

REVIEW

Open access

Generative Artificial Intelligence in Precision Agriculture: A Critical Review of Synthetic Data, Decision Support, Model Reliability, and Responsible Deployment

Patrick O'Connor^{1*}, Grace Murphy¹, Niamh Walsh²

Abstract

Generative artificial intelligence is emerging as a potentially important component of precision agriculture through its capacity to create synthetic data, generate agricultural imagery, retrieve and reformulate knowledge, support conversational advisory interfaces, and assist decision-support workflows. However, evidence of technical feasibility is developing faster than evidence of agronomic validity, reliability, farmer-centred usefulness, and responsible deployment. This critical review evaluates generative AI as a distinct agricultural technology class, with particular attention to generative adversarial networks, diffusion models, variational autoencoders, large language models, vision-language models, foundation models, fine-tuning, prompting, domain adaptation, and retrieval-augmented generation. It critically examines the proposed use of synthetic crop, disease, weed, pest, and remote-sensing data for augmentation, class-imbalance mitigation, and label-scarcity support, while emphasizing image realism, biological plausibility, source-data quality, and synthetic-real domain gaps. The review also assesses generative decision support and advisory systems, distinguishing accessible language generation and knowledge retrieval from agronomically validated recommendations. Reliability concerns include hallucination, factual inaccuracy, uncertainty underreporting, poor local adaptation, robustness limitations, benchmark insufficiency, and failures of human oversight. Responsible deployment further requires data governance, privacy protection, accountability, transparency, explainability, farmer trust, digital inclusion, risk monitoring, and assessment of computational and sustainability costs. The review contributes an integrated opportunity-risk interpretation and identifies priorities for domain-specific benchmarks, synthetic-data validation, hallucination measurement, uncertainty-aware assistance, human-in-the-loop evaluation, farmer-centred design, governance documentation, and cautious field-readiness assessment.

Keywords Precision agriculture, Decision support, Generative AI, Synthetic data, Hallucination, Model reliability

*Correspondence:

Patrick O'Connor

patrick.oconnor@ucc.ie

¹ Department of Generative AI for Agriculture, School of Agriculture, University College Cork, Cork, Ireland

² Department of Synthetic Data and Reliable AI Deployment, Faculty of Agriculture, University of Galway, Galway, Ireland

Introduction

Generative artificial intelligence has expanded the functional scope of agricultural AI beyond the classification, prediction, and optimization tasks commonly associated with conventional machine learning. In precision agriculture, generative systems may create candidate images or structured samples, produce natural-language explanations, retrieve and reformulate technical knowledge, support multimodal interactions, and generate conditional scenarios for research or decision-support purposes. Agriculture-focused reviews consequently describe a broad landscape encompassing synthetic data, agricultural image generation, language-based advisory interfaces, knowledge services, and analytical

assistance. These opportunities are nevertheless accompanied by uneven evidence maturity, substantial variation among methods, and unresolved questions concerning whether generated outputs remain valid outside the datasets and experimental conditions in which they were evaluated [1-3].

The need for critical evaluation is particularly apparent in agricultural applications of large language models. Agricultural language generation can make complex information more accessible, facilitate interaction with technical resources, and support communication between data-intensive systems and non-specialist users. Yet the same systems can produce inaccurate, incomplete, outdated, or context-insensitive content in fluent and persuasive language. In food-production settings, such failures may affect how users interpret crop conditions, production risks, management information, or institutional guidance. Additional concerns arise when agricultural data are incorporated into proprietary or poorly documented systems, because unclear data practices, labour implications, and premature policy dependence may create risks that cannot be assessed through output quality alone [4].

A critical review is therefore required to distinguish technological possibility from demonstrated agricultural validity. Synthetic imagery may appear visually convincing without reproducing biologically meaningful structures, and improved performance on a selected image dataset may not transfer to different farms, seasons, cultivars, sensors, or environmental conditions. Similarly, a conversational system may retrieve relevant information or generate a coherent explanation without accounting for local soils, climate, crop stage, management history, resource constraints, or uncertainty. Generative AI should consequently be examined as a collection of methods and workflow components whose value depends on the quality of their source data, the intended task, the validation design, the human decision context, and the governance arrangements surrounding use.

This article critically reviews generative AI in precision agriculture across four connected domains: synthetic data and agricultural image generation; generative decision support and advisory systems; model reliability, hallucination, and validation risks; and responsible deployment. Its contribution is not to present generative systems as autonomous agricultural intelligence or as substitutes for farmers, agronomists, extension professionals, or local knowledge. Instead, it evaluates where generative methods may support research and carefully governed workflows, where available evidence remains preliminary or task-bounded, and which validation, oversight, documentation, and governance requirements should precede claims of reliability or deployment readiness.

Generative AI as an Emerging Precision Agriculture Technology

Generative AI refers to computational methods that learn patterns from data and produce new candidate content, representations, or responses rather than only assigning labels or estimating predefined outputs. In precision agriculture, this category can include models that generate images, text, multimodal descriptions, structured data, embeddings, simulations, or conditional scenarios. It should be distinguished from conventional discriminative learning, in which a model primarily maps an input to a class, score, forecast, or control variable. Foundation models represent a related but broader architectural category: they are pretrained on extensive datasets and may be adapted to multiple downstream tasks, but not every foundation model is generative, and not every generative agricultural system qualifies as a foundation model. Agricultural foundation-model research includes language, vision, and multimodal systems, while continuing to face unresolved problems involving domain representation, generalization, computational requirements, and evaluation [5].

Large language models are generative foundation models designed to predict and compose sequences of text, enabling question answering, summarization, explanation generation, translation, information extraction, and conversational interaction. Their relevance to precision agriculture lies partly in their ability to provide a natural-language interface to heterogeneous information that may otherwise remain distributed across sensor records, farm-management systems, scientific publications, extension materials, and local documentation. However, the usefulness of such interfaces depends

on the availability of relevant knowledge, suitable infrastructure, linguistic and regional adaptation, and mechanisms for identifying unsupported outputs. In smart-farming environments, conversational accessibility may reduce some technical barriers, but it does not itself demonstrate agronomic accuracy, equitable access, or operational suitability [6].

Prompting, fine-tuning, retrieval-augmented generation, and domain adaptation represent different approaches to modifying model behaviour. Prompting conditions a model through instructions or examples without necessarily changing its parameters, whereas fine-tuning modifies model parameters using additional task- or domain-specific data. Domain adaptation is a broader process intended to improve performance under a target agricultural distribution, which may include regional terminology, crop-specific knowledge, local practices, multimodal inputs, or production constraints. Retrieval-augmented generation supplements model generation with information retrieved from an external corpus, while agricultural knowledge graphs may provide structured relationships among crops, symptoms, conditions, and management concepts. These approaches can improve relevance under defined conditions, but their effectiveness remains dependent on corpus quality, provenance, updating, retrieval accuracy, and external evaluation [7].

Generative adversarial networks, variational autoencoders, and diffusion models are prominent generative approaches for visual or structured agricultural data. Generative adversarial networks use competition between a generator and discriminator to learn candidate samples, although instability, mode collapse, and incomplete distribution coverage may limit diversity. Variational autoencoders learn probabilistic latent representations from which new samples can be drawn, but reconstructed or generated outputs may smooth or omit task-relevant details. Diffusion models generate content through iterative denoising and may provide improved sample control or diversity in some contexts, although they can be computationally demanding and remain dependent on representative training data. Vision-language models connect visual and linguistic representations, potentially supporting image description, visual question answering, or multimodal decision assistance. These method classes should not be treated as interchangeable because they create different outputs, encode different assumptions, and require distinct validation strategies [8].

