmers, agronomists, extension workers, and farm managers with different expertise, languages, and digital access conditions. At the farm level, the system must adapt to different field sizes, crop types, machinery configurations, and management practices. At the regional level, it must cope with differences in climate, soil, pest pressure, policies, and advisory infrastructures. At the platform level, it must integrate with existing farm management information systems, sensor platforms, remote-sensing services, machinery terminals, digital twins, and agricultural databases. If each agent is tightly coupled to a customized toolset or local data format, scaling will require repeated engineering work and will be difficult to sustain. Future agricultural intelligent agents should therefore be designed with modular architectures, standardized interfaces, configurable knowledge bases, and human-review mechanisms that support gradual deployment rather than one-off demonstrations.

6. Future Directions

The reviewed studies suggest that agricultural intelligent agents are developing unevenly across knowledge services, crop-state understanding, resource management,

and farm-level coordination. Advisory and diagnostic systems have begun to show practical value, but many still depend on limited context modeling, fragmented data access, and weak process-level verification. More advanced frameworks involving tool execution, digital twins, formal verification, or farm-level workflows indicate possible routes toward deployable systems, yet these routes remain technically demanding and insufficiently validated across crops, regions, and production cycles. Future research should therefore focus less on adding more agent functions in isolation and more on improving the conditions under which agents can understand agricultural contexts, integrate heterogeneous data and tools, and execute task chains in a reliable and inspectable manner. Three directions are particularly important.

6.1. Developing Multimodal and Context-Aware Agricultural Intelligent Agents

Multimodality should be a central direction for agricultural intelligent agents, but it should not be reduced to the simple addition of image, speech, remote-sensing, or sensor inputs. Agricultural problems usually appear as incomplete and heterogeneous evidence rather than as well-formed textual tasks. A farmer may describe leaf symptoms orally while uploading a field image; a regional manager may need to combine remote-sensing observations with recent weather anomalies; and a farm decision may depend on sensor readings, crop stage, historical operations, and local management rules. In such situations, the agent must determine how different forms of evidence relate to the same agricultural object, whether they refer to the same time and spatial scale, and which contextual factors are necessary before a judgment can be made. If multimodal inputs are only concatenated without agricultural context modeling, the system may become more information-rich but not necessarily more reliable.

Existing studies have begun to demonstrate the potential of multimodal agricultural agents. AgriDoctor integrates image recognition, object detection, knowledge retrieval, and large language model reasoning within a unified diagnostic framework, promoting agricultural diagnosis from single visual classification toward multimodal reasoning [68]. MA3 further proposes a multimodal agricultural agent architecture that incorporates classification, detection, visual question answering, tool selection, and agent evaluation into a unified framework [112]. Meanwhile, AgriGPT-Omni attempts to unify speech, vision, and text inputs, and constructs multilingual agricultural speech data and a trimodal evaluation benchmark, indicating that future agricultural agents should be able not only to “see” crops, but also to “listen to” farmers and understand multilingual production contexts [113].

Recent progress in generative perception and annotation workflows further extends this discussion. Rana et al. proposed a generative semantic reconstruction workflow for spectrally heterogeneous rangeland imagery, combining prompt-constrained reconstruction, ResNet101-based scene parsing, exclusion-mask generation, vegetation-prior refinement, and SAM-based final object-mask generation to produce annotation-ready vegetation priors. The study also showed that generated data may introduce domain shifts, suggesting that generated visual evidence should not be accepted into downstream analysis only because it is visually clearer or easier to annotate. This work should not be regarded as an agricultural intelligent agent system itself. Rather, it provides an enabling perception and annotation workflow for future multimodal agricultural agents, and it illustrates why generated visual evidence should be examined for semantic preservation, architecture-dependent effects, and downstream reliability before being incorporated into agent reasoning [114].

Therefore, the multimodal development of agricultural intelligent agents should not be understood simply as the addition of more input modalities. Rather, it should support

context-aware agricultural cognition by coordinating multimodal perception models, generative foundation models, region-specific knowledge retrieval, agricultural knowledge graphs, and verification mechanisms. In this way, agents can establish a continuous chain among observing agricultural objects, understanding user expressions, locating applicable knowledge, assessing the reliability of multimodal evidence, and forming context-aware judgments, thereby evolving from language-centered systems toward scenario-aware systems.

6.2. Toward Data-Integrated Agricultural Decision-Making and Collaborative Management Systems

The further development of agricultural intelligent agents depends not only on advances in model reasoning and task execution mechanisms, but also on whether agricultural data infrastructures can achieve a higher level of integration and organization. In real-world agricultural scenarios, information is often distributed across remote-sensing imagery, field sensors, meteorological services, soil attributes, operation records, geospatial data, and farm management platforms, with substantial differences in temporal scale, spatial resolution, sampling frequency, and semantic representation. Without stable data fusion mechanisms, even agents with strong language understanding and tool-use capabilities may only switch passively among local data sources, rather than form a holistic understanding of agricultural contexts. Conventional crop-monitoring studies have already shown this requirement at the feature level; for example, UAV-based LAI estimation can depend on the fusion of multispectral, spatial, and vegetation-index information rather than on a single image source [115]. Future research should promote the transition of multi-source heterogeneous agricultural data from isolated coexistence to coordinated representation.

A prerequisite for this transition is the standardization and unification of agricultural data. Agricultural data are often recorded using different units, naming conventions, spatial references, temporal intervals, crop-stage definitions, and management vocabularies. For example, the same field operation may be described differently in farm records, machinery logs, extension documents, and sensor platforms. Without unified metadata and semantic descriptions, agricultural agents may retrieve or combine data that appear relevant but actually refer to different spatial units, time windows, crop varieties, or management conditions. Future data foundations should therefore standardize core agricultural entities, including field boundaries, crop species and varieties, growth stages, soil attributes, weather records, sensor observations, input applications, machinery operations, and management events. Such standardization should also include data quality labels, provenance records, uncertainty descriptions, and access permissions, so that agents can judge not only what a data item says, but also whether it is reliable and usable for a specific task.

