

ORIGINAL RESEARCH

Open access

Responsible Foundation Models for Precision Agriculture: A Governance and Technical Safeguard Framework for Trustworthy Agricultural AI

Santiago Morales^{1*}, Laura Rojas², Camila Vega¹, Diego Molina³

Abstract

Foundation models are becoming relevant to precision agriculture because they can connect sensing, remote sensing interpretation, multimodal data integration, knowledge retrieval, advisory communication, and farm decision-support workflows. Their broad capability, however, does not establish agronomic validity, trustworthy performance, operational safety, or deployment readiness. This article develops an original governance and technical safeguard framework for responsible agricultural foundation models. The framework treats these models as potential components of decision infrastructure and integrates trustworthy artificial intelligence requirements with data governance, validation, human oversight, farmer participation, monitoring, and explicit claim boundaries. It identifies requirements concerning data quality, provenance, documentation, transparency, explainability, interpretability, uncertainty communication, robustness, generalization, bias, fairness, privacy, cybersecurity, accountability, auditability, and traceability. It also maps risks arising from opaque pretraining data, uncertain data control, hallucination, factual error, agronomic invalidity, platform dependency, model drift, data drift, and unclear responsibility. Technical safeguards include data screening, domain-adaptation assessment, external and local validation, robustness testing, bias and fairness review, factual and agronomic checking, human-in-the-loop review, and continuing monitoring. Responsible deployment principles add purpose limitation, context sensitivity, incident response, withdrawal planning, and non-autonomy boundaries. The contribution is a structured conceptual framework that connects governance questions with technical evidence and accountable actors without claiming legal compliance, guaranteed trustworthiness, farmer approval, or deployment readiness.

Keywords Precision agriculture, Responsible AI, Foundation models, Trustworthy AI, AI governance, Technical safeguards

*Correspondence:

Santiago Morales

santiago.morales@agro.uba.ar

¹ Department of Responsible AI Governance for Agriculture, Faculty of Agriculture, University of Buenos Aires, Buenos Aires, Argentina

² Department of Technical Safeguards and Trustworthy AI, Faculty of Agriculture, Pontifical Catholic University of Chile, Santiago, Chile

³ Department of Foundation Model Governance, Faculty of Agriculture, National University of La Plata, La Plata, Argentina

Introduction

Foundation models are broad AI systems that can be adapted to many downstream tasks. In precision agriculture, they may connect image analysis, textual knowledge, spatial information, temporal observations, sensor streams, and management records. Current agricultural foundation-model research spans language, vision, multimodal analysis, decision support, and robotics, while also identifying unresolved problems involving data availability, domain adaptation,

distribution shift, resource demands, and contextual validity [1, 2]. Their significance therefore lies not only in model scale, but in their possible role as reusable infrastructure for agricultural information processing.

Large language models illustrate both the promise and fragility of this infrastructure. They may support agricultural question answering, information retrieval, summarisation, and advisory communication, yet fluent language can hide unsupported claims, outdated knowledge, or missing local context. Agricultural LLM applications consequently raise concerns about hallucination, bias, factual reliability, privacy, and the suitability of general-purpose language capability for specialised agronomic settings [3]. Similar concerns apply when broad visual or multimodal representations are transferred into crop, soil, disease, or remote-sensing workflows without adequate evaluation.

Responsible use requires more than technical performance. Agricultural AI is embedded in relationships among farmers, advisors, developers, data stewards, cooperatives, platform providers, extension services, and public institutions. Ethical and responsible agricultural AI therefore requires accountability, transparency, explainability, privacy, fairness, data governance, stakeholder participation, and meaningful human oversight alongside model accuracy [4]. These issues become more consequential when outputs may affect resource allocation, crop-management choices, risk interpretation, or control over agricultural data.

This article proposes a governance and technical safeguard framework for responsible foundation models in precision agriculture. It does not certify models or prescribe universal deployment procedures. Instead, it links use-case definition, stakeholder mapping, data governance, validation, uncertainty communication, human review, monitoring, accountability, and decision boundaries. The central argument is that foundation models may support agricultural decision infrastructure only when their role, evidence base, limitations, responsible actors, technical safeguards, and residual risks are made visible and subject to continuing review.

Foundation Models as Agricultural Decision Infrastructure

Vision foundation models may provide reusable visual representations for crop, disease, weed, soil-surface, or field-condition analysis. Their value depends on the relationship between pretraining data, downstream agricultural tasks, adaptation data, and evaluation context. Agricultural vision-foundation-model research indicates that generic representations require task-specific specialisation and downstream assessment before they can be interpreted as suitable for agricultural use [5]. Reusability should therefore be treated as a technical opportunity rather than evidence of agronomic validity.