Synthetic Data and Agricultural Image Generation

Synthetic agricultural data are artificially generated samples intended to approximate selected characteristics of observed agricultural data. Potential applications include supplementing small datasets, increasing representation of minority classes, testing analytical workflows, and supporting model development where observation or annotation is difficult. In agricultural image analysis, generative adversarial networks and variational autoencoders have been widely discussed as mechanisms for producing or transforming crop, disease, weed, pest, and environmental imagery, while systematic evidence on adversarial augmentation shows substantial diversity in datasets, model configurations, evaluation methods, and downstream tasks. The resulting literature supports investigation of synthetic augmentation but does not justify treating synthetic data as interchangeable with field observations or assuming that augmentation will improve every model [9, 10].

Plant-disease image generation provides an instructive example of both potential and limitation. Adversarial image-translation methods have been used to transform available images and supplement underrepresented disease classes in selected recognition tasks. Such studies indicate that generated examples can contribute to classifier training under controlled conditions, particularly where class imbalance constrains learning. Nevertheless, a downstream improvement within one dataset does not demonstrate that synthetic lesions reproduce disease biology, severity progression, cultivar responses, illumination variability, or field backgrounds. Validation must therefore separate visual acceptability from pathological plausibility and must test the resulting model on independent real images rather than evaluating generated samples only through the model that helped produce or select them [11].

Conditional generative adversarial networks have similarly been applied to tomato-leaf disease imagery in combination with transfer learning. This approach may increase the number of training examples available for selected disease categories and can provide a controlled setting for comparing real-only and augmented training designs. Its interpretation must remain conditional on the source dataset, image-acquisition conditions, disease labels, augmentation ratio, classifier

architecture, and evaluation procedure. A model that performs better on a held-out portion of a related public dataset may still rely on background patterns, acquisition artefacts, or generated features that do not transfer to commercial fields, other cultivars, different cameras, or naturally varying symptom severity. Synthetic image usefulness is consequently a property of a specified workflow and validation context rather than an intrinsic feature of the generated image [12].

Weed-image generation extends the same problem into a domain where taxonomic structure, growth stage, occlusion, canopy arrangement, and field background are central to identification. Channel-attention generative adversarial methods have been investigated as a means of creating additional weed images for precise identification, potentially addressing the cost of collecting and annotating diverse field examples. However, a generated plant may appear coherent while blending structures from different species, omitting diagnostic features, or presenting unrealistic relationships among leaves, stems, scale, soil, and neighbouring crops. Evaluation should therefore include botanical review, species-specific error analysis, diversity assessment, and independent field testing. Label scarcity may justify synthetic-data research, but it does not justify converting generated labels into unquestioned ground truth [13].

Diffusion-based augmentation has also been investigated for long-tailed pest classification, where a small number of common classes may dominate training while rare pests remain poorly represented. Conditional generation may expand minority classes and support more balanced model development, but aggregate benchmark improvement cannot establish taxonomic validity or generalization to real scouting conditions. Pest images require attention to anatomy, life stage, pose, scale, damage context, imaging device, and background complexity. Comparable concerns apply to proposed synthetic crop imagery, remote-sensing imagery, and livestock imagery, for which perceptual realism may diverge from spectral, spatial, temporal, physiological, or behavioural validity. Across these domains, source-data quality, biological plausibility, distributional comparison, real-only external testing, and transparent reporting are essential to prevent synthetic samples from being interpreted as observational equivalents [14].

Table 1 summarises synthetic data and agricultural image generation opportunities, risks, and validation requirements in precision agriculture.

Table 1. Synthetic Data and Agricultural Image Generation in Precision Agriculture: Generative Adversarial Networks, Diffusion Models, Variational Autoencoders, Data Augmentation, Class Imbalance, Label Scarcity, Image Realism, Biological Plausibility, Synthetic-Real Domain Gap, Validation Needs, and Claim Boundaries

Synthetic data component	Core critical question	Required evidence or input	Generative function	Potential use	Risk of overinterpretation	Validation need
Synthetic crop images	Do generated images preserve crop morphology, growth stage, canopy structure, and environmental context?	Diverse field images, crop-stage labels, acquisition metadata, and independent real test data	Generate or translate crop imagery	Supplement limited training data and explore controlled variation	Visually plausible crops may contain anatomically or phenologically impossible features	Agronomic review, feature-level comparison, diversity analysis, and independent real-data testing
Synthetic disease images	Are lesion shape, colour, distribution, severity, and host response	Expert-labelled disease images, severity information, cultivar	Generate or translate symptomatic images	Support minority disease classes and controlled augmentation studies	Generated lesions may encode visual shortcuts rather than disease processes	Plant-pathology review, lesion-specific assessment, cross-dataset

	biologically credible?	metadata, and multi-environment observations				testing, and external field validation
Synthetic weed images	Are taxonomic features retained across growth stages, occlusion patterns, and field backgrounds?	Botanically verified images, species labels, scale information, and field-context diversity	Generate weed specimens or scenes	Address species-level annotation scarcity	Images may blend species traits or omit diagnostic structures	Botanical review, species-level error analysis and independent field evaluation
Synthetic pest images	Are anatomy, life stage, scale, pose, and damage context accurately represented?	Taxonomically verified pest images and long-tailed class distributions	Generate minority-class pest examples	Rebalance pest-classification datasets	Improved class balance may conceal unrealistic organisms or backgrounds	Entomological review, per-class metrics and out-of-distribution testing
Synthetic remote-sensing imagery	Are spectral, spatial, radiometric, and temporal relationships physically credible?	Calibrated sensor observations, geospatial metadata, atmospheric information, and phenological records	Generate, reconstruct, or translate remote-sensing scenes	Explore missing-data reconstruction or rare-condition augmentation	Texture realism may be confused with radiometric or agronomic validity	Spectral-fidelity testing, spatial consistency checks, temporal validation, and real-scene downstream evaluation
Synthetic livestock imagery	Are anatomy, behaviour, welfare indicators, and housing contexts represented without artefacts?	Multi-farm imagery, veterinary annotations, behavioural labels, and acquisition metadata	Generate animal images or behavioural scenes	Potentially supplement rare-event observations	A visually coherent animal may contain invalid posture, behaviour, or welfare signals	Veterinary and ethological review with independent multi-farm testing
Data augmentation	Does synthetic augmentation add useful variation rather than reproduce source-data bias?	Real-only baselines, controlled augmentation ratios, provenance records, and	Add generated samples to a training distribution	Expand sample diversity under defined conditions	Any observed gain may be dataset-specific or caused by regularization rather than realistic generation	Controlled comparisons, confidence analysis, and independent real-only test sets

