

REVIEW

Open access

Foundation Models and Multimodal AI in Agriculture: A Scoping Review of Emerging Applications, Data Requirements, and Deployment Risks

Kwame Mensah^{1*}, Kojo Asante¹, Linda Owusu²

Abstract

Foundation models and multimodal artificial intelligence are beginning to reshape agricultural AI by enabling systems that can integrate images, text, sensor streams, remote-sensing data, environmental records, and management information. These approaches differ from earlier narrow models because they can reuse learned representations across related tasks and support richer forms of agricultural interpretation. Their potential relevance spans crop monitoring, disease detection, advisory services, automation, forecasting, and agri-food decision support. Despite this promise, the transfer of large-scale AI systems into agriculture remains complex. Agricultural environments are biologically variable, spatially heterogeneous, seasonally dynamic, and shaped by management decisions that are often poorly captured in available datasets. As a result, questions remain about data quality, annotation adequacy, domain adaptation, validation, explainability, governance, and deployment risk. This scoping review maps emerging applications, data requirements, and deployment risks associated with foundation models and multimodal AI in agriculture and agri-food systems. It focuses on how the reviewed literature frames model architectures, data modalities, application domains, validation needs, and governance concerns. The purpose is not to estimate pooled effectiveness but to chart the evidence landscape and clarify where knowledge is advancing or remains unsettled. The review uses a scoping approach to organise literature around key concepts, model types, data inputs, agricultural tasks, implementation settings, and responsible deployment concerns. It charts relationships among remote-sensing foundation models, vision-language systems, large language models, generative AI, multimodal data fusion, benchmark datasets, and agricultural decision-support systems. It also identifies gaps in field validation, interoperability, low-resource contexts, participatory governance, and trustworthy deployment. The review contributes a structured synthesis of how foundation models and multimodal AI may extend agricultural sensing, prediction, automation, and decision support. It highlights that technical capability alone is insufficient for responsible agricultural use. Robust data infrastructures, context-aware adaptation, transparent validation, and accountable governance will be central to whether these systems support sustainable and equitable agri-food transformation.

Keywords Remote sensing, Agriculture, Generative AI, Agri-food systems, Foundation models, Multimodal AI

*Correspondence:

Kwame Mensah

kwame.mensah@gmail.com

¹ Department of Agricultural Engineering, Faculty of Engineering, University of Ghana, Accra, Ghana

² Department of Food Systems Technology, Faculty of Agriculture, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Introduction

Artificial intelligence in agriculture has developed from isolated prediction tools into broader data-intensive systems that connect sensing, analytics, and decision support. Smart farming research has shown how agricultural data infrastructures increasingly combine farm-management data, machine data, environmental observations, and platform-based analytics to support operational decisions [1]. Deep learning accelerated this shift by enabling learned visual and temporal representations for crop monitoring, disease recognition, yield prediction, livestock analysis, and agricultural automation [2]. These developments created the technical foundation for newer interest in foundation models and multimodal AI.

However, agricultural AI is not only a computational transition. Digital agriculture is embedded in advisory systems, platform markets, labour relations, institutional settings, farmer knowledge, and uneven access to infrastructure, which means that model deployment must be understood as a socio-technical change process [3]. Earlier machine-learning applications in agriculture already covered crop management, livestock monitoring, irrigation, disease detection, and yield estimation, but they often remained task-specific and dependent on carefully bounded datasets [4]. Foundation models and multimodal systems now extend this trajectory by promising more reusable representations across agricultural tasks while also amplifying long-standing concerns about data quality, context, and accountability.

The emergence of large-scale pretrained models raises a central scoping question: what exactly is being transformed in agricultural AI? In principle, these systems may integrate satellite imagery, drone imagery, field photographs, sensor streams, soil and weather records, textual advice, and management information into more context-aware decision-support workflows. Yet agriculture is marked by regional crop diversity, seasonal instability, variable management practices, and uneven data infrastructures, which limits assumptions that models trained in one setting will generalise safely elsewhere. A scoping review is therefore appropriate because the field requires mapping of concepts, applications, data needs, risks, and gaps rather than a narrow assessment of intervention effectiveness.

This article maps the reviewed literature on foundation models and multimodal AI in agriculture and agri-food systems. It first clarifies how foundation models are being defined in agricultural AI, then describes the scoping review mapping approach, charts data requirements, maps emerging applications, analyses deployment risks, synthesises cross-cutting patterns, and proposes a research and governance agenda. The review treats foundation models, multimodal learning, language models, remote-sensing models, generative AI, and data-fusion systems as related but distinct evidence streams. This structure supports a broad evidence map while preserving careful alignment between cited sources and the claims they substantiate.

Foundation Models in Agricultural AI

Foundation models in agricultural AI can be understood as large-scale pretrained models whose learned representations can be adapted to multiple downstream tasks. In smart agriculture, these models are discussed as a way to move from single-task prediction toward reusable representations for sensing, classification, segmentation, monitoring, and decision support [5]. Their relevance derives from training approaches such as transfer learning, self-supervised learning, multimodal representation learning, and large-scale pretraining on heterogeneous datasets. In this sense, foundation models are not a single architecture but a family of approaches that seek greater adaptability across agricultural contexts.