Adaptation techniques include linear probing, fine-tuning, parameter-efficient methods, transfer learning, and knowledge distillation. These methods may reduce the resources needed to specialise models, but their performance remains sensitive to dataset composition, labels, crop coverage, and differences between development and intended-use settings. Comparative agricultural evidence suggests that adapted visual representations can support disease and weed identification, yet benchmark or cross-dataset results do not independently establish field readiness [6]. Adaptation must remain an assessable stage with explicit evidence requirements.

Geospatial and remote-sensing foundation models extend decision infrastructure to satellite and other spatial observations. Potential applications include crop monitoring, land interpretation, environmental assessment, and regional analysis. However, sensor variation, spatial resolution, temporal frequency, geographic transfer, modality heterogeneity, and scale dependence create substantial constraints. Reviews of remote-sensing foundation models identify continuing barriers involving data harmonisation, architecture, downstream reasoning, and transfer across regions and tasks [7]. Agricultural use therefore requires context-specific evaluation rather than assumptions of universal transferability.

Large language, vision-language, multimodal, and time-series foundation models may combine imagery, spatial data, weather, sensors, management histories, and textual knowledge. Prompting and retrieval-augmented generation may support knowledge access, while multimodal architectures may connect evidence across farm workflows. Reviews of

large vision and language models in agriculture nevertheless identify challenges involving grounding, bias, privacy, reliability, and responsible use [8]. These systems should thus be framed as bounded decision-support components, not autonomous agricultural authorities.

Trustworthy Agricultural AI Requirements

Trustworthy agricultural AI combines technical reliability, social legitimacy, institutional responsibility, and stakeholder experience. Trust cannot be reduced to benchmark accuracy. It also depends on whether responsibility is assigned, farmers can challenge outputs, data practices are transparent, and risks and benefits are fairly distributed. Agricultural research on responsibility, trust, and ethical impact shows that trustworthy use requires explicit accountability and attention to consequences beyond model performance [9, 10]. Trust should therefore be assessed and calibrated rather than assumed.

Transparency, explainability, and interpretability serve different functions. Transparency concerns system purpose, data flows, limitations, and responsible actors. Explainability concerns whether relevant output factors can be communicated to users, while interpretability concerns whether qualified reviewers can inspect model behaviour. Explainable-AI research shows that methods differ in objective, fidelity, scope, and suitability for different audiences [11]. An explanation is insufficient when it is misleading, unstable, difficult to understand, or detached from uncertainty.

Validation, robustness, generalization, privacy, cybersecurity, and accountability are also essential. Validation tests fitness for a defined use; robustness tests stability under plausible disruption; and generalization examines transfer across farms, crops, seasons, sensors, and regions. Privacy and cybersecurity protect sensitive data and system integrity, while accountability identifies actors with authority to respond to failure. Trustworthy-AI scholarship treats these as interdependent technical, organisational, and social conditions rather than isolated checklist items [12].

Human oversight, farmer participation, autonomy, auditability, and traceability complete the requirement set. Human review must provide authority to question, override, or stop model-supported action. Farmer participation should influence use-case design, risk identification, evaluation, and feedback. Auditability requires inspectable evidence, while traceability links outputs to data, models, prompts, retrieval sources, and workflow decisions. None of these requirements guarantees trustworthiness; together they structure evidence, scrutiny, and residual-risk disclosure. **Table 1** summarises trustworthy agricultural AI requirements for responsible foundation models in precision agriculture.

Table 1. Trustworthy Agricultural AI Requirements for Responsible Foundation Models in Precision Agriculture: Data Quality, Provenance, Documentation, Transparency, Explainability, Interpretability, Uncertainty Communication, Validation, Robustness, Bias, Fairness, Privacy, Cybersecurity, Human Oversight, Farmer Participation, Accountability, Auditability, Traceability, and Trust Boundaries

Trustworthy AI requirement	Core governance question	Required evidence or practice	Trust function	Potential output	Failure risk	Evaluation or oversight need	Fra
Data quality	Are inputs accurate and representative?	Quality, missingness, label, sensor, and coverage checks	Reduces input error	Quality report	Biased or unstable output	Data and domain review	Er
Data provenance	Can sources and	Source, custody, processing,	Supports accountability	Provenance register	Unverifiable lineage	Data-governance audit	Tra