		ablation designs				
Class imbalance mitigation	Do minority-class improvements occur without degrading calibration or other classes?	Transparent class distributions and class-specific evaluation	Generate targeted minority-class samples	Improve representation of rare diseases, weeds, or pests	Aggregate accuracy may conceal unequal class performance	Per-class recall, precision, calibration, confusion analysis, and repeated testing
Label scarcity support	Can generated or transferred labels reduce annotation burden without amplifying error?	Expert labels, label uncertainty, adjudication procedures, and representative samples	Conditional generation or image-to-image translation	Supplement sparsely labelled classes	Synthetic labels may be treated as verified ground truth	Expert label auditing and disagreement analysis
Generative adversarial networks	Do adversarial training and sampling provide adequate stability and distribution coverage?	Representative source data, reproducible training details, and diversity measures	Adversarial image or data generation	Crop, disease, weed, and other image augmentation	Mode collapse and training instability may produce narrow or misleading diversity	Fidelity, diversity, stability, and real-domain downstream testing
Diffusion models	Does iterative denoising improve controllability and minority-class coverage in agricultural data?	Condition labels, sufficient computation, representative real comparators, and reproducible settings	Denoising-based conditional generation	Long-tailed pest or disease augmentation	High visual quality may be mistaken for biological accuracy	Class-conditional fidelity, diversity, computational reporting, and external testing
Variational autoencoders	Does latent-space sampling preserve rare but diagnostically important variation?	Representative real data, latent-space analysis, and reconstruction diagnostics	Probabilistic encoding, reconstruction, and sampling	Structured synthetic data or image generation	Over-smoothed outputs may remove diagnostic details	Reconstruction analysis, distribution comparison and task-specific validation
Image realism assessment	Which aspects of realism are relevant to the	Expert assessments, perceptual	Evaluate candidate	Screen obviously unusable samples	A single realism score may be	Multi-dimensional assessment

	intended agricultural task?	measures, feature statistics, and task evidence	synthetic outputs		treated as proof of validity	combining technical and domain expertise
Biological plausibility	Could the generated crop, symptom, weed, pest, or animal state occur biologically?	Domain knowledge, phenological rules, taxonomic criteria, and causal constraints	Constrain, filter, or audit generation	Remove impossible or misleading samples	Coherent visual composition may conceal impossible biology	Specialist review and explicit rejection criteria
Synthetic-real domain gap	Does synthetic-supported performance transfer to unseen real farms and devices?	Independent farms, seasons, cultivars, sensors, and environmental conditions	Compare synthetic and real distributions and model behaviour	Estimate transferability	Related-dataset evaluation may substantially underestimate domain shift	External, geographic, temporal, and device-level validation
Data quality	Are source data representative, correctly labelled, traceable, and free from leakage?	Provenance, metadata, quality audits, and sampling documentation	Learn and reproduce patterns in source data	Potentially expose or supplement data gaps	Generators may reproduce noise, bias, artefacts, and restricted variation	Upstream data audit before generation and downstream error analysis
No synthetic-equivalence claim boundary	Can generated data be treated as interchangeable with field observations?	Direct evidence across technical, biological, distributional, and task-relevant properties	Interpretive boundary rather than a generation function	Prevent unsupported claims	"Realistic" may be interpreted as "real" or "observationally equivalent"	Explicit non-equivalence language and context-specific validation
No performance-guarantee boundary	Does adding synthetic data necessarily improve model performance or reliability?	Controlled real-only comparisons, repeated studies, external tests, and negative results	Interpretive boundary rather than a generation function	Prevent universal augmentation recommendations	Positive selected studies may be generalized across tasks and farms	Replication, ablation, uncertainty reporting, and independent evaluation

Generative Decision Support and Advisory Systems

Large language models may support agricultural extension and advisory workflows by translating technical material, responding to questions, summarizing evidence, and enabling interaction through natural language. Such functions could help users navigate complex information systems and may provide preliminary explanations or candidate responses where extension capacity is limited. Their role must nevertheless be interpreted as assistance rather than replacement. Agricultural decisions are shaped by local knowledge, tacit experience, crop and livestock conditions, institutional rules, economic constraints, and rapidly changing environmental circumstances that may not be represented in a language model. Extension professionals, agronomists, and farmers therefore remain necessary for contextual interpretation, verification, and decision authority, particularly where generated content could influence consequential production choices [15].

Agricultural question-answering systems illustrate the importance of the knowledge architecture beneath the conversational interface. These systems may draw from textual corpora, databases, knowledge graphs, retrieval engines, language models, or combinations of these components. Their usefulness depends on whether relevant agricultural knowledge has been represented at the appropriate temporal, regional, seasonal, and production-system level. A question about disease symptoms, irrigation, nutrient status, or weather risk cannot be answered reliably from generic terminology alone; the system may need crop stage, location, recent observations, measurement units, and source provenance. Agricultural question answering can therefore improve access to organized information, but the generation layer cannot compensate automatically for incomplete corpora, inconsistent terminology, missing metadata, or outdated knowledge [16].

Hybrid systems that connect large language models with agricultural knowledge graphs or retrieval resources may provide a more structured basis for answering plant-health questions. Structured relationships can help identify candidate crops, symptoms, pathogens, and explanatory concepts, and retrieved records can reduce exclusive reliance on parametric model memory. However, a retrieved statement may still be irrelevant, incomplete, outdated, or unsuitable for the farm context, while the generator may combine correct evidence into an unsupported conclusion. Vision-language models introduce additional possibilities by linking images with textual interaction, but uncertain image quality, ambiguous symptoms, and visually similar disorders can propagate errors into the generated explanation. Knowledge retrieval and multimodal input should consequently be treated as evidence channels requiring validation rather than as automatic guarantees of diagnostic or advisory correctness [17].

Fine-tuning and parameter-efficient adaptation can improve the vocabulary and task alignment of agricultural language models. A plant-protection question-answering model adapted through low-rank techniques may respond more consistently within the subject matter represented in its corpus, but its apparent specialization remains conditional on corpus quality, coverage, labelling, and evaluation design [18]. Agriculture-oriented conversational assistants similarly demonstrate how natural-language interfaces may connect users with descriptions of practices or technical concepts, yet an interface demonstration does not establish field reliability, farmer approval, or safe decision performance [19]. Prompting, fine-tuning, retrieval-augmented generation, scenario generation, and explanation generation may therefore support carefully bounded advisory workflows, provided that outputs are presented as reviewable candidates, uncertainty is communicated, sources can be inspected, and qualified humans retain authority. No agronomic-advice claim or autonomous-decision claim should be inferred from fluent text, domain adaptation, retrieved evidence, or conversational usability alone.

Model Reliability, Hallucination, and Validation Risks

Direct evaluation of agricultural language-model outputs demonstrates why fluency must not be used as a proxy for reliability. When ChatGPT responses to farmers' questions about irrigated lowland rice cultivation were assessed against domain knowledge, answer quality varied in correctness, completeness, specificity, and contextual suitability. Some responses could support preliminary information access, but others required correction or qualification before they could be considered useful within an extension setting. The central reliability issue is therefore not whether a model can

produce grammatically coherent agricultural text, but whether each substantive claim is factually accurate, locally relevant, appropriately bounded, and supported by sufficient evidence. Crop-specific evaluation by qualified experts remains necessary because a generally plausible answer may omit decisive details concerning production ecology, management history, input availability, or regional practice [20].

Similar caution applies when generative models are evaluated for plant-disease risk communication. Comparative assessments of ChatGPT outputs indicate that model versions may differ in concision, contextual detail, linguistic consistency, and the organization of explanatory content, yet these differences do not independently establish agronomic validity. A response may be well structured while failing to capture disease epidemiology, host susceptibility, environmental thresholds, or the uncertainty inherent in forecasting. Domain adaptation may improve terminology and task alignment, but deployment claims require more than stylistic or lexical evaluation. Expert review, source verification, explicit uncertainty communication, and testing against realistic disease-risk scenarios are needed before generated material can be incorporated into consequential plant-protection workflows [21].