In addition to unified data descriptions, agricultural intelligent agents require unified data exchange protocols between different systems and platforms. Future systems should support machine-readable interfaces through which agents can query, update, and verify information across farm management information systems, sensor networks, remote-sensing platforms, agricultural machinery terminals, digital twins, and advisory databases. These protocols should define common data schemas, metadata vocabularies, tool descriptions, authentication rules, permission scopes, logging mechanisms, and error-handling procedures. For agent systems, the purpose of data exchange protocols is not only to move data across platforms, but also to make tool calls traceable, parameter use checkable, and data flows auditable. Without such protocols, agricultural agents may remain dependent on customized connectors and local engineering work, which would limit their scalability and weaken their reliability in multi-platform production environments.

Existing studies have begun to demonstrate the potential of this direction. Farm-LightSeek incorporates multi-source data, including images, weather information, and geo-

graphic information, into an edge-side agricultural analysis framework, exploring cross-modal monitoring and low-latency decision-making under resource-constrained conditions [116]. UniCrop, oriented toward crop yield prediction, automatically performs the acquisition, cleaning, alignment, and feature construction of multi-source environmental data, including satellite, meteorological, soil, and topographic information [117]. The bottleneck of intelligent agricultural analysis lies not only in the model itself, but also in whether a reusable and scalable data engineering pipeline can be established. Meanwhile, research on multimodal foundation models for remote sensing suggests that agricultural and geospatial tasks are shifting from single-image interpretation toward joint modeling of optical imagery, radar, LiDAR, text, and geographic information, although data alignment, cross-modal transfer, and large-scale deployment remain key challenges [118]. These advances indicate that future agricultural intelligent agents require not simply more data, but an integrated agricultural data foundation capable of supporting comprehensive reasoning.

To achieve this goal, future research may further develop data fusion frameworks specifically designed for agricultural intelligent agents, with a focus on spatiotemporal alignment, semantic unification, quality control, provenance tracking, and interface standardization of multi-source data. Such frameworks should provide agents with organized representations of agricultural states, enabling more robust analysis of crop growth, environmental change, resource allocation, and risk assessment. As multi-source agricultural data evolve into a shared, computable, and callable foundation, agricultural intelligent agents are expected to move beyond isolated tools and local observations toward continuous judgment and context-aware decision-making in complex agricultural information environments.

6.3. Building Reliable Harnesses for Agricultural Intelligent Agents

Model scaling and additional modules alone are insufficient for reliable agricultural agent deployment. Greater attention should be paid to harness engineering that supports their stable operation. Many existing agricultural agent systems are still constructed in a compositional manner, typically combining a foundation model, a retrieval module, and several external tools. Although such systems can demonstrate functionality in specific tasks, they often expose limitations in cross-tool workflows, multi-stage processes, long task chains, and high-responsibility scenarios, including insufficient state maintenance, unstable tool invocation, limited error recovery, and unclear safety boundaries. Recent agent research suggests that practical agent capabilities increasingly depend on the organization of capabilities outside the model itself, including memory, skills, interaction protocols, and the harness constructed around the execution process [119].

For agricultural intelligent agents, the role of a harness is to elevate complex agricultural tasks from local tool invocation to observable, constrained, and recoverable process execution. Future agricultural agents may construct a unified harness layer around knowledge bases, remote-sensing analysis services, weather interfaces, crop models, digital twin modules, databases, and equipment-control interfaces. This layer should support context management, tool registration, invocation permissions, parameter validation, intermediate-evidence preservation, failure retry, exception rollback, and human takeover. AutoHarness shows that a well-designed harness can prevent agents from performing explicitly prohibited behaviors in the execution environment by automatically generating code-level constraints, thereby reducing prohibited actions [120]. SafeHarness further incorporates input filtering, hierarchical verification, tool-permission isolation, and state rollback into the agent lifecycle, providing a system-level safety perspective for high-risk deployment scenarios [121].

This direction is particularly important for agriculture because agricultural tasks often depend on multi-source data, involve substantial real-world consequences, and have long

feedback cycles. Future agricultural intelligent agents should develop harness-centered operational architectures and integrate them with digital twins, formal constraints, and intermediate verification mechanisms, thereby establishing a closed-loop task execution process of “goal understanding–resource invocation–evidence generation–scheme verification–controlled output”. In this way, the development focus of agricultural intelligent agents will shift from whether the model is sufficiently intelligent to whether the system is sufficiently reliable for real-world agricultural production environments.

Responsible deployment also requires explicit attention to ethical and governance issues. In agriculture, ethical risks are not limited to general concerns about algorithmic bias or privacy. They also involve farmer data ownership, informed consent for data reuse, responsibility for agronomic recommendations, unequal access to digital services, and the consequences of automated advice on pesticide use, irrigation, machinery operations, or input purchasing. Agricultural intelligent agents should therefore expose evidence sources, uncertainty, and decision constraints in forms that farmers and agronomists can inspect. Human authorization should be required for high-cost or safety-sensitive actions, and users should have mechanisms to contest, revise, or roll back agent-generated recommendations or records. Ethical governance should be treated as part of system architecture rather than as an external policy layer added after deployment.

7. Conclusions

This review examined agricultural intelligent agents as an emerging system-level paradigm in smart agriculture. The focus was not on all agricultural applications that use large language models, but on systems that connect foundation-model reasoning with agricultural tasks, external resources, workflow organization, feedback, or collaborative mechanisms. Under this boundary, agricultural intelligent agents differ from traditional MAS/ABM, standalone agricultural foundation models, static RAG-based question-answering systems, and task-specific agricultural AI models. Their defining feature is not the presence of a language interface, but the ability to organize agricultural knowledge, dynamic data, analytical tools, and decision processes around task progression.

The studies reviewed in this paper suggest that agricultural intelligent agents are developing unevenly across application types and maturity dimensions. Knowledge-service systems are relatively mature as entry-level applications, but many still remain close to enhanced advisory tools when they lack task tracking, verification, or links to downstream decisions. Monitoring, diagnosis, and phenotyping systems show stronger workflow characteristics because they connect perception results with reasoning and uncertainty assessment. Resource-management and farm-level systems face a higher bar, since their outputs may affect irrigation, fertilization, crop protection, records, devices, or operational plans. The comparative assessment in this review further shows that existing studies differ substantially in the extent to which they evaluate task decomposition, tool-use validity, workflow reliability, memory consistency, multi-agent coordination, practical validation, deployment cost, and operational outcomes. For this reason, agricultural intelligent agents should be judged less by whether they use LLMs and more by whether they provide evidence for tool-grounded reasoning, task-chain control, process verification, and deployment readiness.

