

ORIGINAL RESEARCH

Open access

Trustworthy AI Certification for Agri-Food Systems: A Framework for Auditability, Safety, Fairness, and Sustainability Claims

Amal Al-Balushi^{1*}, Saif Al-Hinai¹

Abstract

Agri-food systems increasingly rely on artificial intelligence for farm decision support, automation, quality control, sustainability monitoring, logistics, supply-chain coordination, risk prediction, and resource management. These applications promise productivity gains and improved responsiveness to environmental and market pressures. Yet their growing influence also creates new governance demands because AI systems can shape decisions that affect farms, firms, workers, consumers, ecosystems, and public institutions. Trustworthy AI in agri-food systems cannot be established through technical performance claims alone. A model may be accurate in a narrow test environment while remaining opaque, unsafe under changing field conditions, unfair across producer groups, or unable to substantiate sustainability claims. Certification therefore needs to move beyond reputational signalling and provide credible assurance that AI systems can be examined, challenged, monitored, and held accountable. This article develops a trustworthy AI certification framework for agri-food systems centred on auditability, safety, fairness, and sustainability-claim verification. The framework treats certification as a governance mechanism rather than a simple label or checklist. It identifies the evidence, assurance procedures, monitoring expectations, and institutional safeguards needed to make certification credible in technically complex and socially consequential agri-food settings. The article contributes a structured approach for regulators, certification bodies, technology providers, agri-food firms, farmer organisations, sustainability auditors, and public institutions. It argues that trustworthy AI certification should depend on verifiable evidence, independent auditability, lifecycle risk assurance, distributional accountability, environmental claim verification, and clear responsibilities across the AI value chain. Certification becomes meaningful only when it supports systems that are auditable, safe, fair, environmentally credible, and publicly accountable.

Keywords Agri-food systems, Safety assurance, Trustworthy AI, AI certification, Auditability, Fairness

*Correspondence:

Amal Al-Balushi
amal.balushi@squ.edu.om

¹ Department of Agricultural and Marine Sciences, College of Agricultural and Marine Sciences, Sultan Qaboos University, Muscat, Oman

Introduction

Artificial intelligence is becoming embedded in agri-food systems through digital agriculture, smart farming platforms, predictive analytics, sensor-based monitoring, robotics, food-quality inspection, sustainability dashboards, logistics optimisation, and decision-support tools. These systems are not isolated technical instruments; they are increasingly connected to data infrastructures that influence production planning, resource allocation, environmental management, and supply-chain coordination. Because smart farming depends on large-scale data collection, integration, analytics, and

platform-mediated decision support, trustworthy certification must begin with the data-intensive character of the systems being certified [1].

The governance challenge is socio-technical because AI systems in agri-food settings interact with farmers, advisers, technology vendors, food processors, retailers, regulators, consumers, and ecological conditions. Certification that focuses only on model accuracy would miss questions of adoption, control, labour, advisory authority, data ownership, market dependence, and institutional trust. Digital agriculture scholarship shows that smart farming and agriculture-oriented automation raise social, organisational, and political questions that must be addressed alongside technical performance [2].

The need for certification also arises because agri-food AI is often promoted through promises of productivity, efficiency, environmental optimisation, and resilience. These promises matter, but they can become weak governance claims if they are not connected to verifiable evidence and accountability. Agriculture-oriented digital transformation should therefore be assessed against people, production, and planetary goals, making certification a mechanism for aligning technical innovation with social and environmental responsibility [3].

This article develops a trustworthy AI certification framework for agri-food systems based on four interdependent dimensions: auditability and evidence requirements, safety and risk assurance, fairness and distributional accountability, and sustainability-claim verification. The framework responds to a gap between high-level trustworthy AI principles and the institutional mechanisms needed to verify them in practice. Global AI ethics scholarship shows that principles such as transparency, justice, responsibility, privacy, and accountability are widely invoked, but certification requires those principles to be converted into assessable and enforceable governance requirements [4]. **Figure 1** illustrates the proposed certification architecture, showing how agri-food AI systems should move from technical evidence and governance review toward auditable, safe, fair, and sustainability-verified certification.