	transformations be traced?	and version records					
Data documentation	Are scope and limitations visible?	Dataset description and known exclusions	Exposes assumptions	Dataset dossier	Hidden gaps	Independent review	Risk
Model documentation	Are model design, adaptation, and limitations recorded?	Version, training, prompts, retrieval, and intended-use records	Supports reproducibility	Model dossier	Undisclosed change	Technical sign-off	Acco ex
Transparency	Do stakeholders understand purpose and limits?	Use disclosure and limitation statement	Supports informed use	Transparency notice	Hidden automation	Governance review	Cro req
Explainability	Are explanations suitable and faithful?	Stakeholder-matched explanation and fidelity test	Supports scrutiny	User explanation	Misleading rationale	Technical and usability review	C sa
Interpretability	Can specialists inspect behaviour?	Diagnostic and expert analysis	Supports investigation	Interpretation record	False confidence	Specialist assessment	Di sa
Uncertainty communication	Are limits and low-confidence conditions visible?	Calibration, warnings, and abstention rules	Reduces overreliance	Uncertainty statement	Confident misuse	Calibration and comprehension review	Cons use
Validation	Is the model fit for its defined use?	Prespecified tests, baselines, and errors	Tests fitness	Validation dossier	Benchmark overclaim	Independent review	Re
Robustness	Does performance withstand disruption?	Noise, missing-data, and stress tests	Tests stability	Robustness profile	Brittle behaviour	Scenario review	Relie
Generalization	Where does performance transfer?	External, local, temporal, and subgroup tests	Defines boundaries	Generalization statement	Context failure	External/local validation	Us
Bias assessment	Which systematic distortions exist?	Disaggregated testing and data-gap analysis	Identifies bias	Bias register	Unequal errors	Multidisciplinary review	iden

Fairness assessment	Are disparities unjustified?	Contextual criteria and impact analysis	Supports equity	Fairness assessment	Metric-only fairness	Stakeholder review	Go
Privacy	Are sensitive data protected?	Minimisation, access, retention, and privacy tests	Protects interests	Privacy assessment	Leakage or inference	Privacy review	sa
Cybersecurity	Are data and model pathways protected?	Threat modelling, controls, and monitoring	Protects integrity	Security record	Manipulation or disruption	Security review	Te sa
Human oversight	Can qualified people intervene?	Roles, escalation, override, and workload design	Preserves control	Oversight protocol	Rubber-stamping	Human-factors review	Non- bo
Farmer participation	Have affected farmers shaped evaluation?	Participation and documented feedback	Supports relevance	Participation record	Tokenism	Inclusion review	Fe me
Farmer trust	Is reliance calibrated to evidence?	Trust and communication assessment	Supports appropriate reliance	Trust record	Overtrust or rejection	Longitudinal review	As o
Farmer autonomy	Is meaningful choice preserved?	Opt-out, alternatives, and contestability	Protects agency	Autonomy statement	Dependency	Governance review	De bo
Accountability	Are responsible actors identifiable?	Responsibility and remedy matrix	Prevents diffusion	Accountability map	Unowned failure	Organisational review	gov ft
Auditability	Can evidence be independently examined?	Logs, documents, and retained tests	Enables assurance	Audit trail	Unverifiable claims	Independent audit	As me
Traceability	Can outputs be reconstructed?	Versioned data, model, prompt, and workflow logs	Supports investigation	Trace record	Failed diagnosis	Technical review	It fo
No trustworthiness-guarantee boundary	Is residual uncertainty explicit?	Limitation and residual-risk statement	Prevents overclaim	Boundary declaration	Framework treated as certification	Independent final review	Ma lin

Governance Risks and Accountability Gaps

Governance risk begins with opaque pretraining and adaptation data. Agricultural models may combine satellite observations, sensors, images, weather records, soil data, crop information, and textual knowledge obtained through multiple organisations. When origin, processing, representation, or permitted use are unclear, it becomes difficult to judge whether outputs rest on suitable evidence or whether contributors retain meaningful control. Smart-farming ethics research has identified persistent concerns involving privacy, autonomy, inequality, data power, and responsible innovation [13].

Farm-data access and control create a second accountability gap. Farmers may contribute operational and commercially sensitive information without clear visibility into how it is retained, reused, combined, or transferred. Uncertainty over ownership, access, and secondary use can weaken willingness to participate in digital agricultural systems [14]. Privacy must therefore be considered together with control and autonomy, while cybersecurity must address manipulation, insecure integrations, poisoned retrieval sources, and compromised model components.

Responsibility is further complicated by distributed development and deployment. One organisation may pretrain a model, another may adapt it, a platform may integrate it, an institution may procure access, and an advisor or farmer may interpret the output. Responsibility can be shifted unless authority, evidence duties, monitoring obligations, and escalation pathways are specified. Research on responsible smart-farming innovation argues for integrated processes that connect technology development, stakeholder participation, and continuing socio-ethical assessment [15]. Human oversight must be assessed within this structure, because assigning a reviewer does not ensure competence, time, independence, or authority.

Platform dependency, weak auditability, limited traceability, model drift, data drift, and readiness overclaiming add further risk. Proprietary platforms may control data access, interfaces, updates, interoperability, pricing, and withdrawal options. Research on agricultural platformisation shows how concentrated infrastructures can reinforce dependency and asymmetric control [16]. The framework therefore preserves four limits: it does not determine legal compliance; formal roles do not establish accountability sufficiency; human review does not guarantee error prevention; and benchmark or framework conformity does not establish deployment readiness.