Benchmarking is necessary for reproducible evaluation, but agricultural benchmarks must reflect the intended decision context rather than merely replicate generic language-model tests. The AgXQA benchmark illustrates an effort to construct question-answering tasks around agricultural extension knowledge, thereby enabling more relevant comparison of model capabilities. Even domain-specific benchmarks, however, capture only a bounded selection of questions, crops, regions, languages, difficulty levels, and answer forms. Strong performance on a static benchmark may coexist with weak robustness under paraphrasing, missing context, contradictory evidence, unfamiliar cultivars, or changing environmental conditions. Benchmark scores should therefore be interpreted as evidence about performance on a specified test distribution, not as proof of broad reliability, agronomic competence, or readiness for field use [22].

Hallucination is a broader family of generation failures in which model outputs contain unsupported, internally inconsistent, fabricated, or factually inaccurate content. It can arise from limitations in training data, model representations, decoding behaviour, prompting, or weak alignment between the user's question and the available evidence. Retrieval-augmented generation may reduce exclusive dependence on model memory by introducing external documents, but it also creates additional failure points involving corpus quality, retrieval relevance, document freshness, ranking, evidence selection, citation fidelity, and synthesis. A model may retrieve an authoritative source and still misrepresent it, omit a decisive qualification, or combine unrelated passages into an unsupported conclusion. Consequently, neither fluent generation nor retrieval grounding guarantees agricultural factuality; hallucination and retrieval-generation failures require component-level evaluation, source traceability, and continuing monitoring [23, 24].

Reliability risks also propagate across synthetic-data and advisory workflows. Poor-quality or unrepresentative agricultural inputs can shape synthetic images, model adaptation, retrieved evidence, and generated explanations. Synthetic-real domain gaps may then weaken downstream recognition, while locally inappropriate knowledge can contribute to factual or agronomic errors in advisory text. Bias, inadequate subgroup representation, prompt sensitivity, uncertainty underreporting, weak explainability, and human automation bias can compound these failures. Validation must therefore include source-data audits, realism and biological-plausibility assessment, domain-specific benchmarks, robustness testing, external validation, local review, uncertainty communication, and explicit human authority. A research prototype or favourable benchmark result does not establish reliability, and a human reviewer cannot function as an effective safeguard without sufficient expertise, time, evidence access, and power to reject the output. **Figure 1** illustrates the risk pathway linking generative AI inputs, synthetic data, advisory outputs, hallucination, validation gaps, human oversight, and deployment boundaries in precision agriculture. **Table 2** maps generative decision-support functions, model reliability concerns, hallucination risks, and validation requirements for precision agriculture.

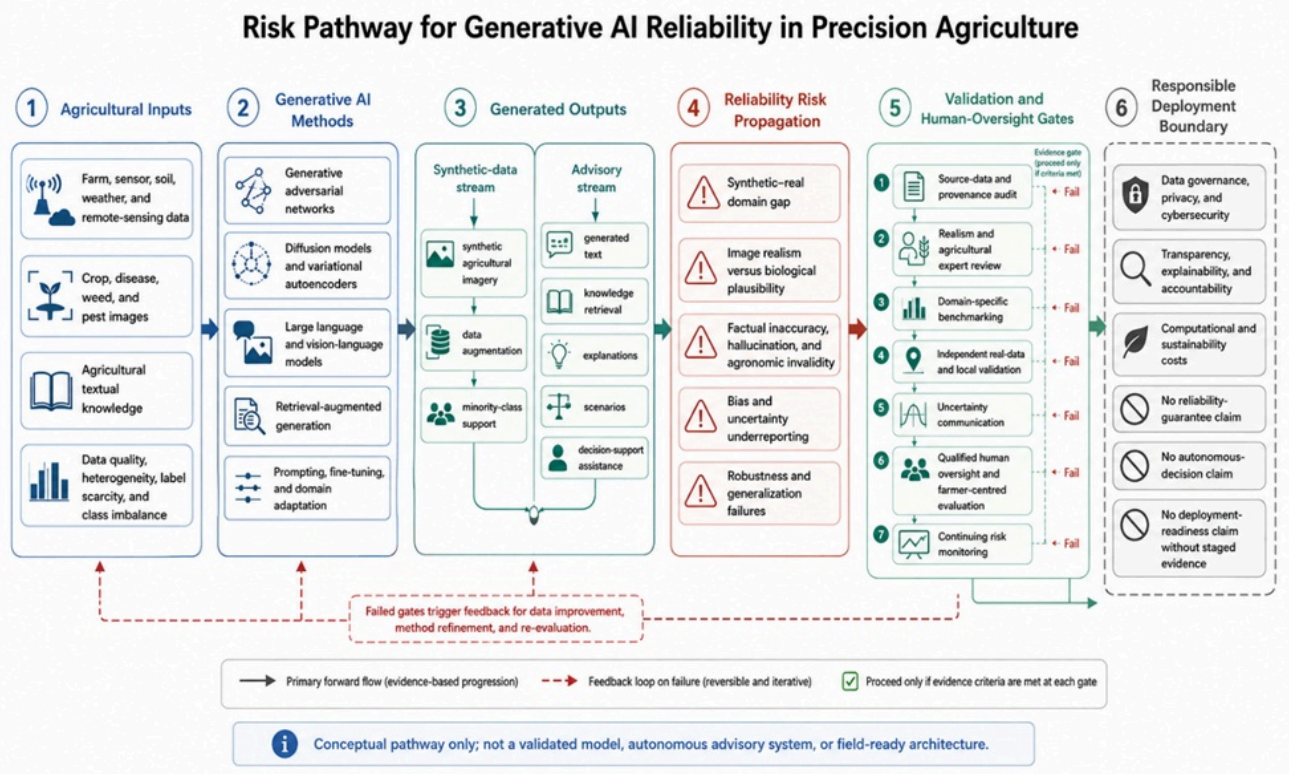


Figure 1. Risk Pathway for Generative AI Reliability in Precision Agriculture

Alt-text description: A conceptual risk pathway showing agricultural data and generative AI models producing synthetic data and advisory outputs, with risks including hallucination, factual inaccuracy, domain gaps, bias, uncertainty, validation gaps, human oversight failures, and deployment-readiness boundaries.

Table 2. Generative Decision Support, Model Reliability, Hallucination, and Validation Risks in Precision Agriculture: Large Language Models, Vision-Language Models, Retrieval-Augmented Generation, Advisory Interfaces, Factual Accuracy, Agronomic Validity, Robustness, Generalization, Uncertainty, Human Oversight, and Deployment Boundaries

Decision-support or reliability component	Core risk question	Required evidence or context	Generative AI function or risk mechanism	Potential output or risk manifestation	Validation or safeguard need	Human oversight need
Large language models	Does the model understand the agricultural question and its local context?	Crop, location, season, production system, recent observations, and validated knowledge	Generates text from learned language patterns and prompts	Coherent but generic, incomplete, outdated, or inaccurate response	Crop-specific factual review, contextual testing, and source verification	Qualified expert review for consequential use
Vision-language models	Does the system interpret visual evidence without creating	Calibrated images, image-quality metadata, field	Connects visual features with generated text	Incorrect symptom description or unsupported visual explanation	Image-quality checks, expert visual confirmation, and	Agronomist, pathologist, botanist, entomologist