The agricultural foundation-model literature also shows that these systems are being explored across crop monitoring, disease diagnosis, segmentation, remote sensing, advisory applications, and robotics. A comprehensive review of foundation models in agriculture frames them as tools that may support more flexible agricultural AI pipelines, while also emphasising the need for task-specific adaptation and validation [6]. This matters because agricultural data are shaped by biological growth, climate variability, soil heterogeneity, farm management, and sensor conditions. General-purpose model capacity therefore needs to be translated into domain-specific reliability before it can be used in production settings.

Vision-language models are especially relevant because many agricultural tasks require linking visual evidence with language-based interpretation. Large vision and language models may allow agricultural systems to connect crop images,

disease symptoms, field observations, user questions, and explanatory outputs in a single interaction loop [7]. This capability could support plant-health diagnosis, advisory communication, robotic perception, and decision support, but the agricultural meaning of an image depends on crop stage, local conditions, management history, and expert interpretation. Vision-language systems therefore require careful grounding in agronomic context rather than reliance on generic visual-language associations.

Multimodal large language models extend this agenda by combining text, images, sensor information, and contextual knowledge in more interactive agricultural AI systems. The literature on multimodal LLMs in agriculture identifies opportunities for advisory services, monitoring, automation, and decision support, while also noting challenges associated with multimodal alignment, evaluation, hallucination, and deployment readiness [8]. These systems may become interfaces between farmers, agronomists, machines, and data platforms, but they also introduce new forms of dependence on model outputs. Agricultural foundation-model deployment should therefore be understood as an adaptation, validation, and governance problem, not merely as a model-scaling achievement.

Review Scope and Mapping Approach

The review scope covers agricultural and agri-food AI literature that directly informs foundation models, multimodal systems, large-scale pretrained models, language models, generative AI, remote-sensing representation learning, data fusion, and related decision-support architectures. Large language models are included because agricultural intelligence increasingly involves knowledge retrieval, natural-language explanation, advisory communication, and decision support rather than only image-based detection [9]. Agricultural question-answering systems are treated as a distinct application category because they translate agricultural knowledge into farmer-facing or advisor-facing responses [10]. This scope allows the review to connect model architectures with practical agricultural decision contexts.

The mapping approach charts concepts, model types, data modalities, application areas, validation challenges, deployment risks, and governance concerns. Agricultural data fusion is central to this logic because multimodal systems often aim to reduce ambiguity, close data gaps, and combine complementary evidence from images, sensors, environmental records, and textual sources [11]. Remote-sensing data fusion is also included as a key evidence stream because precision agriculture increasingly relies on satellite, drone, temporal, and spectral information that can be integrated through transformer-based or foundation-model approaches [12]. The mapping therefore follows the structure of the evidence landscape rather than imposing a single hierarchy of model types.

Eligibility was defined in conceptual terms: the literature needed clear relevance to agricultural or agri-food AI and to foundation-model or multimodal-AI development, application, validation, or deployment. Articles were charted according to the agricultural context addressed, the AI method described, the data modality used, the application domain mapped, and the risk or validation issue raised. This scoping logic avoided claims of pooled effectiveness and instead focused on how the field is being organised, what concepts are emerging, and where uncertainty remains. It also distinguished between directly agriculture-specific foundation-model studies and broader remote-sensing or governance work that requires careful wording when applied to agricultural contexts. **Figure 1** visualises the scoping evidence map by linking foundation-model types, multimodal agricultural data inputs, agri-food application domains, validation requirements, and deployment-risk categories.

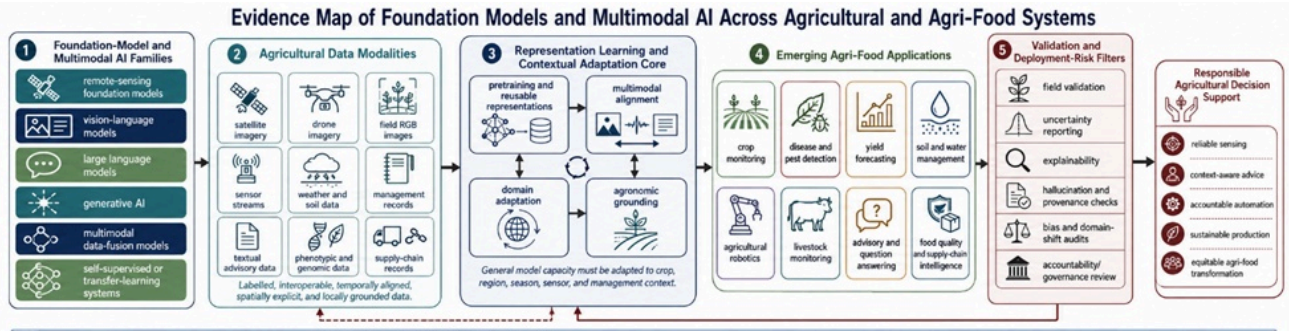


Figure 1. Evidence Map of Foundation Models and Multimodal AI Across Agricultural and Agri-Food Systems

The evidence-charting process was designed to preserve alignment between each source and the manuscript claim it supports. The resulting map organises the literature into foundation-model concepts, multimodal data requirements, application domains, deployment risks, synthesis themes, and governance priorities. **Table 1** summarises the scoping review mapping approach, eligibility logic, and evidence-charting dimensions used to organise the literature. This framework ensures that the review remains a scoping evidence map rather than a systematic effectiveness assessment or a generic narrative essay.