The main constraints are not isolated technical defects. Agricultural knowledge is highly context-dependent, and recommendations that appear scientifically reasonable may still fail under specific regional, crop-stage, or management conditions. Heterogeneous sensors, remote-sensing products, crop models, APIs, farm records, and equipment interfaces also make it difficult for agents to build a stable task context. In longer workflows, small errors in retrieval, tool invocation, intermediate reasoning, or inter-agent handoff

may be carried forward before they become visible. When memory, databases, external platforms, or equipment-related interfaces are involved, reliability further intersects with data governance, permission control, security, and cyber–physical safety.

Future research should move from capability demonstration toward dependable system construction. Larger models and more modules may improve local performance, but they cannot replace context modeling, economic feasibility analysis, scalable deployment design, standardized data and tool interfaces, unified data exchange protocols, process-level evaluation, reliable harnesses, evidence preservation, ethical governance, human review, and rollback mechanisms for high-risk actions. The value of agricultural intelligent agents will depend on whether their evidence sources, tool calls, assumptions, uncertainty, and decision constraints remain inspectable before outputs are used in production. At present, most systems are still exploratory architectures rather than mature deployment solutions; their next test is whether they can remain reliable, safe, and transparent under changing field conditions and incomplete information.

Author Contributions: Conceptualization, W.T., Q.S. and Q.M.; methodology, W.T.; investigation, W.T. and Q.S.; writing—original draft preparation, W.T.; writing—review and editing, W.T. and Q.M.; visualization, W.T. and Q.S.; supervision, C.C. and Q.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Project of Faculty of Agricultural Engineering of Jiangsu University under Grant NGXB20240101.

Institutional Review Board Statement: Not applicable

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, J.; Xu, M.; Xiang, L.; Chen, D.; Zhuang, W.; Yin, X.; Li, Z. Foundation models in smart agriculture: Basics, opportunities, and challenges. *Comput. Electron. Agric.* **2024**, *222*, 109032. [[CrossRef](#)]
2. Yin, S.; Xi, Y.; Zhang, X.; Sun, C.; Mao, Q. Foundation models in agriculture: A comprehensive review. *Agriculture* **2025**, *15*, 847. [[CrossRef](#)]
3. Wang, L.; Ma, C.; Feng, X.; Zhang, Z.; Yang, H.; Zhang, J.; Chen, Z.; Tang, J.; Chen, X.; Lin, Y.; et al. A survey on large language model based autonomous agents. *Front. Comput. Sci.* **2024**, *18*, 186345. [[CrossRef](#)]
4. Xi, Z.; Chen, W.; Guo, X.; He, W.; Ding, Y.; Hong, B.; Zhang, M.; Wang, J.; Jin, S.; Zhou, E.; et al. The rise and potential of large language model based agents: A survey. *Sci. China Inf. Sci.* **2025**, *68*, 121101. [[CrossRef](#)]
5. FAO. *The Future of Food and Agriculture: Alternative Pathways to 2050*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2018; p. 228.
6. Ge, C.; Zhang, G.; Wang, Y.; Shao, D.; Song, X.; Wang, Z. Research Status and Development Trends of Artificial Intelligence in Smart Agriculture. *Agriculture* **2025**, *15*, 2247. [[CrossRef](#)]
7. Wooldridge, M.; Jennings, N.R. Intelligent agents: Theory and practice. *Knowl. Eng. Rev.* **1995**, *10*, 115–152. [[CrossRef](#)]
8. Jennings, N.R.; Sycara, K.; Wooldridge, M. A roadmap of agent research and development. *Auton. Agents Multi-Agent Syst.* **1998**, *1*, 7–38. [[CrossRef](#)]
9. Brooks, R.A. Intelligence without representation. *Artif. Intell.* **1991**, *47*, 139–159. [[CrossRef](#)]
10. Rao, A.S.; Georgeff, M.P. BDI Agents: From Theory to Practice. In Proceedings of the Icmas, San Francisco, CA, USA, 12–14 June 1995; Volume 95, pp. 312–319.
11. Maes, P. Agents that reduce work and information overload. *Commun. ACM* **1994**, *37*, 30–40. [[CrossRef](#)]
12. Macal, C.; North, M. Tutorial on agent-based modelling and simulation. *J. Simul.* **2010**, *4*, 151–162. [[CrossRef](#)]
13. Yao, S.; Zhao, J.; Yu, D.; Du, N.; Shafran, I.; Narasimhan, K.; Cao, Y. React: Synergizing Reasoning and Acting in Language Models. In Proceedings of the 11th International Conference on Learning Representations, ICLR, Kigali, Rwanda, 1–5 May 2023.
14. Schick, T.; Dwivedi-Yu, J.; Dessì, R.; Raileanu, R.; Lomeli, M.; Hambro, E.; Zettlemoyer, L.; Cancedda, N.; Scialom, T. Toolformer: Language models can teach themselves to use tools. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 68539–68551. [[CrossRef](#)]