	false associations?	context, and expert labels			independent multimodal testing	or veterinarians as appropriate
Retrieval-augmented generation	Are retrieved sources authoritative, current, and relevant to the question?	Curated corpus, versioning, provenance, retrieval logs, and source-access records	Retrieves documents before generating a response	Answer grounded in irrelevant, stale, incomplete, or conflicting evidence	Retriever evaluation, citation-fidelity testing, corpus governance, and claim-level verification	Reviewer must be able to inspect the retrieved evidence
Prompting	Is model behaviour stable across equivalent instructions and user phrasing?	Prompt variants, user-language diversity, adversarial cases, and task definitions	Conditions generation without changing core parameters	Prompt-sensitive or inconsistent recommendations	Prompt-robustness testing, structured input requirements, and red-team evaluation	Users need clear escalation when information is missing
Fine-tuning	Does adapted performance extend beyond the fine-tuning corpus?	Representative domain corpus, held-out tasks, and baseline comparisons	Updates model parameters for agricultural terminology or tasks	Improved in-domain answers with overfitting or inherited corpus errors	External evaluation, regression testing, and corpus-quality audit	Domain experts must review training content and failures
Domain adaptation	Does the model remain valid across regions, crops, languages, and practices?	Regional data, local terminology, management systems, and environmental variation	Adjusts representations or model behaviour to a target domain	Poor performance outside the adapted distribution	Geographic, temporal, linguistic, and production-system validation	Local institutions and users should participate in evaluation
Knowledge retrieval	Can each generated claim be traced to suitable agricultural evidence?	Structured databases, knowledge graphs, extension materials, and provenance	Searches and ranks candidate knowledge	Unsupported synthesis or omission of decisive qualifications	Source-level relevance testing and evidence-to-claim mapping	Human review for high-impact interpretation
Text generation	Does wording preserve factual content, limitations, and uncertainty?	Verified facts, communication objectives, and user-	Composes summaries, explanations, and answers	Fabricated detail, excessive certainty, or omitted caveats	Claim verification, uncertainty-language testing, and	Human approval before consequential communication

		comprehension requirements			editorial review	
Scenario generation	Are generated scenarios explicit conditional possibilities rather than forecasts?	Defined assumptions, ranges, causal constraints, and uncertainty information	Produces hypothetical or conditional narratives	Plausible scenario interpreted as a prediction	Assumption disclosure, plausibility review, and sensitivity analysis	Advisor interpretation and user clarification
Advisory interface	Does the interface collect sufficient context and permit correction or challenge?	User needs, literacy, language, connectivity, and interaction records	Conducts conversational elicitation and presentation	Premature response, omitted context, or automation bias	Usability testing, clarification requirements, and contestability controls	Users must be able to reject and escalate output
Farm decision support	Does generated information improve a defined decision process without displacing authority?	Decision objective, alternatives, constraints, observations, and accountable roles	Summarizes evidence or compares candidate options	Incomplete decision framing or unsupported prioritization	Workflow evaluation, error-impact assessment, and decision logging	Farmer and qualified advisor retain final authority
Factual accuracy	Is each statement correct, current, and appropriately sourced?	Authoritative evidence and date-sensitive knowledge	Failure may arise from stale parameters or incorrect synthesis	Wrong facts, units, dates, causal claims, or source attributions	Claim-level checking and version-controlled evidence	Expert review where errors could affect management
Agronomic validity	Is the information suitable for the specific biological and production context?	Soil, crop stage, cultivar, weather, management history, and constraints	Generic language may omit decisive local variables	Generally plausible but locally unsuitable advice	Local expert assessment and context-specific evaluation	Agronomist or extension professional review
Synthetic-real domain gap	Does synthetic-supported performance transfer to real observations?	Independent farms, seasons, devices, cultivars, and environmental conditions	Generated samples differ from the target real distribution	Performance decline under real field conditions	Real-only external, temporal, and geographic testing	Human review cannot compensate for unmeasured model shift
Hallucination	Does the system create	Expert-verified questions,	Parametric generation, weak	Fabricated facts, sources,	Agricultural hallucination	Reviewers need evidence

	unsupported agricultural information?	evidence sets, and error taxonomy	grounding, or decoding failure	symptoms, causal links, or practices	measurement and claim verification	access and stopping authority
Data-quality failure	Are source inputs accurate, representative, labelled, and traceable?	Provenance, metadata, quality controls, and leakage checks	Model learns and reproduces defects in input data	Confident outputs derived from biased, noisy, or incomplete information	Upstream data audit and downstream error analysis	Data steward and domain experts share responsibility
Bias	Are errors systematically concentrated in particular crops, regions, languages, or farm types?	Representative subgroup data and stratified evaluation	Unequal representation in training or retrieval sources	Lower-quality outputs for underrepresented contexts	Subgroup testing, dataset review, and continuing monitoring	Affected users should participate in defining relevant groups
Fairness concern	Are benefits, burdens, errors, and access distributed acceptably?	User, farm, regional, linguistic, and infrastructure context	Technical and structural disparities interact	Exclusion, unequal error exposure, or concentrated benefits	Distributional assessment and participatory evaluation	Governance bodies must address structural as well as model-level issues
Robustness failure	Does output change under benign variation in prompts, images, formats, or measurements?	Perturbation sets and realistic input variation	Sensitivity to wording, noise, compression, or missing fields	Inconsistent explanations or classifications	Prompt, corruption, sensor, and formatting stress tests	Humans require warnings when inputs are outside validated conditions
Generalization failure	Does performance extend beyond development data and benchmark tasks?	Unseen regions, seasons, farms, cultivars, sensors, and question forms	Overfitting or narrow representation	Strong benchmark performance with weak real-context performance	External and temporal validation by independent evaluators	Local reviewers should confirm applicability
Uncertainty underreporting	Does linguistic confidence exceed evidence strength?	Calibration data, ambiguous cases, and confidence labels	Generation objectives favour complete-sounding answers	Definitive wording despite incomplete or conflicting evidence	Calibration, abstention, uncertainty-language, and user-interpretation testing	Humans need clear escalation routes for uncertain cases
Explainability gap	Does the explanation accurately	Traceable sources, model records, and	Generates plausible rationales that	False reassurance or	Evidence-linked explanations	Reviewers must be able to inspect

	expose evidence, assumptions, and limitations?	user requirements	may not reflect actual evidence use	misleading causal narrative	and explanation-fidelity testing	underlying sources
Benchmark limitation	Does the test represent intended agricultural use and consequences?	Documented task coverage, contamination checks, and diverse cases	Static test design may omit operational variability	High average score with unmeasured subgroup or workflow failure	Multi-domain benchmarks and transparent limitation statements	Independent interpretation of benchmark relevance
External validation gap	Has performance been reproduced outside the development team or dataset?	Independent evaluators and distinct data sources	Developer-centred evaluation may preserve hidden assumptions	Uncertain reproducibility and transfer	Independent replication and transparent reporting	Deployment institutions should require external evidence
Field evaluation gap	Has the system been evaluated within realistic agricultural workflows?	Prospective tasks, accountable users, monitoring, and predefined stopping rules	Offline testing excludes workflow and consequence effects	Unknown usability, workload, trust, and error impact	Staged, supervised, context-specific field evaluation planning	Qualified users should retain decision authority throughout testing
Human oversight failure	Can human reviewers detect, challenge, correct, and reject model errors?	Role clarity, expertise, time, evidence access, and override mechanisms	Automation bias or inadequate interface design	Rubber-stamping, delayed correction, or responsibility diffusion	Human-factors testing, escalation design, logging, and audit	A named competent person must hold final authority
No reliability-guarantee boundary	Do selected evaluations justify describing the system as reliable?	Evidence across factuality, robustness, generalization, uncertainty, and context	Overgeneralization from averages or demonstrations	Unconditional reliability language	Formal claim audit and context-specific limitation reporting	Authors, reviewers, and deployers share responsibility
No deployment-readiness boundary	Do technical demonstrations establish operational suitability?	Integrated technical, human, organizational, and governance evidence	Conflation of feasibility with readiness	Premature field, commercial, or autonomous-use claim	Staged readiness assessment, independent review, and stop criteria	Institutions require explicit authorization and accountability

Responsible Deployment Challenges

Responsible deployment requires explicit allocation of responsibility across the generative-AI lifecycle. Agricultural systems may involve data providers, model developers, platform operators, knowledge curators, extension organizations, farm-management vendors, advisors, farmers, researchers, and public institutions. When an inaccurate output contributes to a poor decision, responsibility can become diffuse if each actor treats another component or organization as the source of authority. Agricultural AI governance must therefore specify who is responsible for data quality, model evaluation, retrieval-corpus maintenance, user communication, monitoring, incident response, model updating, and the authorization of consequential use. Trustworthy deployment cannot be achieved by assigning responsibility to the model itself or by assuming that the final user can identify every hidden technical failure [25].