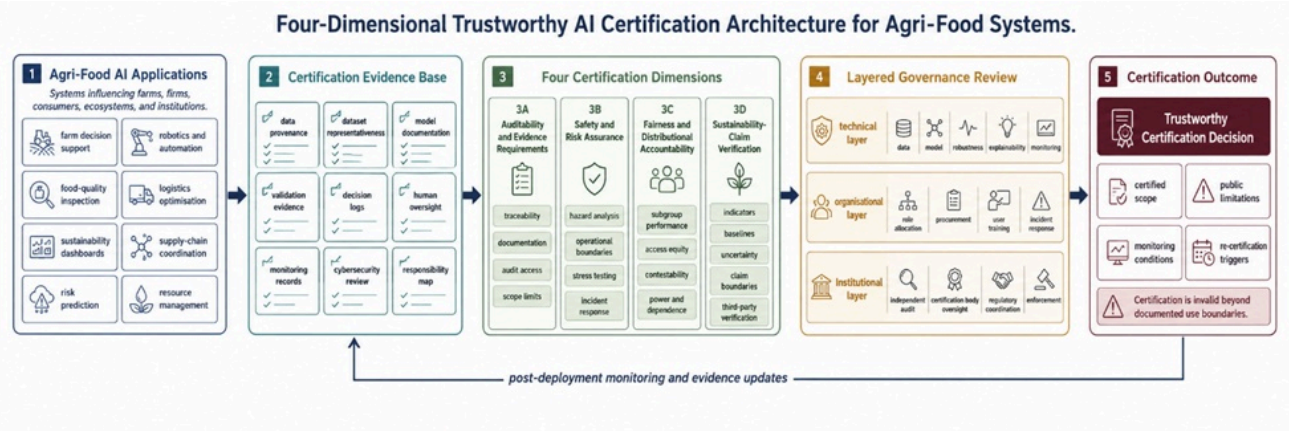


Figure 1. Four-Dimensional Trustworthy AI Certification Architecture for Agri-Food Systems.

Trustworthy AI in Agri-Food Systems

Trustworthy AI in agri-food systems can be defined as AI that is technically reliable, socially accountable, environmentally credible, and institutionally governable. It must be capable of evaluation before deployment, monitoring after deployment, and justification when its outputs influence consequential decisions. A certification framework should therefore translate ethical principles into assessable obligations concerning evidence, oversight, safety, fairness, and accountability rather than treating trust as a general aspiration [5].

Agri-food AI creates distinctive trust challenges because it operates across farms, ecological variability, supply chains, input markets, advisory systems, sustainability claims, and food-sector regulation. Farmers and other users may need to

depend on AI recommendations while lacking full visibility into data sources, model assumptions, update procedures, or platform incentives. Trust in agricultural AI therefore depends on transparency, responsibility, perceived usefulness, risk communication, and fair distribution of benefits and burdens [6].

Trustworthiness also requires attention to impacts that are not captured by narrow efficiency measures. AI systems may affect privacy, autonomy, labour, market access, decision authority, environmental management, and accountability for harm. The agri-food AI literature shows that social and ethical impacts must be included in governance design if certification is to reflect the real contexts in which agricultural AI is developed, procured, deployed, and contested [7].

Certification can help make trustworthy AI more operational, but only if it is evidence-based and context-sensitive. In agri-food systems, responsibility is shared among technology developers, data intermediaries, platform providers, agribusinesses, auditors, public authorities, and users who interpret AI outputs in practice. Responsible AI in agriculture and food-system technology therefore requires institutional trust-building, clear communication, and governance arrangements that clarify who is responsible for what across the AI lifecycle [8].