15. Singh, N.; Wang'ombe, J.; Okanga, N.; Zelenska, T.; Repishti, J.; Mishra, S.; Manokaran, R.; Singh, V.; Rafiq, M.I.; Gandhi, R.; et al. Farmer. Chat: Scaling AI-Powered agricultural services for smallholder farmers. *arXiv* **2024**, arXiv:2409.08916.
16. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.T.; Rocktäschel, T.; et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 9459–9474.
17. Yang, S.; Liu, Z.; Mayer, W.; Ding, N.; Wang, Y.; Huang, Y.; Wu, P.; Li, W.; Li, L.; Zhang, H.Y.; et al. Shizishangpt: An agricultural large language model integrating tools and resources. In *Proceedings of the Web Information Systems Engineering—WISE 2024*; Springer: Singapore, 2025; Volume 15439, pp. 284–298.
18. Kandamali, D.F.; Porter, W.M.; Porter, E.; McLemore, A.; Rains, G.C. CottonBot: An AI-Driven Cotton Farming Assistant and Irrigation Advisor Using LLM-RAG and Agentic AI Tools. *Smart Agric. Technol.* **2025**, *12*, 101640. [\[CrossRef\]](#)
19. Zhang, Z.; Zhang, J.; Liu, H.; Lv, Q.; Yang, J.; Cai, K.; Wang, K. AgriWorld: A World Tools Protocol Framework for Verifiable Agricultural Reasoning with Code-Executing LLM Agents. *arXiv* **2026**, arXiv:2602.15325.
20. Summers, T.R.; Yao, S.; Narasimhan, K.; Griffiths, T.L. Cognitive Architectures for Language Agents. *arXiv* **2024**, arXiv:2309.02427.
21. Park, J.S.; O'Brien, J.; Cai, C.J.; Morris, M.R.; Liang, P.; Bernstein, M.S. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual Acm Symposium on User Interface Software and Technology*, San Francisco, CA, USA, 29 October–1 November 2023; pp. 1–22.
22. Packer, C.; Wooders, S.; Lin, K.; Fang, V.; Patil, S.G.; Stoica, I.; Gonzalez, J.E. MemGPT: Towards LLMs as Operating Systems. *arXiv* **2023**, arXiv:2310.08560.
23. Wang, G.; Xie, Y.; Jiang, Y.; Mandlekar, A.; Xiao, C.; Zhu, Y.; Fan, L.; Anandkumar, A. Voyager: An Open-Ended Embodied Agent with Large Language Models. *arXiv* **2023**, arXiv:2305.16291.
24. Wang, L.; Xu, W.; Lan, Y.; Hu, Z.; Lan, Y.; Lee, R.K.W.; Lim, E.P. Plan-and-solve prompting: Improving zero-shot chain-of-thought reasoning by large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 2609–2634.
25. Yao, S.; Yu, D.; Zhao, J.; Shafran, I.; Griffiths, T.; Cao, Y.; Narasimhan, K. Tree of thoughts: Deliberate problem solving with large language models. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 11809–11822. [\[CrossRef\]](#)
26. Patil, S.G.; Zhang, T.; Wang, X.; Gonzalez, J.E. Gorilla: Large language model connected with massive apis. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 126544–126565. [\[CrossRef\]](#)
27. Qin, Y.; Liang, S.; Ye, Y.; Zhu, K.; Yan, L.; Lu, Y.; Lin, Y.; Cong, X.; Tang, X.; Qian, B.; et al. Toolllm: Facilitating large language models to master 16000+ real-world apis. *Proc. Int. Conf. Learn. Represent.* **2024**, *2024*, 9695–9717.
28. Zhang, Y.; Zhang, B.; Shen, C.; Liu, H.; Huang, J.; Tian, K.; Tang, Z. Review of the field environmental sensing methods based on multi-sensor information fusion technology. *Int. J. Agric. Biol. Eng.* **2024**, *17*, 1–13. [\[CrossRef\]](#)
29. Shinn, N.; Cassano, F.; Gopinath, A.; Narasimhan, K.; Yao, S. Reflexion: Language agents with verbal reinforcement learning. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 8634–8652. [\[CrossRef\]](#)
30. Gou, Z.; Shao, Z.; Gong, Y.; Yang, Y.; Duan, N.; Chen, W. Critic: Large language models can self-correct with tool-interactive critiquing. *Proc. Int. Conf. Learn. Represent.* **2024**, *2024*, 57734–57811.
31. Tran, K.T.; Dao, D.; Nguyen, M.D.; Pham, Q.V.; O'Sullivan, B.; Nguyen, H.D. Multi-agent collaboration mechanisms: A survey of llms. *arXiv* **2025**, arXiv:2501.06322.
32. Li, G.; Hammoud, H.; Itani, H.; Khizbullin, D.; Ghanem, B. Camel: Communicative agents for" mind" exploration of large language model society. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 51991–52008. [\[CrossRef\]](#)
33. Wu, Q.; Bansal, G.; Zhang, J.; Wu, Y.; Li, B.; Zhu, E.; Jiang, L.; Zhang, X.; Zhang, S.; Liu, J.; et al. Autogen: Enabling next-gen LLM applications via multi-agent conversations. In *Proceedings of the First Conference on Language Modeling*, Philadelphia, PA, USA, 7–9 October 2024.
34. Hong, S.; Zhuge, M.; Chen, J.; Zheng, X.; Cheng, Y.; Wang, J.; Zhang, C.; Yau, S.; Lin, Z.; Zhou, L.; et al. MetaGPT: Meta programming for a multi-agent collaborative framework. *Proc. Int. Conf. Learn. Represent.* **2024**, *2024*, 23247–23275.
35. Chen, W.; Su, Y.; Zuo, J.; Yang, C.; Yuan, C.; Chan, C.M.; Yu, H.; Lu, Y.; Hung, Y.H.; Qian, C.; et al. Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors. *Proc. Int. Conf. Learn. Represent.* **2024**, *2024*, 20094–20136.
36. Qian, C.; Liu, W.; Liu, H.; Chen, N.; Dang, Y.; Li, J.; Yang, C.; Chen, W.; Su, Y.; Cong, X.; et al. Chatdev: Communicative agents for software development. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, Bangkok, Thailand, 11–16 August 2024; Volume 1: Long Papers, pp. 15174–15186.
37. Du, Y.; Li, S.; Torralba, A.; Tenenbaum, J.B.; Mordatch, I. Improving factuality and reasoning in language models through multiagent debate. In *Proceedings of the Forty-First International Conference on Machine Learning*, Vienna, Austria, 21–27 July 2024.
38. Chan, C.M.; Chen, W.; Su, Y.; Yu, J.; Xue, W.; Zhang, S.; Fu, J.; Liu, Z. Chateval: Towards better llm-based evaluators through multi-agent debate. *Proc. Int. Conf. Learn. Represent.* **2024**, *2024*, 9079–9093.
39. Cemri, M.; Pan, M.Z.; Yang, S.; Agrawal, L.A.; Chopra, B.; Tiwari, R.; Keutzer, K.; Parameswaran, A.; Klein, D.; Ramchandran, K.; et al. Why do multi-agent llm systems fail? *Adv. Neural Inf. Process. Syst.* **2026**, *38*, 135676–135729. [\[CrossRef\]](#)