Ethical and responsible agricultural AI also requires attention to privacy, transparency, fairness, accountability, and the conditions under which data are reused. Generative systems may process farm records, sensor streams, images, geospatial information, production histories, user questions, and other context that can reveal commercially or personally sensitive information. Retrieval systems and conversational interfaces introduce further risks through data retention, unintended disclosure, prompt injection, insecure integrations, or secondary use beyond the original purpose. Bias can arise from unequal representation of crops, regions, languages, technologies, and farm structures, while opaque model updating can alter behaviour without adequate revalidation. Responsible practice consequently requires provenance, access controls, data minimization, security testing, subgroup evaluation, documentation, change control, and mechanisms for audit and redress [26].

Farmer trust must be treated as an empirical and relational outcome rather than an automatic consequence of technical performance. Research on smart-farming data practices indicates that trust is weakened when farmers do not understand what data are collected, how they are processed, who benefits, or whether information may be reused beyond the stated purpose. Generative AI may intensify these concerns because derived embeddings, synthetic records, generated summaries, and conversational logs can be difficult for users to inspect. Transparency should therefore include understandable explanations of system purpose, source-data categories, knowledge provenance, limitations, update practices, and responsible actors. Benefit sharing, contestability, meaningful consent, and the ability to refuse or discontinue participation are also relevant. Farmer-centred evaluation must assess whether the technology supports users' goals and agency rather than merely measuring whether they are willing to use it [27].

A responsible-deployment framework must be anticipatory, inclusive, responsive, and reflexive throughout development and use. Anticipation requires examining plausible technical, social, economic, and environmental consequences before large-scale adoption. Inclusion requires participation from farmers, advisors, workers, local institutions, and underrepresented users whose needs may not be visible in development datasets. Responsiveness requires the capacity to revise, suspend, or withdraw a system when evidence changes or harms emerge, while reflexivity requires institutions and developers to examine their own assumptions, incentives, and definitions of success. These principles extend to computational and sustainability costs, digital inclusion, local validation, documentation, risk monitoring, and cybersecurity. Because safeguards remain context-dependent and incomplete, their presence cannot support an unconditional responsible-deployment guarantee or a field-readiness claim [28].

Critical Synthesis

The reviewed evidence reveals a persistent imbalance between the breadth of proposed generative-AI opportunities and the maturity of validation supporting them. Generative adversarial networks, diffusion models, variational autoencoders, large language models, vision-language models, and retrieval-supported systems may contribute to synthetic-data experiments, agricultural image augmentation, knowledge access, explanation generation, and conversational decision assistance. Yet responsible agricultural innovation requires more than technical novelty or performance on bounded datasets. It requires attention to who defines the problem, which farmers and production systems are represented, how benefits and burdens are distributed, and whether affected stakeholders can influence development and use. Agriculture

4.0 scholarship therefore supports a broader interpretation in which social purpose, inclusion, and accountability are integral to assessing technological value [29].

The synthesis also distinguishes user acceptance from evidence of model validity. Perceived usefulness, ease of use, compatibility, social expectations, and behavioural intention can influence whether agricultural users adopt AI-enabled tools. These factors are important because systems that are inaccessible, confusing, or poorly aligned with farm workflows are unlikely to support meaningful decisions. Nevertheless, a system may be accepted because it is convenient, novel, persuasive, or institutionally promoted while still producing inaccurate or weakly contextualized outputs. Conversely, a technically valid research tool may remain unacceptable because it imposes excessive workload, lacks transparency, or conflicts with user priorities. Farmer-centred evaluation must therefore examine calibrated trust and appropriate reliance rather than aiming simply to maximize adoption [30].

Computational cost represents another area where opportunity claims exceed available agricultural evidence. Generative models can require substantial resources for training, adaptation, retrieval, and repeated inference, and these costs may affect affordability, latency, infrastructure dependence, and access for smaller organizations or low-resource farming communities. Accuracy-centred evaluation can obscure energy use, hardware demand, financial cost, and the opportunity cost of pursuing increasingly large models when smaller or task-specific alternatives may be sufficient. Green AI principles support reporting efficiency and resource consumption alongside performance, but agriculture-specific evidence remains limited. Claims that generative AI contributes to environmental sustainability should therefore distinguish potential agricultural benefits from the unmeasured environmental burdens of the digital system itself [31]

Across synthetic data and advisory systems, the strongest evidence supports bounded research and assistance functions rather than autonomous agricultural intelligence. Synthetic crop, disease, weed, and pest images may help investigate class imbalance and label scarcity, but their utility depends on source-data quality, biological plausibility, diversity, and transfer to independent real observations. Language and multimodal models may improve access to agricultural knowledge or generate candidate explanations, but reliability depends on factual verification, local context, evidence provenance, uncertainty communication, and human review. Retrieval, fine-tuning, prompting, and domain adaptation can modify performance but do not remove hallucination, bias, robustness, or generalization risks. The central evidence gap is therefore not a lack of possible applications; it is the absence of sufficiently integrated validation connecting technical output quality to agronomic validity, human decision processes, governance, and monitored use.

A cautious research trajectory should link opportunity assessment to explicit validation and governance gates. Promising technical functions should first be evaluated under controlled conditions, followed by independent real-data tests, subgroup analysis, local contextual review, human-factors assessment, and monitored field-evaluation planning where justified. Data governance, privacy, cybersecurity, accountability, documentation, computational efficiency, digital inclusion, and farmer participation must be considered as interacting conditions rather than separate ethical additions. Research priorities should focus on measurable weaknesses: synthetic-real domain gaps, agricultural hallucination, uncertainty communication, retrieval fidelity, underrepresented crops and regions, effective human oversight, calibrated trust, and lifecycle monitoring. **Figure 2** synthesises the critical review findings by connecting generative AI methods, precision agriculture applications, opportunity pathways, risk domains, validation requirements, responsible deployment safeguards, and future research priorities.



Future Research Agenda

Synthetic-data research requires stronger methods for assessing realism, biological plausibility, and synthetic-real transfer. Future studies should combine perceptual and statistical measures with crop-, disease-, weed-, pest-, livestock-, and remote-sensing-specific expert criteria. Generated samples should be tested through controlled augmentation studies using real-only independent test sets, multiple environments, repeated training, subgroup metrics, calibration, and negative-result reporting. Domain-gap analysis should examine changes across farms, seasons, cultivars, sensors, image-acquisition protocols, and environmental conditions. Research should also compare generative adversarial networks, diffusion models, variational autoencoders, and non-generative augmentation methods under transparent resource budgets. These studies would clarify when synthetic samples add useful diversity, when they reproduce source bias, and when they create visually plausible but agronomically misleading artefacts.