Auditability and Evidence Requirements

Auditability is the foundation of trustworthy AI certification because it determines whether claims about performance, safety, fairness, and sustainability can be independently examined. Without auditability, certification risks becoming a declaration of intent rather than an assurance process. Principles alone cannot guarantee ethical AI because high-level commitments require implementation structures, evidentiary standards, and accountability mechanisms that make them testable in specific systems [9].

For agri-food AI, certification should require evidence about data provenance, data quality, training and validation conditions, model design assumptions, performance boundaries, human oversight, risk testing, update procedures, monitoring arrangements, and decision logs. These requirements should not be treated as administrative paperwork; they are the core materials through which auditors reconstruct how a system was built, what it was designed to do, and where it may fail. Tools and methods that translate AI ethics into practice show that trustworthy AI depends on practical artefacts, procedures, and organisational routines rather than abstract commitments alone [10].

Auditability must also define the scope and limits of the audit. A certification body should determine whether the audit covers the model, dataset, interface, deployment environment, user training, monitoring system, supply-chain application, or sustainability claim. Ethics-based auditing of automated decision-making systems demonstrates that audits can support accountability only when their nature, scope, evidence base, independence, and limitations are made explicit [11].

In agri-food systems, evidence requirements must extend across farms, platforms, supply chains, sustainability analytics, advisory tools, and automated decision systems. Dataset documentation is especially important because training data may encode geographic, seasonal, crop-specific, socio-economic, or operational limitations that affect downstream performance and fairness. Datasheets for datasets provide a practical model for documenting provenance, composition, collection processes, intended uses, limitations, maintenance, and accountability responsibilities, which are essential for certification evidence in data-dependent agri-food AI [12]. Table 1 summarises auditability and evidence requirements for trustworthy AI certification in agri-food systems.

Table 1. Auditability and Evidence Requirements for Trustworthy AI Certification in Agri-Food Systems: Documentation, Data Traceability, Model Records, Performance Evidence, Monitoring Logs, and Verification Responsibilities

Certification evidence domain	Required audit evidence	Agri-food relevance	Verification responsibility	Governance safeguard

Data provenance and traceability	Source records, collection methods, consent or access basis, geographic and temporal coverage, sensor or platform origin, data-cleaning records	Determines whether datasets represent relevant crops, regions, production systems, farm sizes, seasons, and supply-chain conditions	Technology provider prepares documentation; certification body verifies completeness; data holders confirm access basis	Require traceable records before certification and re-check when datasets are replaced or expanded
Dataset quality and representativeness	Missing-data reports, bias checks, coverage analysis, validation splits, known limitations, excluded populations or production contexts	Helps identify whether the AI system underperforms for smallholders, low-resource regions, unusual crops, extreme weather, or underrepresented supply-chain actors	Model developer documents; independent auditor tests samples; user organisations review contextual adequacy	Require limitations to be disclosed in certification scope and user-facing deployment conditions
Model design and assumptions	Model architecture, feature selection, target variables, optimisation goals, decision thresholds, explainability method, intended use boundaries	Shows whether the system optimises yield, input reduction, quality inspection, logistics, sustainability reporting, or risk prediction in ways that may create trade-offs	Developer provides technical file; auditor reviews against intended use; domain experts assess agronomic plausibility	Link certification validity to declared use cases and prohibit unsupported claims beyond the certified scope
Performance and validation evidence	Accuracy metrics, robustness tests, subgroup performance, stress tests, field validation, uncertainty estimates, comparison against baseline practices	Demonstrates whether the AI system performs under variable field, climate, operational, and supply-chain conditions	Developer conducts testing; auditor verifies methods and results; certification body sets minimum evidence thresholds	Require evidence from realistic deployment settings before broad claims are certified
Human oversight and decision logs	User interface records, override logs, advisory explanations, operator training material, escalation procedures, accountability chain	Supports review when AI recommendations influence irrigation, spraying, harvest timing, food-quality decisions, logistics, or compliance	Deploying organisation maintains logs; auditor inspects records; users report failures or contest decisions	Require contestability and escalation pathways for consequential decisions
Monitoring and update records	Drift monitoring, incident reports, model updates, retraining records, cybersecurity events, post-deployment performance reviews	Ensures certification remains valid as crops, climate, pests, markets, sensors, and farm practices change	Deployer monitors; developer reports major changes; certification body triggers re-certification when needed	Make certification conditional on continuing monitoring rather than a one-time approval