40. Liu, X.; Yu, H.; Zhang, H.; Xu, Y.; Lei, X.; Lai, H.; Gu, Y.; Ding, H.; Men, K.; Yang, K.; et al. Agentbench: Evaluating llms as agents. *Proc. Int. Conf. Learn. Represent.* **2024**, 2024, 52989–53046.
41. Mialon, G.; Fourier, C.; Wolf, T.; LeCun, Y.; Scialom, T. Gaia: A benchmark for general ai assistants. *Proc. Int. Conf. Learn. Represent.* **2024**, 2024, 9025–9049. [[CrossRef](#)]
42. Yao, S.; Shinn, N.; Razavi, P.; Narasimhan, K.R. τ -bench: A Benchmark for Tool-Agent-User Interaction in Real-World Domains. In Proceedings of the International Conference on Learning Representations, Singapore, 24–28 April 2025.
43. Chen, F.; Stogiannidis, I.; Wood, A.; Bueno, D.; Williams, D.; Macfarlane, F.; Grieve, B.D.; Wells, D.; Atkinson, J.A.; Hawkesford, M.J.; et al. A conversational multi-agent AI system for automated plant phenotyping. *Nat. Commun.* **2026**, *17*, 6391. [[CrossRef](#)] [[PubMed](#)]
44. Cantonjos, N.A.P.; Biswas, A. AgroAskAI: A Multi-Agent AI Framework for Supporting Smallholder Farmers' Enquiries Globally. *Proc. AAAI Conf. Artif. Intell.* **2026**, *40*, 40217–40223. [[CrossRef](#)]
45. Li, Y.; Wang, J.; Yuan, Z.; Zhang, H. Agent Technology for Agricultural Intelligence: Methodological Framework and Applications. *Electronics* **2026**, *15*, 1547. [[CrossRef](#)]
46. Liu, E.; Yang, H.; Hostens, M. Evaluating Small Language Models for Agentic On-Farm Decision Support Systems. *arXiv* **2025**, arXiv:2512.14043.
47. Taha, M.F.; Mao, H.; Zhang, Z.; Elmasry, G.; Awad, M.A.; Abdalla, A.; Mousa, S.; Elwakeel, A.E.; Elsherbiny, O. Emerging technologies for precision crop management towards agriculture 5.0: A comprehensive overview. *Agriculture* **2025**, *15*, 582. [[CrossRef](#)]
48. Keating, B.A.; Carberry, P.S.; Hammer, G.L.; Probert, M.E.; Robertson, M.J.; Holzworth, D.; Huth, N.I.; Hargreaves, J.N.; Meinke, H.; Hochman, Z.; et al. An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.* **2003**, *18*, 267–288. [[CrossRef](#)]
49. Jones, J.W.; Hoogenboom, G.; Porter, C.H.; Boote, K.J.; Batchelor, W.D.; Hunt, L.A.; Wilkens, P.W.; Singh, U.; Gijsman, A.J.; Ritchie, J.T. The DSSAT cropping system model. *Eur. J. Agron.* **2003**, *18*, 235–265. [[CrossRef](#)]
50. Yu, Z.; Lin, M.; Chen, Y.; He, X.; Lu, K.; Wen, K.; Sun, W.; Fu, W. Predicting and Optimizing Cotton Yield and Nitrogen-Phosphorus Fertilization Using the DSSAT Model and Ecological Stoichiometry Theory. *J. Soil Sci. Plant Nutr.* **2025**, *25*, 8481–8495. [[CrossRef](#)]
51. Chauhdary, J.N.; Li, H.; Akbar, N.; Javaid, M.; Rizwan, M.; Akhlaq, M. Evaluating corn production under different plant spacings through integrated modeling approach and simulating its future response under climate change scenarios. *Agric. Water Manag.* **2024**, *293*, 108691. [[CrossRef](#)]
52. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [[CrossRef](#)] [[PubMed](#)]
53. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
54. Desai, S.V.; Guo, W.; Balasubramanian, V.N.; Chandra, A.L. Computer Vision with Deep Learning for Plant Phenotyping in Agriculture: A Survey. *arXiv* **2020**, arXiv:2006.11391.
55. Mulla, D.J. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* **2013**, *114*, 358–371. [[CrossRef](#)]
56. Tzounis, A.; Katsoulas, N.; Bartzanas, T.; Kittas, C. Internet of Things in agriculture, recent advances and future challenges. *Biosyst. Eng.* **2017**, *164*, 31–48. [[CrossRef](#)]
57. Hasan, M.M.; Islam, M.U.; Sadeq, M.J. Towards the technological adaptation of advanced farming through artificial intelligence, the internet of things, and robotics: A comprehensive overview. *Artif. Intell. Smart Agric. Technol.* **2022**, 21–42. [[CrossRef](#)]
58. Klose, J.; Schröder, M.; Becker, S.; Bernardi, A.; Ruckelshausen, A. Data Preparation in Agriculture Through Automated Semantic Annotation–Basis for a Wide Range of Smart Services. *arXiv* **2019**, arXiv:1911.06606.
59. Fanuel, M.; Mahmoud, M.N.; Marshal, C.C.; Lakhota, V.; Dari, B.; Roy, K.; Zhang, S. AgriRegion: Region-Aware Retrieval for High-Fidelity Agricultural Advice. *arXiv* **2025**, arXiv:2512.10114.
60. Wu, Y.; Chen, L.; Yang, N.; Sun, Z. Research Progress of Deep Learning-Based Artificial Intelligence Technology in Pest and Disease Detection and Control. *Agriculture* **2025**, *15*, 2077. [[CrossRef](#)]
61. Liu, H.; Zhan, B.; Fang, R.; Zhang, Y.; Ma, Y.; Shen, Z.; Mao, Q. Recent advances in pest and disease recognition: A comprehensive review. *J. Agric. Eng.* **2025**, *56*, 1776. [[CrossRef](#)]
62. Cong, C.; Cao, G.; Li, Y.; Liu, D.; Ma, B.; Zhang, J.; Li, L.; Hu, J. Research on Monitoring Methods for the Appropriate Rice Harvest Period Based on Multispectral Remote Sensing. *Discrete Dyn. Nat. Soc.* **2022**, 2022, 1519667. [[CrossRef](#)]
63. Pei, H.; Sun, Y.; Huang, H.; Zhang, W.; Sheng, J.; Zhang, Z. Weed detection in maize fields by UAV images based on crop row preprocessing and improved YOLOv4. *Agriculture* **2022**, *12*, 975. [[CrossRef](#)]
64. Ji, W.; Pan, Y.; Xu, B.; Wang, J. A real-time apple targets detection method for picking robot based on ShufflenetV2-YOLOX. *Agriculture* **2022**, *12*, 856. [[CrossRef](#)]
65. Wang, A.; Xu, Y.; Hu, D.; Zhang, L.; Li, A.; Zhu, Q.; Liu, J. Tomato Yield Estimation Using an Improved Lightweight YOLO11n Network and an Optimized Region Tracking-Counting Method. *Agriculture* **2025**, *15*, 1353. [[CrossRef](#)]