Technical Safeguards for Foundation Models

Technical safeguards begin with data-quality screening, provenance tracking, and dataset and model documentation. Agricultural inputs should be examined for missingness, measurement error, label quality, geographic coverage, seasonal relevance, and compatibility across modalities. Documentation should identify data sources, transformations, model versions, adaptation procedures, prompts, retrieval resources, intended uses, and limitations. Uncertainty methods can support risk communication, but their estimates remain dependent on assumptions and calibration quality [17].

Bias and fairness review should examine agricultural data, labels, model design, evaluation methods, and workflow effects. Aggregate accuracy may conceal unequal performance across crops, regions, farm types, technologies, or user groups. Fairness criteria must therefore be selected according to the affected stakeholders and decision context. Machine-learning research shows that bias can enter at multiple stages and that different fairness definitions may conflict [18].

Benchmarking should be separated from external, local, and field validation. External validation tests transfer beyond development data, while local validation examines the intended crop, location, season, sensor, and workflow. Robustness testing should assess noise, missing inputs, environmental variation, prompt changes, and retrieval failures. Model- and data-drift monitoring remain necessary because agricultural conditions and model behaviour change over time [19].

Generative systems require hallucination monitoring, factual verification, and agronomic validity review. Fluent output may contain unsupported or contextually unsuitable claims. Claim-level checking should therefore be combined with reliable sources and qualified agricultural interpretation. Hallucination research confirms that generated text may remain plausible while being unfaithful or factually incorrect [20].

Human-in-the-loop review should define reviewer competence, evidence access, escalation duties, and override authority. Traceability should connect outputs to relevant data, prompts, retrieval sources, model versions, and reviewer actions. Auditability should examine governance, model, and application layers without implying certification [21]. **Table 2** maps the principal risks and safeguards.

Table 2. Governance Risks, Accountability Gaps, and Technical Safeguards for Agricultural Foundation Models: Opaque Pretraining Data, Data Provenance, Privacy, Cybersecurity, Bias, Fairness, Hallucination, Factual Accuracy, Agronomic Validity, Model Drift, Data Drift, Validation, Explainability, Human Review, Auditability, and Deployment Boundaries

Risk, gap, or safeguard	Core accountability or technical question	Required input or procedure	Governance or safeguard function	Potential output	Residual or failure risk	Validation or oversight need	
Pretraining opacity	What inherited data and limitations shaped the model?	Available training descriptions and limitation review	Exposes inherited uncertainty	Limitation statement	Unavailable information	Independent review	
Data provenance	Can sources and transformations be traced?	Versioned lineage records	Supports accountability	Provenance register	Incomplete metadata	Data audit	
Privacy and cybersecurity	Are sensitive data and system integrity protected?	Minimisation, access controls, threat testing, monitoring	Reduces exposure and manipulation	Risk register	Leakage or attack	Specialist review	
Bias and fairness	Are errors or burdens unequally distributed?	Disaggregated testing and impact review	Detects distortion	Bias/fairness report	Hidden groups or metric conflict	Stakeholder review	
Validation	Is the model fit for the defined setting?	Benchmark, external, local, robustness, and field-validation plans	Tests bounded fitness	Validation dossier	Overgeneralization	Independent review	
Hallucination and factual accuracy	Are generated claims supported?	Prompt tests, grounding, and claim verification	Detects unsupported output	Error log	Undetected claims	Domain review	

Agronomic validity	Is output appropriate for local agricultural conditions?	Crop, soil, weather, and management review	Tests contextual relevance	Agronomic review	Missing context	Qualified review	Co
Explainability and uncertainty	Are explanations faithful and limitations visible?	Fidelity, calibration, and comprehension tests	Supports calibrated reliance	Explanation and warning	Misleading confidence	Technical and user review	Co
Human review	Can a qualified person reject or stop action?	Roles, escalation, override, and workload design	Preserves human control	Review record	Automation bias	Human-factors review	No r
Drift monitoring	Have inputs or model behaviour changed?	Baselines, alerts, versioning, and revalidation	Detects degradation	Drift report	Silent change	Continuing oversight	
Auditability and traceability	Can decisions and evidence be reconstructed?	Logs, documents, versions, and access	Supports investigation	Audit trail	Missing evidence	Independent audit	r
Deployment boundary	Is readiness assessed separately?	Residual-risk synthesis	Prevents overclaim	Boundary statement	Checklist treated as approval	Multidisciplinary review	

Figure 1 illustrates the technical safeguard pipeline for responsible foundation models in precision agriculture, linking data provenance, model documentation, validation, robustness, uncertainty, explainability, human oversight, monitoring, and deployment boundaries.