Table 1. Scoping Review Mapping Framework for Foundation Models and Multimodal AI in Agriculture: Review Scope, Eligibility Logic, Evidence-Charting Categories, and Mapping Dimensions

Mapping Component	Scoping Review Logic	Evidence-Charting Categories	Inclusion Emphasis	Exclusion Emphasis
Review purpose	Map the breadth, range, and structure of emerging evidence	Concepts, model families, applications, data needs, risks, and gaps	Literature that clarifies how foundation models or multimodal AI relate to agricultural or agri-food systems	Literature that only offers generic AI discussion without agricultural or agri-food relevance
Model scope	Capture related large-scale and multimodal AI approaches	Foundation models, LLMs, vision-language models, generative AI, self-supervised learning, transfer learning, data fusion, remote-sensing models	Studies with clear relevance to agricultural AI development, deployment, or evaluation	Studies that discuss agriculture without relevant AI model approaches
Agricultural scope	Cover farm-level and agri-food-system contexts	Crop monitoring, remote sensing, advisory systems, robotics, yield forecasting, water and soil management, food-system intelligence	Work addressing crops, farms, production systems, environmental monitoring, or agri-food decision support	Work outside agricultural or agri-food contexts
Data-modality scope	Identify data infrastructures required for multimodal AI	Satellite data, drone imagery, field images, sensors, weather, soil, management records, text, benchmarks, geospatial labels	Studies that describe data inputs, fusion needs, benchmarks, or interoperability challenges	Studies with insufficient information on data requirements
Risk and validation	Chart deployment concerns beyond	Domain shift, field validation, explainability, bias,	Work identifying validation, ethical, legal,	Risk claims that are superficial or

scope	model accuracy	hallucination, privacy, accountability, governance, trust	social, or governance issues in agricultural AI	disconnected from agricultural deployment
Synthesis logic	Organise evidence by application and deployment readiness	Opportunities, unresolved gaps, governance needs, responsible deployment priorities	Sources that support specific claims in the scoping map	Sources that duplicate stronger evidence or cannot support a planned claim

Multimodal Data Requirements

Foundation models and multimodal AI in agriculture depend on curated datasets that represent both biological variability and field-level operational conditions. Agricultural vision benchmarks illustrate the importance of high-resolution labelled imagery, consistent annotation protocols, and task-specific evaluation for crop detection and phenotyping [13]. Geospatial crop datasets further show that model development requires harmonised labels, spatially aligned records, and interoperable data structures across regions and production systems [14]. These requirements become more demanding when models are expected to transfer across crops, climates, sensors, and management contexts.

Temporal and few-shot learning tasks add another layer of data complexity. Crop-type classification benchmarks show that agricultural models need time-series data capable of capturing phenological stages, seasonal variation, and limited-labelled-data scenarios [15]. Yield forecasting benchmarks similarly demonstrate that prediction depends on integrating production records, weather, soil, remote-sensing indicators, and environmental covariates into coherent data structures [16]. For foundation models, these examples highlight that agricultural data quality is not only about volume but also about temporal coherence, spatial representativeness, and agronomic meaning.

Multimodal crop-stress studies show that combining data streams can improve interpretive richness while also creating validation challenges. In sweet potato water-stress classification, multimodal fusion and explainable AI were used to connect different sensing inputs with interpretable stress-detection outputs [17]. This kind of evidence suggests that multimodal models require modality-level checks to determine whether predictions are driven by agronomically plausible signals. Without such checks, a model may appear accurate while relying on unstable correlations, sensor artefacts, or context-specific cues.

The mapped evidence indicates that agricultural foundation models require data infrastructures that are labelled, interoperable, temporally aligned, spatially explicit, and sensitive to local production conditions. Image datasets, harmonised crop labels, time-series benchmarks, yield data, and multimodal stress-monitoring studies each reveal different parts of this infrastructure challenge [13-17]. Table 2 classifies the multimodal data requirements associated with foundation models and multimodal AI in agricultural contexts. The table emphasises that data quality risks are inseparable from model validation and responsible deployment.

Table 2. Multimodal Data Requirements for Agricultural Foundation Models: Data Modalities, Annotation Needs, Spatial and Temporal Resolution, Integration Challenges, and Data-Quality Risks

Data Modality	Agricultural Use	Annotation or Metadata Needs	Spatial and Temporal Requirements	Integration Challenges	Data-Quality Risks
Satellite imagery	Crop mapping, phenology	Crop labels, geolocation,	Field-to-regional scale; repeated	Sensor harmonisation, cloud	Regional bias, missing