66. Wang, J.; Zhang, Y.; Gu, R. Research status and prospects on plant canopy structure measurement using visual sensors based on three-dimensional reconstruction. *Agriculture* **2020**, *10*, 462. [\[CrossRef\]](#)
67. Zhang, S. Chat Demeter: A multi-agent system for plant disease diagnosis integrating CNN-transformer models. *Front. Plant Sci.* **2025**, *16*, 1695227. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Zhang, M.; Xu, Z.; Wang, P.; Li, R.; Wang, L.; Liu, Q.; Xu, J.; Zhang, X.; Wu, S. Agridoctor: A multimodal intelligent assistant for agriculture. In *Proceedings of the ICASSP 2026–2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: New York, NY, USA, 2026; pp. 2741–2745.
69. Roumeliotis, K.I.; Sapkota, R.; Karkee, M.; Tselikas, N.D. Agentic AI with orchestrator-agent trust: A modular visual classification framework with trust-aware orchestration and RAG-based reasoning. *IEEE Access* **2026**, *14*, 26965–26982. [\[CrossRef\]](#)
70. Lindia, P.; Cantini, R.; Bettucci, F.; Sartori, L.; Trunfio, P. Enhancing the Evaluation of Fault Detection Models in Smart Agriculture Using LLM Agents for Rule-Based Anomaly Generation. In *Proceedings of the GreenAI@ AI* IA*; CEUR-WS.org: Bonn, Germany, 2024; pp. 29–39.
71. Liu, Y.; Yan, H.; Zhang, C.; Zhang, J.; Wang, G.; Zhang, D.; Bao, R.; Han, Y. Application of Irrigation Decision-Making Methods for Smart Irrigation Systems: A Review. *Irrig. Drain.* **2026**, *75*, 1250–1264. [\[CrossRef\]](#)
72. Chen, Y.; Yu, Z.; Han, Z.; Sun, W.; He, L. A decision-making system for cotton irrigation based on reinforcement learning strategy. *Agronomy* **2023**, *14*, 11. [\[CrossRef\]](#)
73. Chen, Y.; Lin, M.; Yu, Z.; Sun, W.; Fu, W.; He, L. Enhancing cotton irrigation with distributional actor–critic reinforcement learning. *Agric. Water Manag.* **2025**, *307*, 109194. [\[CrossRef\]](#)
74. Zheng, W.; Lan, R.; Zhangzhong, L.; Yang, L.; Gao, L.; Yu, J. A hybrid approach for soil total nitrogen anomaly detection integrating machine learning and spatial statistics. *Agronomy* **2023**, *13*, 2669. [\[CrossRef\]](#)
75. Swati, N.L.P.; Gupta, S.V.; Duddela, N.S.; Parvathy, L.R. Agentic AI-driven autonomous decision support system for smart agriculture. *Sci. Rep.* **2026**, *16*, 9972. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Wei, L.; Yang, H.; Niu, Y.; Zhang, Y.; Xu, L.; Chai, X. Wheat biomass, yield, and straw-grain ratio estimation from multi-temporal UAV-based RGB and multispectral images. *Biosyst. Eng.* **2023**, *234*, 187–205. [\[CrossRef\]](#)
77. Chauhdary, J.N.; Li, H.; Jiang, Y.; Pan, X.; Hussain, Z.; Javaid, M.; Rizwan, M. Advances in sprinkler irrigation: A review in the context of precision irrigation for crop production. *Agronomy* **2023**, *14*, 47. [\[CrossRef\]](#)
78. Srinivasu, P.N.; Pavate, A.; JayaLakshmi, G.; Shafi, J.; Choi, J.; Ijaz, M.F. Agentic AI for smart and sustainable precision agriculture. *Front. Plant Sci.* **2025**, *16*, 1706428. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Singh, G.; Kumar, G.; Rajput, I.S.; Kumar, A.; Tyagi, S. Agentic AI with Large Language Models for Precision Farming: Advancing Sustainable Resource Optimization in Smart Agritech. *J. Recent Innov. Comput. Sci. Technol.* **2026**, *3*, 45–65. [\[CrossRef\]](#)
80. Zhang, R.; Zhu, H.; Chang, Q.; Mao, Q. A comprehensive review of digital twins technology in agriculture. *Agriculture* **2025**, *15*, 903. [\[CrossRef\]](#)
81. Cao, C.; Das, S.K.; Zhuang, J.; Davis, R. LLM-Orchestrated Digital Twins for Safe, Human-Centered Decision Support in Precision Agriculture. In *Proceedings of the First International Workshop on AI in Agriculture (Agri AI)*, Co-Located with AAAI 2026, Singapore, 26 January 2026.
82. Zhu, Y.; Zhang, S.; Tang, S.; Gao, Q. Research progress and applications of artificial intelligence in agricultural equipment. *Agriculture* **2025**, *15*, 1703. [\[CrossRef\]](#)
83. Jiang, L.; Xu, B.; Husnain, N.; Wang, Q. Overview of agricultural machinery automation technology for sustainable agriculture. *Agronomy* **2025**, *15*, 1471. [\[CrossRef\]](#)
84. Mohamed, T.M.K.; Gao, J.; Tunio, M. Development and experiment of the intelligent control system for rhizosphere temperature of aeroponic lettuce via the Internet of Things. *Int. J. Agric. Biol. Eng.* **2022**, *15*, 225–233. [\[CrossRef\]](#)
85. Qureshi, W.A.; Gao, J.; Elsherbiny, O.; Mosha, A.H.; Tunio, M.H.; Qureshi, J.A. Boosting aeroponic system development with plasma and high-efficiency tools: AI and IoT—A review. *Agronomy* **2025**, *15*, 546. [\[CrossRef\]](#)
86. Murad, M.; Ahmed, M.; Din, N.U.; Shahid, M.F.; Siddiqui, S.; Byers, D.; Tanveer, M.H.; Voicu, R.C. Agentic AI Framework to Automate Traditional Farming for Smart Agriculture. *AgriEngineering* **2026**, *8*, 8. [\[CrossRef\]](#)
87. Basir, M.S.; Pai, A.; Loo, L.Y.; Castiblanco, F.A.; Bailey, J.K.; Balmos, A.; Buckmaster, D.R.; Krogmeier, J.V. Meta Ag 2.0: A Framework for AI-Infused Contextual Agricultural Recordkeeping. In *Proceedings of the 2025 ASABE Annual International Meeting*, Toronto, ON, Canada, 13–16 July 2025; Paper No. 2500039.
88. Akbar, N.A.; Lenzitti, B.; Tegolo, D. A Novel Approach for Leveraging Agent-Based Experts on Large Language Models to Enable Data Sharing Among Heterogeneous IoT Devices in Agriculture. In *Proceedings of the International Conference of the Italian Association for Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany; pp. 12–22.
89. Jin, Y.; Liu, J.; Xu, Z.; Yuan, S.; Li, P.; Wang, J. Development status and trend of agricultural robot technology. *Int. J. Agric. Biol. Eng.* **2021**, *14*, 1–19. [\[CrossRef\]](#)
90. Luo, Y.; Wei, L.; Xu, L.; Zhang, Q.; Liu, J.; Cai, Q.; Zhang, W. Stereo-vision-based multi-crop harvesting edge detection for precise automatic steering of combine harvester. *Biosyst. Eng.* **2022**, *215*, 115–128. [\[CrossRef\]](#)