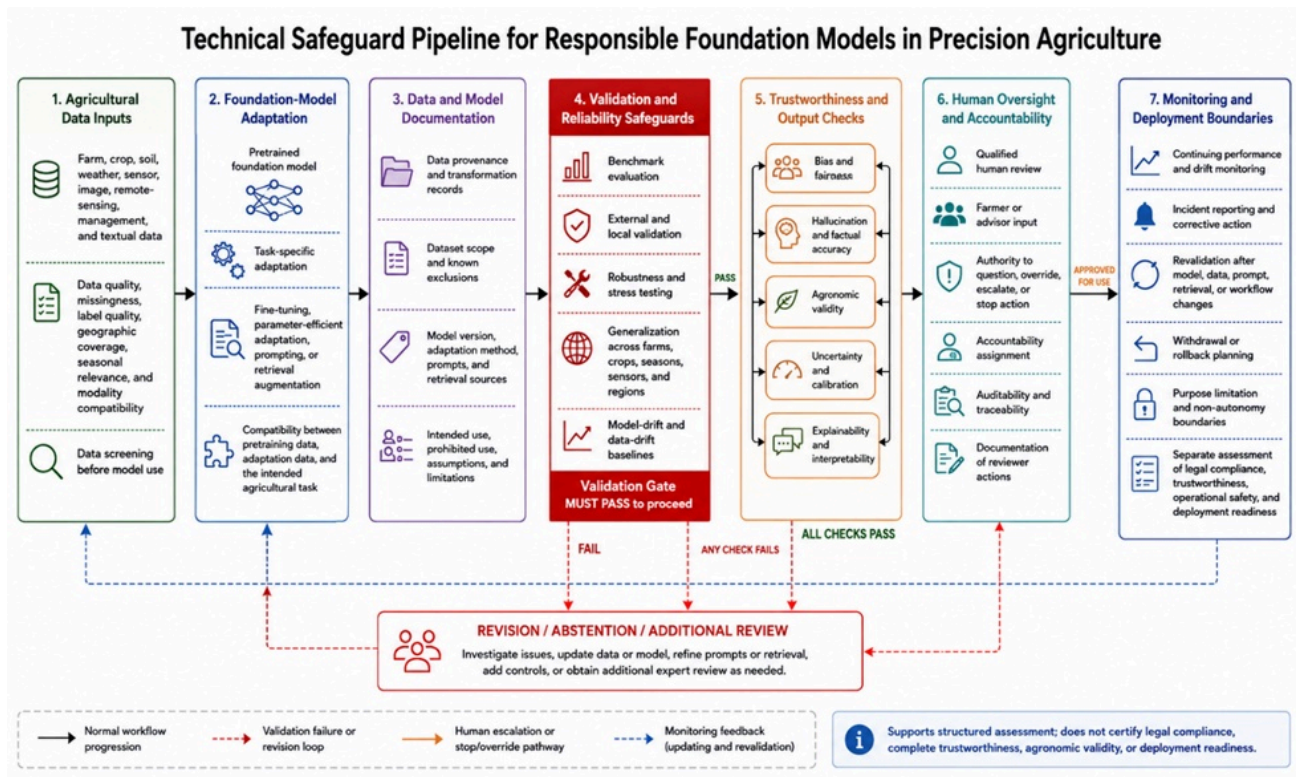


Figure 1. Technical Safeguard Pipeline for Responsible Foundation Models in Precision Agriculture

Alt text: A left-to-right pipeline connecting agricultural data, model adaptation, documentation, validation, risk checks, human oversight, monitoring, and deployment boundaries.

Responsible Deployment Principles

Responsible deployment begins with purpose limitation, use-case definition, and risk classification. Each system should have a defined task, user, output, consequence, and prohibited-use statement. Higher-consequence applications require stronger evidence and more restrictive boundaries. Human-centred AI emphasises meaningful human control rather than automation as the default objective [22].

Human oversight should assign complementary human and machine roles. Reviewers need sufficient context, competence, time, and authority to question or reject outputs. Hybrid intelligence similarly treats value as dependent on appropriate allocation of human and computational capabilities [23]. Human presence alone does not guarantee effective oversight.

Farmer participation should influence use-case selection, risk identification, evaluation, and feedback. Farmers should be able to contest unsuitable assumptions and understand data practices, limitations, and alternatives. Responsible-innovation research in farming supports participation, inclusion, and contextual assessment [24]. Consultation should not be represented as universal farmer approval.

Deployment assessment must also address local validation, privacy, cybersecurity, explainability, uncertainty, auditability, monitoring, incident reporting, and withdrawal capacity. Digital agriculture is embedded in organisational, institutional, and power relationships that affect adoption and accountability [25]. These principles provide conceptual guidance rather than legal, procurement, agronomic, or operational instructions.

Proposed Governance and Safeguard Framework

The proposed framework connects cooperative governance, organisational AI governance, and practical safeguard implementation. Responsible AI requires coordinated actors, lifecycle structures and processes, and procedures that translate principles into evidence and reviewable outputs [26-28]. Governance and technical safeguards are therefore paired at every stage.