Generative decision-support research should develop agricultural hallucination taxonomies, factuality tests, retrieval-quality metrics, and uncertainty-aware interaction designs. Evaluation should distinguish errors caused by model memory, prompting, retrieved evidence, corpus age, source ranking, multimodal interpretation, and unsupported synthesis. Retrieval-augmented agricultural systems require versioned corpora, evidence provenance, claim-to-source traceability, and tests of whether citations actually support generated statements. Human-in-the-loop studies should measure whether farmers and advisors can detect errors, interpret uncertainty, challenge outputs, and avoid automation bias under realistic workloads. Farmer-centred usability and trust research should examine comprehension, contestability, language, accessibility, benefit distribution, and calibrated reliance rather than treating satisfaction or adoption intention as proof of decision quality.

Responsible-deployment research should integrate privacy-preserving methods, cybersecurity testing, bias and fairness assessment, sustainable-computing measurement, governance documentation, model-update controls, incident reporting, and post-release risk monitoring. Proposed documentation standards should record data provenance, generation methods, retrieval sources, prompt structures, fine-tuning procedures, evaluation contexts, known limitations, subgroup performance, computational requirements, and responsible actors. Field-readiness assessment should use staged, reversible, and independently reviewed evidence gates rather than treating a benchmark, prototype, or pilot as sufficient. Local validation, farmer participation, qualified human authority, suspension procedures, and continuing monitoring should remain mandatory conditions for any progression toward consequential use. **Table 3** presents a future research agenda for responsible generative AI in precision agriculture.

Table 3. Future Research Agenda for Responsible Generative AI in Precision Agriculture: Agricultural Benchmarks, Synthetic Data Realism, Domain Gap Evaluation, Domain Adaptation, Hallucination Measurement, Knowledge Retrieval Validation, Uncertainty-Aware Decision Support, Human-in-the-Loop Evaluation, Farmer Trust, Bias, Privacy, Sustainable Computing, Governance, and Field-Readiness Boundaries

Future research priority	Problem addressed	Required methodological advance	Required data or infrastructure	Validation need	Governance or user need	Expected contribution
Agricultural generative-AI benchmarks	Generic benchmarks omit agricultural context, multimodality, and consequence sensitivity	Multi-task, multilingual, multimodal, contamination-aware benchmark design	Versioned agricultural questions, images, metadata, evidence sources, and expert answers	Independent testing across crops, regions, languages, and difficulty levels	Transparent benchmark governance and access procedures	More relevant and comparable model evaluation
Synthetic-data realism assessment	Visual plausibility is often treated as evidence of validity	Multi-dimensional technical, perceptual, biological, and task-based realism framework	Paired real and synthetic data with expert annotations and provenance	Cross-site assessment, expert review, and real-only downstream testing	Disclosure of generation, filtering, and rejection criteria	Clearer identification of usable and misleading synthetic samples
Synthetic-real domain-gap evaluation	Synthetic-supported performance may not	Geographic, temporal, sensor, cultivar, and	Multi-farm, multi-season, multi-device real datasets	External and out-of-distribution validation	Data-sharing safeguards and documentation	Better evidence on transferability

	transfer to real agricultural environments	environmental domain-shift protocols				
Domain-specific generative-model adaptation	Generic models may lack regional terminology, practices, and biological knowledge	Comparative evaluation of prompting, fine-tuning, retrieval, and modular adaptation	Curated crop-, region-, language-, and production-specific corpora	Held-out regions, tasks, and temporal periods	Local expert stewardship and version control	Improved contextual relevance
Low-resource crop and region representation	Dominant crops and well-resourced regions shape model behaviour	Purposeful sampling, participatory data design, and representation audits	Multilingual and underrepresented-crop data with contextual metadata	Stratified evaluation and distributional error analysis	Fair data partnerships and benefit-sharing arrangements	Reduced visibility and performance gaps
Hallucination measurement in agricultural advisory systems	Agricultural factuality failures lack standardized definitions and tests	Claim-level taxonomy covering fabrication, contradiction, omission, and invalid synthesis	Expert-verified questions, evidence sets, and adversarial cases	Repeated testing across models, prompts, crops, and regions	Incident-reporting and escalation procedures	Measurable agricultural hallucination risk
Retrieval-augmented agricultural knowledge validation	Retrieval can introduce irrelevant, stale, or conflicting evidence	Component-level evaluation of corpus, retriever, ranking, citation, and generator	Versioned authoritative corpora with provenance and update records	Retrieval relevance, answer correctness, citation fidelity, and freshness testing	Corpus ownership, maintenance, and correction policy	Stronger evidence-grounding practices
Uncertainty-aware generative decision support	Generated answers may sound more certain than available evidence	Calibration, abstention, ambiguity detection, and uncertainty-language design	Labelled uncertain, conflicting, and insufficient-context cases	Calibration and user-interpretation evaluation	Safe escalation and non-response options	More appropriate user reliance
Human-in-the-loop advisory evaluation	Human oversight is often asserted without evidence of effectiveness	Controlled human-factors and workflow experiments	Realistic advisory tasks, evidence displays, logs, and error scenarios	Error detection, correction, workload, and automation-bias measurement	Clear authority, training, and stopping rights	Evidence on when oversight reduces or amplifies risk
Farmer-centred usability and	Technical evaluation omits farmer	Participatory, longitudinal, multilingual, and	Diverse farmer panels and locally relevant	Comprehension, burden, contestability,	Feedback rights, accessibility,	More legitimate and usable

trust evaluation	goals, comprehension, and decision practices	mixed-method studies	interaction scenarios	trust calibration, and sustained-use assessment	and benefit sharing	system design
Bias and fairness assessment	Model quality and access may vary across farms, crops, languages, and regions	Intersectional subgroup testing and distributional evaluation	Representative user, farm, crop, regional, and infrastructure data	Per-group factuality, robustness, access, and error-impact assessment	Community-defined fairness objectives and remediation processes	Visibility of unequal performance and benefits
Privacy-preserving generative AI	Context-rich agricultural systems may expose sensitive farm or user data	Data minimization, federated methods, confidential processing, and leakage testing	Secure infrastructure, threat models, and controlled data-access environments	Utility-privacy trade-off and re-identification evaluation	Meaningful consent, access control, deletion, and secondary-use restrictions	Reduced unnecessary data exposure
Sustainable-computing evaluation	Energy, hardware, financial, and access costs are underreported	Workload-normalized and lifecycle-oriented resource measurement	Energy, latency, hardware, utilization, and model-size records	Replication across training, adaptation, retrieval, and inference settings	Transparent resource and cost disclosure	Evidence-based comparison of technical alternatives
Responsible-deployment monitoring	Model behaviour, knowledge, and user context change after release	Drift detection, incident taxonomy, feedback analysis, rollback, and suspension protocols	Version control, secure logs, monitoring dashboards, and reporting channels	Sensitivity of monitoring and effectiveness of corrective action	Independent oversight and stop authority	Earlier detection of deterioration and misuse
Governance and documentation standards	Assurance information is inconsistent across agricultural generative systems	Agriculture-specific minimum reporting and audit framework	Records covering data, models, prompts, retrieval, evaluation, users, and updates	Auditability, completeness, and inter-reviewer consistency testing	Multi-stakeholder development and periodic revision	Comparable evidence packages and clearer accountability
Field-readiness assessment without overclaiming	Prototypes and benchmark results may be misrepresented as deployment evidence	Staged readiness levels integrating technical and socio-technical criteria	Controlled sandboxes, supervised pilots, local expertise, and independent evaluation	Prospective local effectiveness, reliability, usability, and safety assessment	Farmer participation, human authorization, exit routes, and incident response	More defensible decisions about progression or suspension