91. Zuzuarregui, M.; Carpin, S. As You Wish: Mission Planning with Formal Verification Using LLMs in Precision Agriculture. In *Proceedings of the 2026 IEEE International Conference on Robotics and Automation (ICRA)*; IEEE: New York, NY, USA, 2026.
92. Liu, E.; Yang, H.; Sharma, S.; van Leerdam, M.; Niu, P.; VandeHaar, M.; Hostens, M. Agents are all you need: Pioneering the use of agentic artificial intelligence to embrace large language models into dairy science. *J. Dairy Sci.* **2025**, *108*, 14038–14049. [[CrossRef](#)] [[PubMed](#)]
93. Yan, L.; Wang, H.; Tang, C.; Liu, H.; Sun, T.; Liu, L.; Guan, Y.; Jiang, J. Agrieval: A comprehensive chinese agricultural benchmark for large language models. *Proc. AAAI Conf. Artif. Intell.* **2026**, *40*, 34205–34213. [[CrossRef](#)]
94. da Silva, M.P.; Correa, S.P.; Oliveira, H.N.; Nunes, I.M.; Santos, J.A.D. Data-Centric AI for Tropical Agricultural Mapping: Challenges, Strategies and Scalable Solutions. *arXiv* **2025**, arXiv:2510.16207.
95. Lin, X.; Ning, Y.; Zhang, J.; Dong, Y.; Liu, Y.; Wu, Y.; Qi, X.; Sun, N.; Shang, Y.; Wang, K.; et al. Llm-based agents suffer from hallucinations: A survey of taxonomy, methods, and directions. *arXiv* **2025**, arXiv:2509.18970.
96. El-Dosuky, M. Real-time Framework for Interoperable Semantic-driven Internet-of-Things in Smart Agriculture. *arXiv* **2025**, arXiv:2510.05187.
97. Xiong, W.; Wang, K.; Song, Y.; Liu, H.; Zhou, S.; Peng, W.; Li, S. More Vulnerable than You Think: On the Stability of Tool-Integrated LLM Agents. *arXiv* **2025**, arXiv:2506.21967.
98. Chakraborty, T.; Ghosh, U.; Zhang, X.; Niloy, F.F.; Dong, Y.; Li, J.; Roy-Chowdhury, A.; Song, C. HEAL: An Empirical Study on Hallucinations in Embodied Agents Driven by Large Language Models. *Proc. Find. Assoc. Comput. Linguist. EMNLP* **2025**, *2025*, 21226–21243. [[CrossRef](#)]
99. Ma, Z.; Yang, S.; Li, J.; Qi, J. Research on slam localization algorithm for orchard dynamic vision based on YOLOD-SLAM2. *Agriculture* **2024**, *14*, 1622. [[CrossRef](#)]
100. Ma, Z.; Wang, X.; Chen, X.; Hu, B.; Li, J. Advances in Crop Row Detection for Agricultural Robots: Methods, Performance Indicators, and Scene Adaptability. *Agriculture* **2025**, *15*, 2151. [[CrossRef](#)]
101. Syed, T.N.; Zhou, J.; Lakhari, I.A.; Marinello, F.; Gemechu, T.T.; Rottok, L.T.; Jiang, Z. Enhancing autonomous orchard navigation: A real-time convolutional neural network-based obstacle classification system for distinguishing ‘real’ and ‘fake’ obstacles in agricultural robotics. *Agriculture* **2025**, *15*, 827. [[CrossRef](#)]
102. Khanal, A.; Tao, Y.; Zhou, J. Beyond pass@1: A reliability science framework for long-horizon LLM agents. *arXiv* **2026**, arXiv:2603.29231.
103. Wang, X.J.; Bai, H.; Sun, Y.; Wang, H.; Zhang, S.; Hu, W.; Schroder, M.; Mutlu, B.; Song, D.; Nowak, R.D. The Long-Horizon Task Mirage? Diagnosing Where and Why Agentic Systems Break. *arXiv* **2026**, arXiv:2604.11978.
104. Zaremehrjerdi, H.; Ganguly, S.; Rairdin, A.; Tranel, E.; Feuer, B.; Di Salvo, J.I.; Panthulugiri, S.; Pacin, H.T.; Moser, V.; Jones, S.; et al. Towards large reasoning models for agriculture. *arXiv* **2025**, arXiv:2505.19259.
105. Gauba, A.; Pi, I.; Man, Y.; Pang, Z.; Adve, V.; Wang, Y.X. Agmmu: A comprehensive agricultural multimodal understanding benchmark. *Adv. Neural Inf. Process. Syst.* **2026**, *38*, 84185–84217. [[CrossRef](#)]
106. Yehudai, A.; Eden, L.; Li, A.; Uziel, G.; Zhao, Y.; Bar-Haim, R.; Cohan, A.; Shmueli-Scheuer, M. Survey on evaluation of llm-based agents. *arXiv* **2025**, arXiv:2503.16416.
107. Zhan, Q.; Liang, Z.; Ying, Z.; Kang, D. Injecagent: Benchmarking indirect prompt injections in tool-integrated large language model agents. *Proc. Find. Assoc. Comput. Linguist. ACL* **2024**, *2024*, 10471–10506. [[CrossRef](#)]
108. Debenedetti, E.; Zhang, J.; Balunovic, M.; Beurer-Kellner, L.; Fischer, M.; Tramèr, F. Agentdojo: A dynamic environment to evaluate prompt injection attacks and defenses for llm agents. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 82895–82920. [[CrossRef](#)]
109. Chen, Z.; Xiang, Z.; Xiao, C.; Song, D.; Li, B. Agentpoison: Red-teaming llm agents via poisoning memory or knowledge bases. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 130185–130213. [[CrossRef](#)]
110. Zhang, H.; Huang, J.; Mei, K.; Yao, Y.; Wang, Z.; Zhan, C.; Wang, H.; Zhang, Y. Agent security bench (asb): Formalizing and benchmarking attacks and defenses in llm-based agents. *Proc. Int. Conf. Learn. Represent.* **2025**, *2025*, 35331–35366.
111. Al Asif, M.R.; Hasan, K.F.; Islam, M.Z.; Khondoker, R. STRIDE-based cyber security threat modeling for IoT-enabled precision agriculture systems. In *Proceedings of the 2021 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI)*; IEEE: New York, NY, USA, 2021; pp. 1–6.
112. Xu, Z.; Xu, J.; Zhang, M.; Wang, P.; Deng, C.; Liu, C.L. Multimodal Agricultural Agent Architecture (MA3): A New Paradigm for Intelligent Agricultural Decision-Making. *arXiv* **2025**, arXiv:2504.04789.
113. Yang, B.; Feng, L.; Chen, Y.; Zhang, Y.; Zhang, J.; Xu, X.; Aierken, N.; Li, S. AgriGPT-Omni: A Unified Speech–Vision–Text Framework for Multilingual Agricultural Intelligence. In *Proceedings of the ACM Web Conference 2026*, Dubai, United Arab Emirates, 29 June–3 July 2026; pp. 9800–9810.
114. Rana, S.; Hensel, O.; Nasirahmadi, A. Generative semantic reconstruction for annotation-ready vegetation priors in spectrally heterogeneous imagery. *MethodsX* **2026**, *16*, 103967. [[CrossRef](#)] [[PubMed](#)]
115. Zhang, L.; Zhang, B.; Zhang, H.; Yang, W.; Hu, X.; Cai, J.; Wu, C.; Wang, X. Multi-source feature fusion network for LAI estimation from UAV multispectral imagery. *Agronomy* **2025**, *15*, 988. [[CrossRef](#)]