The first stage defines the agricultural use case and maps stakeholders, decision context, data sources, provenance, pretraining information, adaptation procedures, data-control arrangements, privacy, cybersecurity, risks, and accountable actors. Its purpose is to clarify scope and responsibility before technical evaluation begins.

The second stage selects trustworthy-AI requirements and technical safeguards. It plans validation, bias and fairness review, explanations, uncertainty communication, human oversight, and farmer participation. Requirements should be proportionate to the use case and produce inspectable evidence.

The third stage reviews readiness, monitoring, drift, incidents, withdrawal, feedback, and framework updating. Ethics-based auditing can identify governance gaps and support correction but does not certify trustworthiness [29]. Changes to data, models, prompts, retrieval resources, or workflows require renewed assessment.

The framework also recognises organisational incentives and public-interest effects. Governance must consider whether providers or deployers possess adequate authority, resources, and motivation to disclose limitations and respond to failures [30]. **Table 3** presents the framework stages.

Table 3. Proposed Governance and Technical Safeguard Framework for Responsible Foundation Models in Precision Agriculture: Use-Case Definition, Stakeholder Mapping, Data Provenance, Model Documentation, Data Governance, Privacy, Cybersecurity, Risk Mapping, Validation, Bias and Fairness Review, Explainability, Uncertainty, Human Oversight, Farmer Feedback, Monitoring, Incident Response, Framework Updating, and Claim Boundaries

Framework stage	Core purpose	Required input	Governance function	Technical safeguard function	Potential output	Validation or assessment need	Failure risk
Use-case definition	Define task and limits	User, workflow, output, consequence	Establishes scope	Selects relevant tests	Use-case charter	Domain review	Function creep
Stakeholder mapping	Identify affected and responsible actors	Farmer, advisor, provider, institution roles	Allocates participation and authority	Identifies workflow constraints	Stakeholder map	Inclusion review	Exclusion
Data and provenance review	Assess source, quality, and control	Data inventory and lineage	Supports stewardship	Enables quality screening	Data register	Data audit	Opaque data
Model documentation	Record model and adaptation choices	Versions, prompts, retrieval, fine-tuning	Assigns disclosure duties	Supports reproducibility	Model dossier	Technical review	Hidden change

Privacy and security review	Identify exposure and integrity risks	Data flows and threat model	Allocates responsibility	Tests controls	Risk register	Specialist testing	Leakage attack
Risk and accountability mapping	Connect failures to responsible actors	Hazard and role analysis	Prevents responsibility diffusion	Links risks to controls	Accountability map	Multidisciplinary review	Unknown risk
Safeguard and validation planning	Select tests and evidence	Risk profile and endpoints	Assigns safeguard ownership	Defines validation and monitoring	Safeguard plan	Independent review	Weakened evidence
Human oversight and participation	Preserve control and contextual input	Roles, authority, farmer feedback	Supports legitimacy and accountability	Implements review and override	Oversight protocol	Human-factors review	Tokenism or rubber stamping
Readiness review	Integrate evidence and unresolved risk	Governance and technical records	Enables bounded decision	Assesses residual limitations	Readiness report	Independent review	Checklist approval
Monitoring and incident response	Detect change and respond	Logs, thresholds, escalation, rollback	Sustains accountability	Detects drift and supports withdrawal	Monitoring record	Periodic revalidation	Silent degradation
Framework updating	Revise safeguards and responsibilities	Audits, incidents, research, feedback	Maintains relevance	Updates tests and controls	Revised framework	Change assessment	Static governance
Claim boundaries	Prevent legal, trust, autonomy, or readiness overclaim	Residual-risk statements	Limits assurance claims	Requires separate evidence	Boundary declaration	Final independent review	Framework treated as certificate

Figure 2 presents the proposed governance and technical safeguard framework for responsible foundation models in precision agriculture, integrating agricultural use-case definition, trustworthy AI requirements, governance safeguards, technical safeguards, responsible deployment principles, monitoring, and feedback updating.