	monitoring, yield forecasting, drought assessment	acquisition date, sensor metadata, cloud masks	seasonal observations	contamination, mixed pixels, cross-region transfer	observations, coarse resolution, domain shift
Drone imagery	Field scouting, canopy assessment, weed detection, plant counting	Plot boundaries, flight parameters, crop stage, disease or stress labels	High spatial resolution; campaign-specific or repeated field surveys	Variable lighting, flight altitude, orthomosaic quality, field calibration	Overfitting to local fields, inconsistent annotation, limited temporal coverage
Field RGB imagery	Disease detection, pest monitoring, plant organ detection, phenotyping	Bounding boxes, segmentation masks, class labels, growth stage, cultivar	Plant-to-plot scale; growth-stage-dependent sampling	Background variation, occlusion, device differences, class imbalance	Label noise, sampling bias, low representation of rare symptoms
Sensor streams	Irrigation, livestock monitoring, greenhouse control, machinery analytics	Sensor type, calibration, sampling frequency, location, maintenance records	Continuous or near-continuous time series	Missing values, sensor drift, heterogeneous sampling intervals	Drift, outages, calibration errors, unreliable temporal alignment
Weather and soil data	Yield forecasting, risk prediction, water management, climate-smart decision support	Station metadata, soil properties, interpolation methods, uncertainty estimates	Local-to-regional scale; historical and forecast time horizons	Spatial interpolation, scale mismatch, incomplete station coverage	Sparse observations, regional inequity, uncertainty propagation
Management records	Fertilisation, irrigation, crop rotation, input optimisation, advisory systems	Operation dates, input rates, field identifiers, machinery logs	Field-level records across seasons	Proprietary formats, incomplete farmer records, inconsistent terminology	Missing management context, privacy concerns, ownership disputes
Textual advisory data	LLM-based decision support, question answering, agronomic knowledge retrieval	Source provenance, expert validation, local relevance, language, update status	Region- and crop-specific knowledge over changing seasons	Ambiguous terminology, multilingual variation, outdated recommendations	Hallucination risk, weak provenance, poor fit to local practice
Phenotypic and genomic data	Breeding, trait prediction, stress response modelling, cultivar recommendation	Trait protocols, genotype identifiers, trial design, environmental metadata	Plot-to-programme scale; multi-environment trials	Linking genotype, phenotype, environment, and management records	Unbalanced trials, limited population diversity, weak interoperability

Supply-chain data	Traceability, quality assessment, logistics, food-system monitoring	Product identifiers, transaction records, quality metrics, custody events	Product-to-network scale; real-time or batch updates	Data-sharing barriers, platform lock-in, confidentiality	Fragmentation, strategic reporting, weak auditability
-------------------	---	---	--	--	---

Emerging Applications across Agri-Food Systems

Remote sensing is one of the clearest application areas for foundation models in agriculture. CropSTS illustrates how a remote-sensing foundation model can use spatiotemporal representation learning for cropland classification, supporting the emergence of foundation-model approaches in agricultural land monitoring [18]. SpectralGPT shows that spectral remote-sensing foundation models can learn reusable representations from spectral data that may be adapted for crop and land-surface monitoring when task-specific agricultural validation is performed [19]. Together, these studies position remote sensing as a major entry point for agricultural foundation-model research.

Language-based decision support is another emerging application stream. Agricultural question-answering systems based on large language models have been developed and evaluated as tools for agricultural knowledge access and advisory support [20]. Evaluations of ChatGPT responses to farmers' rice-cultivation questions show that farmer-facing language-model outputs must be judged against local agronomic context, production constraints, and domain expectations [21]. These studies suggest that agricultural LLMs may support advisory workflows, but their reliability depends on grounding, expert review, and context-specific validation.

Foundation-model ideas are also beginning to interact with optimisation and control. Work integrating reinforcement learning and large language models for crop production management shows how language-based reasoning may be combined with process optimisation in agricultural decision workflows [22]. This application direction moves beyond static classification toward planning, control, and management support. It also raises demanding validation questions because crop-management recommendations can influence input use, costs, labour, environmental outcomes, and production risk.

Across the mapped literature, emerging applications include crop monitoring, cropland classification, disease and stress interpretation, yield forecasting, advisory systems, decision support, and production-process optimisation. These applications differ in data intensity, user interaction, deployment setting, and tolerance for error, so they should not be treated as equivalent use cases [18-22]. **Table 3** maps emerging applications of foundation models and multimodal AI across agri-food systems. The table distinguishes between model functions, data inputs, potential users, and deployment settings to clarify how application readiness varies across domains.

Table 3. Emerging Applications of Foundation Models and Multimodal AI Across Agri-Food Systems: Tasks, Data Inputs, Model Functions, Potential Users, and Deployment Settings

Application Area	Typical Tasks	Data Inputs	Model Functions	Potential Users	Deployment Settings
Crop monitoring	Crop classification, phenology tracking, stress detection, biomass estimation	Satellite imagery, drone imagery, field images, weather, crop calendars	Representation learning, segmentation, temporal classification, anomaly detection	Farmers, agronomists, insurers, public agencies	Field platforms, remote-sensing dashboards, advisory services