116. Jiang, D.; Shen, Z.; Zheng, Q.; Zhang, T.; Xiang, W.; Jin, J. Farm-LightSeek: An edge-centric multimodal agricultural IoT data analytics framework with lightweight LLMs. *IEEE Internet Things Mag.* **2025**, *8*, 72–79. [[CrossRef](#)]
117. Khidirova, E.; Karakuş, O. UniCrop: A Universal, Multi-Source Data Engineering Pipeline for Scalable Crop Yield Prediction. *Appl. Sci.* **2026**, *16*, 4724. [[CrossRef](#)]
118. Huang, Z.; Yan, H.; Zhan, Q.; Yang, S.; Zhang, M.; Zhang, C.; Lei, Y.; Liu, Z.; Liu, Q.; Wang, Y. A survey on remote sensing foundation models: From vision to multimodality. *arXiv* **2025**, arXiv:2503.22081.
119. Zhou, C.; Chai, H.; Chen, W.; Guo, Z.; Shan, R.; Song, Y.; Xu, T.; Yang, Y.; Yu, A.; Zhang, W.; et al. Externalization in llm agents: A unified review of memory, skills, protocols and harness engineering. *arXiv* **2026**, arXiv:2604.08224.
120. Lou, X.; Lázaro-Gredilla, M.; Dedieu, A.; Wendelken, C.; Lehrach, W.; Murphy, K.P. Autoharness: Improving llm agents by automatically synthesizing a code harness. *arXiv* **2026**, arXiv:2603.03329.
121. Lin, X.; Liu, Y.; Chen, Y.; Wu, Y.; Ning, Y.; Liu, Y.; Sun, N.; Zhang, S.; Chong, B.; Zhou, C.; et al. SafeHarness: Lifecycle-Integrated Security Architecture for LLM-based Agent Deployment. *arXiv* **2026**, arXiv:2604.13630.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.