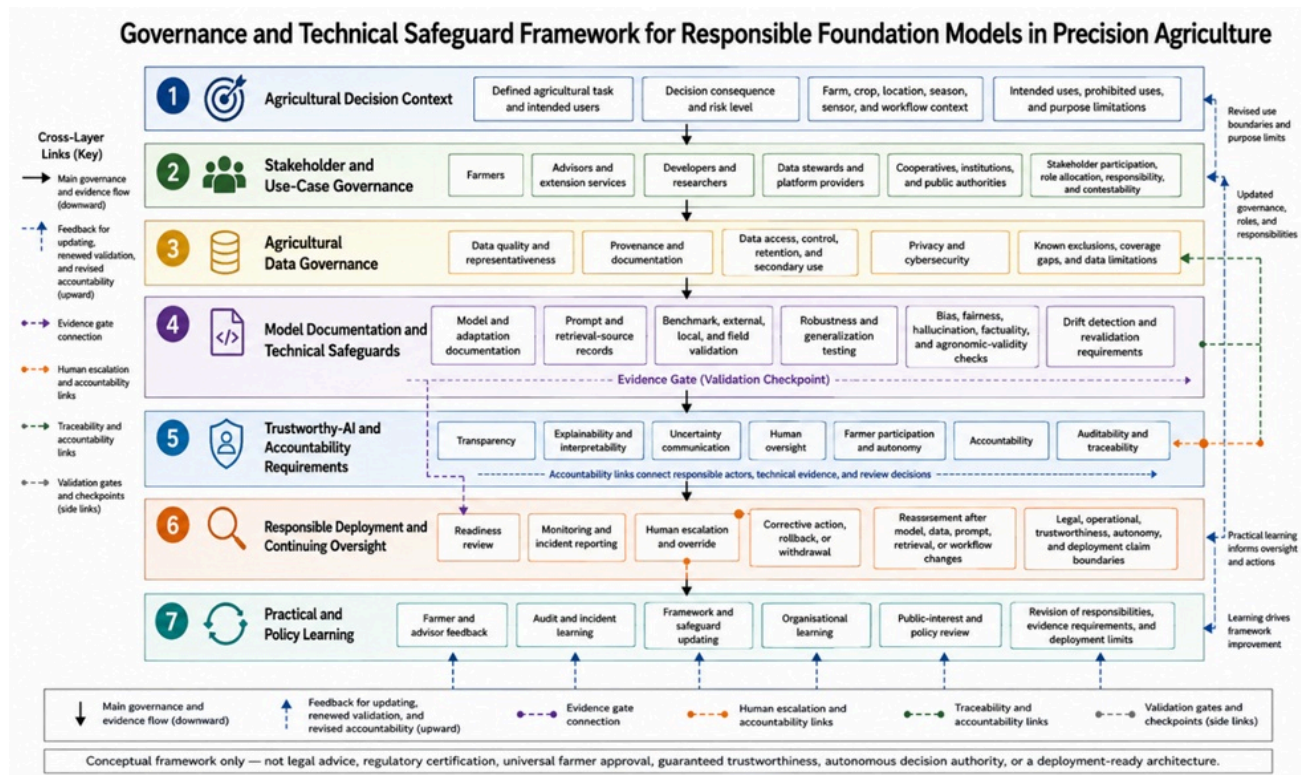


Figure 2. Governance and Technical Safeguard Framework for Responsible Foundation Models in Precision Agriculture

Alt text: A layered framework connecting agricultural decision context, stakeholders, data governance, technical validation, trustworthy AI, responsible deployment, monitoring, and policy learning.

Practical and Policy Implications

Farmers, advisors, cooperatives, and extension services need understandable uncertainty communication, meaningful review rights, and visible data-control arrangements. Unclear access and reuse conditions can weaken farmer participation and trust [14]. Farmer-centred evaluation should therefore assess comprehension, autonomy, contestability, and dependence on platforms.

Developers, platform providers, data stewards, and researchers should create documentation and validation infrastructure together. Provenance records, model histories, disaggregated errors, security findings, and monitoring evidence should remain independently inspectable. Layered auditing supports assurance but should not be treated as certification [21].

Public institutions and policymakers may need capacity for validation, participation, audit access, incident learning, and accountability review. Responsible innovation requires inclusion [24], while ethics-based auditing can support corrective governance learning [29]. Public-interest review should also examine organisational incentives and platform dependency [30]. These implications are conceptual, not legal or regulatory prescriptions.

Conclusion

The proposed framework links agricultural foundation models, trustworthy-AI requirements, governance risks, technical safeguards, responsible deployment principles, accountability, monitoring, and feedback updating. It emphasises evidence, validation, transparency, farmer participation, and meaningful human oversight before readiness claims are

made. The framework supports structured evaluation but does not establish legal compliance, complete trustworthiness, autonomous decision authority, farmer approval, operational safety, or deployment readiness.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 13 Nov 2025

Revised: 10 Jan 2026

Accepted: 09 Feb 2026

Published online: 25 May 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

- 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.* 2025;15(8):847.
<https://doi.org/10.3390/agriculture15080847>.
- De Clercq D, Nehring E, Mayne H, Mahdi A. Large language models can help boost food production, but be mindful of their risks. *Front Artif Intell.* 2024;7:1326153.
<https://doi.org/10.3389/frai.2024.1326153>.
- 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>.

Ye J, Yu Z, Lin J, Li H, Lin L. Vision foundation model for agricultural applications with efficient layer aggregation network. *Expert Syst Appl.* 2024;257:124972.
<https://doi.org/10.1016/j.eswa.2024.124972>.