Disease and pest detection	Symptom recognition, severity estimation, pest scouting, treatment triage	Field images, expert labels, textual descriptions, local weather, management history	Vision-language diagnosis, image classification, retrieval-augmented advisory support	Farmers, extension agents, crop-protection advisors	Mobile apps, extension workflows, decision-support tools
Yield forecasting	Seasonal prediction, sub-national forecasting, risk mapping	Yield records, weather, soil, satellite time series, management data	Multimodal forecasting, uncertainty estimation, transfer learning	Farmers, governments, traders, food-security organisations	Forecasting platforms, risk dashboards, planning systems
Soil and water management	Irrigation scheduling, water-stress detection, soil-condition assessment	Soil sensors, weather data, thermal imagery, crop stress indicators, management records	Data fusion, control support, stress classification, scenario recommendation	Farmers, irrigation managers, agronomists	Farm decision systems, irrigation controllers, climate-smart advisory tools
Agricultural robotics	Navigation, perception, manipulation, harvesting, targeted intervention	RGB images, depth sensors, LiDAR, text instructions, machine telemetry	Vision-language grounding, perception, task planning, control integration	Robotics developers, growers, machinery operators	Greenhouses, orchards, fields, controlled-environment agriculture
Livestock monitoring	Behaviour recognition, welfare monitoring, disease alerts, productivity assessment	Video, wearable sensors, environmental sensors, farm records	Multimodal pattern recognition, anomaly detection, decision support	Livestock farmers, veterinarians, welfare auditors	Barn monitoring systems, farm-management platforms
Advisory and question answering	Farmer questions, agronomic guidance, knowledge retrieval, explanation	Text, images, local recommendations, crop calendars, expert knowledge bases	LLM response generation, retrieval, summarisation, explanation	Farmers, extension agents, advisors, cooperatives	Chatbots, mobile advisory systems, extension platforms
Food quality and supply chains	Quality grading, traceability, logistics monitoring, sustainability reporting	Product images, sensor data, transaction records, traceability logs, quality measurements	Multimodal classification, anomaly detection, document reasoning, risk flagging	Processors, retailers, regulators, logistics managers	Processing facilities, traceability platforms, quality-control systems

Deployment Risks and Validation Challenges

Deployment risk in agricultural AI cannot be reduced to model accuracy. Responsible AI work in agriculture argues that risks and externalities must be understood systemically, including effects on farming practices, institutions, labour, knowledge systems, environmental outcomes, and market relations [23]. Ethical guidance for digital agriculture further emphasises fairness, transparency, privacy, accountability, and robustness as safeguards that should shape agricultural AI deployment [24]. These concerns are especially important for foundation models because their scale, opacity, and general-purpose interfaces may obscure context-specific failure modes.

Risk management literature identifies agricultural AI concerns that include safety, autonomy, accountability, labour displacement, environmental consequences, and unclear responsibility when automated systems influence decisions [25]. Legal and ethical analysis of AI in climate-smart agriculture similarly highlights accountability, data governance, fairness, and liability challenges in sustainability-oriented deployments [26]. These issues become sharper when multimodal systems produce recommendations that combine remote sensing, text, sensor data, and farm records. A system may appear technically sophisticated while still being poorly governed, insufficiently validated, or misaligned with farmer needs.

Validation is particularly difficult because agricultural conditions vary across crops, cultivars, soils, climates, seasons, sensor types, and management regimes. A model that performs well on a benchmark or in a narrow experimental setting may fail under different phenological stages, rare stress events, mixed cropping systems, or low-resource environments. Responsible deployment therefore requires field validation across production contexts, uncertainty communication, explainability checks, and attention to who is affected by model errors [23–26]. For language-based systems, validation must also address hallucination, outdated knowledge, weak provenance, and overconfident advice.

The mapped evidence suggests that agricultural foundation models require deployment protocols that connect technical evaluation with governance review. Domain shift, regional bias, data ownership, cybersecurity, auditability, liability, environmental cost, and platform power should be assessed before systems are scaled into advisory, monitoring, or control settings [23–26]. Because agricultural decisions affect livelihoods, ecosystems, food security, and public trust, model evaluation should include reliability, safety, accountability, and social legitimacy. Deployment readiness is therefore a multidimensional judgement rather than a single performance threshold.

Scoping Synthesis

The mapped evidence shows that trust is a cross-cutting synthesis theme in agricultural AI. Work on trust in farming AI highlights that adoption depends not only on technical performance but also on transparency, farmer experience, institutional relationships, usability, accountability, and confidence in data practices [27]. Literature mapping of social and ethical impacts further shows that agricultural AI raises recurring concerns about power, justice, labour, sustainability, and responsibility [28]. These themes suggest that foundation models should be evaluated as parts of agri-food systems, not only as software artefacts.

Generative AI extends the application landscape by enabling text generation, synthetic data, advisory tools, knowledge interfaces, and multimodal content generation in digital agri-food contexts [29]. Food-system AI initiatives also point toward interdisciplinary infrastructures that connect data, modelling, domain expertise, institutions, and deployment environments across the agri-food chain [30]. These developments indicate that foundation models and multimodal AI may become a new layer of agricultural intelligence linking sensing, reasoning, communication, and decision support. However, this layer will only be useful if it is grounded in reliable data and accountable use.

Several patterns emerge from the scoping synthesis. Momentum is strongest in remote sensing, vision-based monitoring, benchmark development, and data-fusion tasks, while language-model and generative-AI applications are expanding rapidly into advisory and decision-support spaces [27–30]. At the same time, evidence remains uneven across livestock

systems, low-resource agricultural regions, smallholder contexts, participatory validation, and long-term deployment studies. The field is therefore simultaneously technically dynamic and institutionally underdeveloped.

The synthesis also shows that opportunities and risks are tightly linked. Foundation models may improve transfer, reduce task-specific engineering, and support multimodal reasoning, but they may also intensify opacity, bias, data dependence, and platform concentration if governance is weak [27-30]. **Table 4** synthesises the opportunities, deployment risks, validation gaps, and governance priorities identified across the scoping review. This synthesis frames responsible deployment as a combined problem of data infrastructure, model adaptation, stakeholder trust, and public-interest governance.