Conclusion

This critical review has synthesized generative artificial intelligence in precision agriculture across synthetic data, agricultural image generation, decision support, advisory systems, model reliability, hallucination, validation risks, responsible deployment challenges, and future research priorities. Generative methods may support carefully bounded research, augmentation, knowledge-access, explanation, and decision-assistance workflows, but their outputs do not independently establish biological plausibility, agronomic correctness, farmer trust, environmental benefit, operational safety, or deployment readiness. Synthetic data require source-quality assessment, realism testing, biological review, and independent real-domain validation. Advisory outputs require factual verification, uncertainty communication, local contextualization, and qualified human authority. Responsible use further depends on farmer-centred evaluation, data governance, privacy and cybersecurity protections, accountability, transparency, bias monitoring, computational assessment, documentation, model-update control, and continuing risk monitoring. Future progress should therefore be judged not only by the ability to generate convincing content, but by whether evidence supports reliable, equitable, traceable, context-sensitive, and reversible use within clearly defined boundaries. The appropriate near-term contribution of generative AI is to support critical agricultural research and supervised decision processes rather than to replace farmers, advisors, agronomists, or established validation practices.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 17 Jul 2025

Revised: 09 Sep 2025

Accepted: 02 Nov 2025

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

Pallottino F, Violino S, Figorilli S, Pane C, Aguzzi J, Colle G, et al. Applications and perspectives of generative artificial intelligence in agriculture. *Comput Electron Agric.* 2025;230:109919.
<https://doi.org/10.1016/j.compag.2025.109919>.

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>.

Sai S, Kumar S, Gaur A, Goyal S, Chamola V, Hussain A. Unleashing the power of generative AI in Agriculture 4.0 for smart and sustainable farming. *Cogn Comput.* 2025;17:63.
<https://doi.org/10.1007/s12559-025-10420-6>.

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>.

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>.

Shaikh TA, Rasool T, Mir WA. Fields of the future: Digital transformation in smart agriculture with large language models and generative AI. *Comput Stand Interfaces.* 2025;94:104005.
<https://doi.org/10.1016/j.csi.2025.104005>.

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>.

Min X, Ye Y, Xiong S, Chen X. Computer vision meets generative models in agriculture: Technological advances, challenges and opportunities. *Appl Sci.* 2025;15(14):7663.
<https://doi.org/10.3390/app15147663>.

Akkem Y, Biswas SK, Varanasi A. A comprehensive review of synthetic data generation in smart farming by using variational autoencoder and generative adversarial network. *Eng Appl Artif Intell.* 2024;131:107881.
<https://doi.org/10.1016/j.engappai.2024.107881>.

Lu Y, Chen D, Olaniyi E, Huang Y. Generative adversarial networks for image augmentation in agriculture: A systematic review. *Comput Electron Agric.* 2022;200:107208.
<https://doi.org/10.1016/j.compag.2022.107208>.

Nazki H, Yoon S, Fuentes A, Park DS. Unsupervised image translation using adversarial networks for improved plant disease recognition. *Comput Electron Agric.* 2020;168:105117.
<https://doi.org/10.1016/j.compag.2019.105117>.

Abbas A, Jain S, Gour M, Vankudothu S. Tomato plant disease detection using transfer learning with C-GAN synthetic images. *Comput Electron Agric.* 2021;187:106279.
<https://doi.org/10.1016/j.compag.2021.106279>.

Li T, Asai M, Kato Y, Fukano Y, Guo W. Channel attention GAN-based synthetic weed generation for precise weed identification. *Plant Phenomics.* 2024;6:0122.
<https://doi.org/10.34133/plantphenomics.0122>.

Du M, Wang F, Wang Y, Li K, Hou W, Liu L, et al. Improving long-tailed pest classification using diffusion model-based data augmentation. *Comput Electron Agric.* 2025;234:110244.
<https://doi.org/10.1016/j.compag.2025.110244>.

Tzachor A, Devare M, Richards C, Pypers P, Ghosh A, Koo J, et al. Large language models and agricultural extension services. *Nat Food*. 2023;4(11):941-8.

<https://doi.org/10.1038/s43016-023-00867-x>.

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>.

Zhao X, Chen B, Ji M, Wang X, Yan Y, Zhang J, et al. Implementation of large language models and agricultural knowledge graphs for efficient plant disease detection. *Agriculture*. 2024;14(8):1359.

<https://doi.org/10.3390/agriculture14081359>.

Xiong J, Pan L, Liu Y, Zhu L, Zhang L, Tan S. Enhancing plant protection knowledge with large language models: A fine-tuned question-answering system using LoRA. *Appl Sci*. 2025;15(7):3850.

<https://doi.org/10.3390/app15073850>.

Ravindran DJS, Skarga-Bandurova I, Sivakumar V, Awais M, Mithra S. AgroLLM: Connecting farmers and agricultural practices through large language models for enhanced knowledge transfer and practical application. *AgriEngineering*. 2026;8(1):38.

<https://doi.org/10.3390/agriengineering8010038>.

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

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

Calone R, Raparelli E, Bajocco S, Rossi E, Crecco L, Morelli D, et al. Analysing the potential of ChatGPT to support plant disease risk forecasting systems. *Smart Agric Technol*. 2025;10:100824.

<https://doi.org/10.1016/j.atech.2025.100824>.

Kpodo J, Kordjamshidi P, Nejadhashemi AP. AgXQA: A benchmark for advanced agricultural extension question answering. *Comput Electron Agric*. 2024;225:109349.

<https://doi.org/10.1016/j.compag.2024.109349>.

Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Trans Inf Syst*. 2025;43(2):42.

<https://doi.org/10.1145/3703155>.

Zhao P, Zhang H, Yu Q, Wang Z, Geng Y, Fu F, et al. Retrieval-augmented generation for AI-generated content: A survey. *Data Sci Eng*. 2026;11:1-29.

<https://doi.org/10.1007/s41019-025-00335-5>.

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>.

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>.

Jakku E, Taylor B, Fleming A, Mason C, Fielke S, Sounness C, et al. "If they don't tell us what they do with it, why would we trust them?" Trust, transparency and benefit-sharing in smart farming. *NJAS Wageningen J Life Sci*. 2019;90-91:100285.

<https://doi.org/10.1016/j.njas.2018.11.002>.

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>.

Rose DC, Chilvers J. Agriculture 4.0: Broadening responsible innovation in an era of smart farming. *Front Sustain Food Syst*. 2018;2:87.
<https://doi.org/10.3389/fsufs.2018.00087>.

Mohr S, Kühl R. Acceptance of artificial intelligence in German agriculture: An application of the technology acceptance model and the theory of planned behavior. *Precis Agric*. 2021;22(6):1816-44.
<https://doi.org/10.1007/s11119-021-09814-x>.

Schwartz R, Dodge J, Smith NA, Etzioni O. Green AI. *Commun ACM*. 2020;63(12):54-63.
<https://doi.org/10.1145/3381831>.

Zhang W, Yadav A. Generative AI as an interface between farmers and AI algorithms: Opportunities and challenges. *J Indian Inst Sci*. 2025;105(2):305-21.
<https://doi.org/10.1007/s41745-025-00490-8>.