Espejo-Garcia B, Güldenring R, Nalpantidis L, Fountas S. Foundation vision models in agriculture: DINOv2, LoRA and knowledge distillation for disease and weed identification. *Comput Electron Agric.* 2025;239:110900.
<https://doi.org/10.1016/j.compag.2025.110900>.

Huo C, Chen K, Zhang S, Wang Z, Yan H, Shen J, et al. When remote sensing meets foundation model: A survey and beyond. *Remote Sens.* 2025;17(2):179.
<https://doi.org/10.3390/rs17020179>.

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>.

Alexander CS, Yarborough M, Smith A. Who is responsible for 'responsible AI'? Navigating challenges to build trust in AI agriculture and food system technology. *Precis Agric.* 2024;25(1):146-85.
<https://doi.org/10.1007/s11119-023-10063-3>.

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>.

Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Benetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion.* 2020;58:82-115.
<https://doi.org/10.1016/j.inffus.2019.12.012>.

Thiebes S, Lins S, Sunyaev A. Trustworthy artificial intelligence. *Electron Mark.* 2021;31(2):447-64.
<https://doi.org/10.1007/s12525-020-00441-4>.

van der Burg S, Bogaardt MJ, Wolfert S. Ethics of smart farming: Current questions and directions for responsible innovation towards the future. *NJAS Wagening J Life Sci.* 2019;90-91:100289.
<https://doi.org/10.1016/j.njas.2019.01.001>.

Wiseman L, Sanderson J, Zhang A, Jakku E. Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming. *NJAS Wagening J Life Sci.* 2019;90-91:100301.
<https://doi.org/10.1016/j.njas.2019.04.007>.

Eastwood C, Klerkx L, Ayre M, Dela Rue B. Managing socio-ethical challenges in the development of smart farming: From a fragmented to a comprehensive approach for responsible research and innovation. *J Agric Environ Ethics.* 2019;32(5-6):741-68.
<https://doi.org/10.1007/s10806-017-9704-5>.

Sauvagerd M, Mayer M, Hartmann M. Digital platforms in the agricultural sector: Dynamics of oligopolistic platformisation. *Big Data Soc.* 2024;11(4):1-16.
<https://doi.org/10.1177/20539517241306365>.

Abdar M, Pourpanah F, Hussain S, Rezazadegan D, Liu L, Ghavamzadeh M, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Inf Fusion.* 2021;76:243-97.
<https://doi.org/10.1016/j.inffus.2021.05.008>.

Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *ACM Comput Surv.* 2021;54(6):115.
<https://doi.org/10.1145/3457607>.

Lu J, Liu A, Dong F, Gu F, Gama J, Zhang G. Learning under concept drift: A review. *IEEE Trans Knowl Data Eng.* 2019;31(12):2346-63.
<https://doi.org/10.1109/TKDE.2018.2876857>.

Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. *ACM Comput Surv.* 2023;55(12):248.
<https://doi.org/10.1145/3571730>.

Mökander J, Schuett J, Kirk HR, Floridi L. Auditing large language models: A three-layered approach. *AI Ethics.* 2024;4(4):1085-115.
<https://doi.org/10.1007/s43681-023-00289-2>.

Shneiderman B. Human-centered artificial intelligence: Reliable, safe & trustworthy. *Int J Hum Comput Interact.* 2020;36(6):495-504.
<https://doi.org/10.1080/10447318.2020.1741118>.

Dellermann D, Ebel P, Söllner M, Leimeister JM. Hybrid intelligence. *Bus Inf Syst Eng.* 2019;61(5):637-43.
<https://doi.org/10.1007/s12599-019-00595-2>.

Gardezi M, Adereti DT, Stock RJ, Ogunyiola A. In pursuit of responsible innovation for precision agriculture technologies. *J Responsible Innov.* 2022;9(2):224-47.
<https://doi.org/10.1080/23299460.2022.2071668>.

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>.

Gianni R, Lehtinen S, Nieminen M. Governance of responsible AI: From ethical guidelines to cooperative policies. *Front Comput Sci.* 2022;4:873437.
<https://doi.org/10.3389/fcomp.2022.873437>.

Mäntymäki M, Minkinen M, Birkstedt T, Viljanen M. Defining organizational AI governance. *AI Ethics.* 2022;2(4):603-9.
<https://doi.org/10.1007/s43681-022-00143-x>.

Morley J, Floridi L, Kinsey L, Elhalal A. From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Sci Eng Ethics.* 2020;26(4):2141-68.
<https://doi.org/10.1007/s11948-019-00165-5>.

Mökander J, Floridi L. Operationalising AI governance through ethics-based auditing: An industry case study. *AI Ethics.* 2023;3(2):451-68.
<https://doi.org/10.1007/s43681-022-00171-7>.

Cihon P, Schuett J, Baum SD. Corporate governance of artificial intelligence in the public interest. *Information.* 2021;12(7):275.
<https://doi.org/10.3390/info12070275>.