Table 4. Scoping Synthesis of Foundation Models and Multimodal AI in Agriculture: Opportunities, Deployment Risks, Validation Gaps, and Governance Priorities

Synthesis Domain	Opportunity Mapped in the Literature	Deployment Risk	Validation Gap	Governance Priority
Remote-sensing foundation models	Reusable representations for crop mapping, phenology, and land monitoring	Domain shift across regions, seasons, sensors, and crop systems	Limited field validation beyond benchmark settings	Transparent benchmarking, uncertainty reporting, regional performance audits
Multimodal data fusion	Integration of images, sensors, weather, soil, and management records	Spurious correlations, missing data, scale mismatch, weak interoperability	Limited modality-level validation and causal interpretation	Data standards, metadata requirements, interoperable farm-data infrastructures
Vision-language systems	Linking images with explanations, diagnosis, and interactive advisory support	Misinterpretation of symptoms, overconfident explanations, weak local grounding	Limited evaluation with agronomists and farmers in real-world settings	Human-in-the-loop review, expert validation, local adaptation protocols
Large language models	Knowledge retrieval, question answering, summarisation, and advisory communication	Hallucination, outdated recommendations, poor provenance, language inequity	Inadequate crop-, region-, and user-specific evaluation	Retrieval governance, source transparency, advisory liability rules
Generative AI	Synthetic data, decision-support content, training materials, scenario generation	Fabricated outputs, bias amplification, misuse, weak traceability	Limited assessment of generated content quality and agronomic validity	Content provenance, audit trails, safeguards against misleading outputs
Agricultural robotics	Multimodal perception, task planning, navigation, and automation support	Safety risks, equipment damage, labour disruption, liability uncertainty	Limited validation under diverse field conditions and edge cases	Safety certification, operator oversight, accountability frameworks
Food-system intelligence	Traceability, quality monitoring, sustainability reporting, logistics support	Data concentration, surveillance, confidentiality risks, market power	Weak cross-chain evaluation and limited transparency	Participatory data governance, data-sharing rules, public-interest safeguards

Responsible deployment	More adaptive, context-aware, and scalable agricultural AI	Trust erosion, exclusion of low-resource users, inequitable benefits	Few long-term studies of social, economic, and environmental outcomes	Equity assessment, participatory design, farmer-centred governance
------------------------	--	--	---	--

Research and Governance Agenda

A research agenda for agricultural foundation models should begin with structured ethical, legal, and social assessment. An ELSA assessment tool for agri-food AI shows how ethical, legal, and social dimensions can be embedded into agricultural AI evaluation rather than added after technical development [31]. Data-sharing policy research further indicates that trusted agricultural AI requires attention to stakeholder rights, data-access rules, governance arrangements, and institutional legitimacy [32]. Future research should therefore combine benchmark performance with assessment of accountability, transparency, rights, and public value.

Participatory data governance is also essential. Data-justice work with agri-food movements shows that agricultural data governance should be shaped from the ground up, including attention to farmer agency, collective rights, power asymmetries, and community priorities [33]. This perspective is important for foundation models because large-scale data aggregation can concentrate value in platforms while shifting risk to producers and local communities. Responsible agricultural AI should therefore include participatory evaluation, transparent data-use agreements, and mechanisms for contesting harmful or inappropriate model outputs.

Technically, the research agenda should prioritise benchmark datasets, multimodal standards, domain adaptation, uncertainty estimation, explainability, and field validation. These priorities align with the need for structured assessment, trusted data-sharing, and data-justice safeguards across agri-food AI development [31-33]. Models should be evaluated across crops, regions, seasons, management systems, sensor conditions, and user groups rather than only in curated benchmark environments. Research should also address privacy-preserving learning, low-resource agricultural settings, multilingual advisory systems, and environmental costs of large-scale model training and inference. **Figure 2** presents a responsible deployment-readiness framework showing how technical validation, data governance, participatory oversight, legal accountability, and sustainability assessment should be combined before foundation models are scaled across agricultural and agri-food systems.

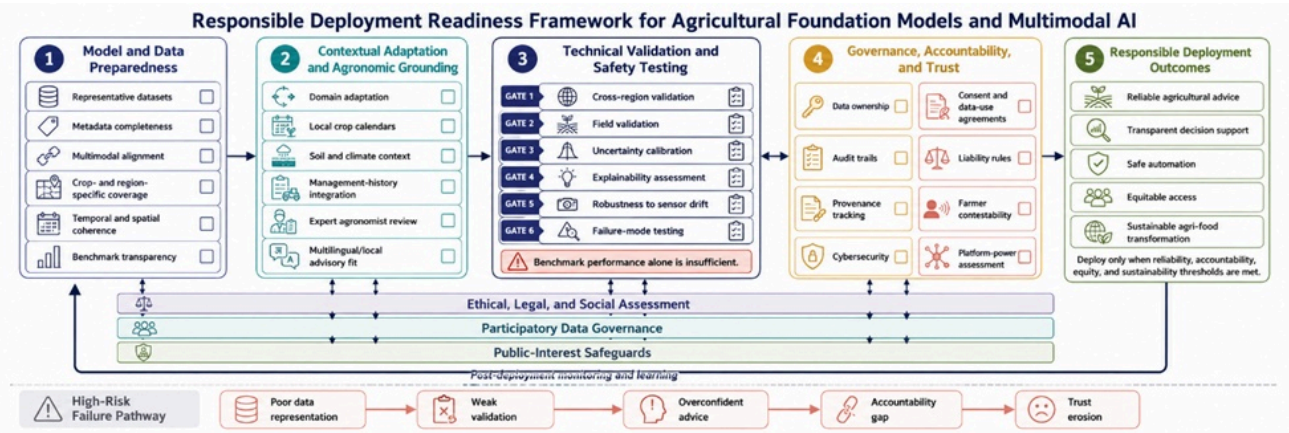


Figure 2. Responsible Deployment Readiness Framework for Agricultural Foundation Models and Multimodal AI

Governance agendas must address data ownership, platform power, model accountability, liability, auditability, transparency, and farmer-centred design. Agricultural foundation models should be assessed not only by predictive

accuracy or generative fluency but also by reliability, safety, equity, sustainability, and trustworthiness [31-33]. This requires collaboration among farmers, agronomists, computer scientists, extension services, regulators, cooperatives, civil society, and food-system actors. The goal is not simply to deploy larger models but to develop agricultural AI systems that are technically robust, context-aware, socially accountable, and aligned with sustainable agri-food transformation.

Conclusion

This scoping review mapped the emergence of foundation models and multimodal AI in agriculture and agri-food systems. It charted how large-scale pretrained models, multimodal data fusion, remote-sensing foundation models, vision-language systems, generative AI, and large language models are being positioned across agricultural sensing, prediction, automation, advisory services, and decision support. The review also clarified that these approaches remain unevenly developed across application domains, data infrastructures, validation settings, and governance arrangements.

The synthesis shows that foundation models and multimodal AI may expand the capabilities of agricultural AI, but their value depends on much more than scale or architectural novelty. High-quality datasets, interoperable data standards, contextual adaptation, robust field validation, interpretable outputs, uncertainty communication, and trusted data governance are central requirements. Without these foundations, advanced models may reproduce bias, generate unreliable advice, obscure accountability, or intensify inequalities across agri-food systems.

Future agricultural foundation models should be developed as responsible socio-technical systems rather than isolated technical products. They should be evaluated in relation to reliability, safety, equity, sustainability, farmer agency, and public-interest outcomes. A context-aware and governance-centred approach can help ensure that foundation models and multimodal AI support sustainable agri-food transformation rather than simply accelerating automation without accountability.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 20 Jul 2025

Revised: 17 Aug 2025

Accepted: 24 Sep 2025

Published online: 15 March 2026

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Wolfert S, Ge L, Verdouw C, Bogaardt MJ. Big data in smart farming – A review. *Agric Syst*. 2017;153:69-80.
<https://doi.org/10.1016/j.agry.2017.01.023>.
- Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: A survey. *Comput Electron Agric*. 2018;147:70-90.
<https://doi.org/10.1016/j.compag.2018.02.016>.
- Klerkx L, Jakku E, Labarthe P. A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda. *NJAS Wagening J Life Sci*. 2019;90-91:100315.
<https://doi.org/10.1016/j.njas.2019.100315>.
- Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: A review. *Sensors (Basel)*. 2018;18(8):2674.
<https://doi.org/10.3390/s18082674>.
- Li J, Xu M, Xiang L, Chen D, Zhuang W, Yin X, et al. Foundation models in smart agriculture: Basics, opportunities, and challenges. *Comput Electron Agric*. 2024;222:109032.
<https://doi.org/10.1016/j.compag.2024.109032>.
- Yin S, Xi Y, Zhang X, Sun C, Mao Q. Foundation models in agriculture: A comprehensive review. *Agriculture (Basel)*. 2025;15(8):847.
<https://doi.org/10.3390/agriculture15080847>.
- Zhu H, Qin S, Su M, Lin C, Li A, Gao J. Harnessing large vision and language models in agriculture: A review. *Front Plant Sci*. 2025;16:1579355.
<https://doi.org/10.3389/fpls.2025.1579355>.
- Sapkota R, Qureshi R, Hadi MU, Hassan SZ, Sadak F, Shoman M, et al. Multi-modal LLMs in agriculture: A comprehensive review. *IEEE Trans Autom Sci Eng*. 2025;22:22510-40.
<https://doi.org/10.1109/TASE.2025.3612154>.
- Li H, Wu H, Li Q, Zhao C. A review on enhancing agricultural intelligence with large language models. *Artif Intell Agric*. 2025;15:671-85.
<https://doi.org/10.1016/j.aiia.2025.05.006>.
- Yang T, Mei Y, Xu L, Yu H, Chen Y. Application of question answering systems for intelligent agriculture production and sustainable management: A review. *Resour Conserv Recycl*. 2024;204:107497.
<https://doi.org/10.1016/j.resconrec.2024.107497>.
- Barbedo JGA. Data fusion in agriculture: Resolving ambiguities and closing data gaps. *Sensors (Basel)*. 2022;22(6):2285.
<https://doi.org/10.3390/s22062285>.
- Saki M, Keshavarz R, Franklin DR, Abolhasan M, Lipman J, Shariati N. A data-driven review of remote sensing-based data fusion in precision agriculture from foundational to transformer-based techniques. *IEEE Access*. 2025;13:166188-209.
<https://doi.org/10.1109/ACCESS.2025.3610649>.
- David E, Madec S, Sadeghi-Tehran P, Aasen H, Zheng B, Liu S, et al. Global Wheat Head Detection (GWHD) Dataset: A large and diverse dataset of high-resolution RGB-labelled images to develop and benchmark wheat head detection methods. *Plant*

Phenomics. 2020;2020:3521852.

<https://doi.org/10.34133/2020/3521852>.

Schneider M, Schelte T, Schmitz F, Körner M. EuroCrops: The largest harmonized open crop dataset across the European Union. *Sci Data*. 2023;10(1):612.

<https://doi.org/10.1038/s41597-023-02517-0>.

Reuss J, Macdonald J, Becker S, Richter L, Körner M. The EuroCropsML time series benchmark dataset for few-shot crop type classification in Europe. *Sci Data*. 2025;12(1):664.

<https://doi.org/10.1038/s41597-025-04952-7>.

Kallenberg M, Paudel D, Ofori-Ampofo S, Baja H, van Bree R, Potze A, et al. CY-Bench: A comprehensive benchmark dataset for sub-national crop yield forecasting. *Earth Syst Sci Data*. 2026;18(6):3997-4018.

<https://doi.org/10.5194/essd-18-3997-2026>.

Choi JW, Cho SB, Hidayat MS, Hwang WH, Cho YS, Lee H, et al. Application of multimodal data fusion and explainable AI for classifying water stress in sweet potatoes. *Front Plant Sci*. 2025;16:1681915.

<https://doi.org/10.3389/fpls.2025.1681915>.

Yan J, Gu X, Chen Y. CropSTS: A remote sensing foundation model for cropland classification with decoupled spatiotemporal attention. *Remote Sens (Basel)*. 2025;17(14):2481.

<https://doi.org/10.3390/rs17142481>.

Hong D, Zhang B, Li X, Li Y, Li C, Yao J, et al. SpectralGPT: Spectral remote sensing foundation model. *IEEE Trans Pattern Anal Mach Intell*. 2024;46(8):5227-44.

<https://doi.org/10.1109/TPAMI.2024.3362475>.

Eldem A, Eldem H. The development and evaluation of agricultural question-answering systems based on large language models. *Sci Rep*. 2026;16(1):5357.

<https://doi.org/10.1038/s41598-026-35003-9>.

Ibrahim A, Senthilkumar K, Saito K. Evaluating responses by ChatGPT to farmers' questions on irrigated lowland rice cultivation in Nigeria. *Sci Rep*. 2024;14(1):3407.

<https://doi.org/10.1038/s41598-024-53916-1>.

Chen D, Huang Y. Integrating reinforcement learning and large language models for crop production process management optimization and control through a new knowledge-based deep learning paradigm. *Comput Electron Agric*. 2025;232:110028.

<https://doi.org/10.1016/j.compag.2025.110028>.

Tzachor A, Devare M, King B, Avin S, Ó hÉigeartaigh S. Responsible artificial intelligence in agriculture requires systemic understanding of risks and externalities. *Nat Mach Intell*. 2022;4(2):104-9.

<https://doi.org/10.1038/s42256-022-00440-4>.

Dara R, Hazrati Fard SM, Kaur J. Recommendations for ethical and responsible use of artificial intelligence in digital agriculture. *Front Artif Intell*. 2022;5:884192.

<https://doi.org/10.3389/frai.2022.884192>.

Sparrow R, Howard M, Degeling C. Managing the risks of artificial intelligence in agriculture. *NJAS Impact Agric Life Sci*. 2021;93(1):172-96.

<https://doi.org/10.1080/27685241.2021.2008777>.

Uddin M, Chowdhury A, Kabir MA. Legal and ethical aspects of deploying artificial intelligence in climate-smart agriculture. *AI Soc*. 2024;39(1):221-34.

<https://doi.org/10.1007/s00146-022-01421-2>.

Gardezi M, Joshi B, Rizzo DM, Ryan M, Prutzer E, Brugler S, et al. Artificial intelligence in farming: Challenges and opportunities for building trust. *Agron J*. 2024;116(3):1217-28.
<https://doi.org/10.1002/agj2.21353>.

Ryan M. The social and ethical impacts of artificial intelligence in agriculture: Mapping the agricultural AI literature. *AI Soc*. 2023;38(6):2473-85.
<https://doi.org/10.1007/s00146-021-01377-9>.

Shahriar S, Corradini MG, Sharif S, Moussa M, Dara R. The role of generative artificial intelligence in digital agri-food. *J Agric Food Res*. 2025;20:101787.
<https://doi.org/10.1016/j.jafr.2025.101787>.

Tagkopoulos I, Earles JM, Lemay DG, Liu X, Nitin N, Smith AD, et al. The AIFS Institute: Building a better food system through AI. *AI Mag*. 2024;45(1):89-93.
<https://doi.org/10.1002/aaai.12164>.

van Hilten M, Ryan M, Blok V, de Roo N. Ethical, legal and social aspects (ELSA) for AI: An assessment tool for agri-food. *Smart Agric Technol*. 2025;10:100710.
<https://doi.org/10.1016/j.atech.2024.100710>.

Ryan M, Atik C, Rijswijk K, Bogaardt MJ, Maes E, Deroo E. The future of agricultural data-sharing policy in Europe: Stakeholder insights on the EU Code of Conduct. *Humanit Soc Sci Commun*. 2024;11:1197.
<https://doi.org/10.1057/s41599-024-03710-1>.

Ruder SL, Wittman H. Agricultural data governance from the ground up: Exploring data justice with agri-food movements. *Big Data Soc*. 2025;12(1):20539517251330182.
<https://doi.org/10.1177/20539517251330182>.