Sustainability-claim evidence	Indicator definitions, measurement methods, uncertainty ranges, baselines, life-cycle boundaries, verification records, claim language	Prevents unsupported claims about emissions, water saving, biodiversity, waste reduction, climate adaptation, or input optimisation	Sustainability auditor verifies indicators; certification body checks claim boundaries; firm maintains evidence file	Require sustainability claims to be specific, measurable, bounded, and independently verifiable
Responsibility and accountability records	Role allocation, contractual responsibilities, audit access rights, complaint mechanisms, remediation plans, regulatory reporting duties	Clarifies who is accountable when AI systems fail, mislead users, create unfair impacts, or support unverifiable claims	Certification body checks governance arrangements; deployer maintains responsibility map; regulators oversee enforcement	Refuse or suspend certification when responsibility is unclear or audit access is blocked

Safety and Risk Assurance

Safety and risk assurance are core certification dimensions because agri-food AI can influence physical operations, biological systems, food quality, environmental exposure, and economic decisions. A certified system should not merely demonstrate average performance; it should show how foreseeable hazards were identified, tested, mitigated, and monitored. Safety assurance for autonomous and AI-enabled systems requires explicit evidence about failure modes, operational boundaries, engineering controls, ethical implications, and legal responsibilities [13].

Risk assurance should clarify who is accountable when AI recommendations or automated actions cause harm. In agri-food settings, harms may arise from incorrect irrigation advice, pesticide misapplication, unsafe robotic movement, contaminated quality-control decisions, cybersecurity breaches, or over-reliance on inaccurate predictions. High-stakes AI safety scholarship shows that accountability must be assigned across design, deployment, oversight, and response processes rather than treated as an afterthought once a system is already in use [14].

Agricultural robotics illustrates why safety certification must include embodied and operational risks as well as software performance. Robots, autonomous tractors, drones, and automated handling systems interact with workers, animals, crops, machinery, terrain, weather, and infrastructure. Certification should therefore evaluate human–machine interaction, emergency stop functions, safe operating zones, operator training, maintenance requirements, and policy safeguards for agricultural automation [15].

Safety and risk assurance should operate across the full AI lifecycle. Before deployment, certification should require hazard analysis, stress testing, validation under realistic field conditions, cybersecurity review, and human-oversight design. After deployment, certification should require incident reporting, model-drift monitoring, escalation pathways, corrective-action procedures, and re-certification triggers when the system, data, or operating environment changes [13].

Fairness and Distributional Accountability

Fairness is a certification requirement because agri-food AI can distribute benefits and burdens unevenly across producers, workers, suppliers, consumers, regions, and institutions. A system may improve aggregate efficiency while disadvantaging smaller farms, tenant farmers, digitally excluded users, women farmers, low-resource regions, or

producers whose data are poorly represented. Fairness certification should therefore identify bias sources, evaluate subgroup performance, specify fairness metrics, and explain the limits of technical mitigation [16].

Distributional accountability extends fairness beyond statistical parity. Agri-food AI systems may reshape dependence on platforms, access to credit or insurance, exposure to compliance scrutiny, bargaining power in supply chains, and the authority of agronomic knowledge. A relational ethics approach to algorithmic injustice supports the view that certification must consider power, dependence, vulnerability, and contextual harm rather than only comparing model outputs across predefined groups [17].

Digital agricultural technologies also raise political questions about who controls data, who captures value, and who bears risk. Certification should ask whether AI systems deepen platform dependency, transfer decision authority away from farmers, privilege capital-intensive farms, or make certain actors more visible to surveillance and compliance pressures. The politics of digital agriculture shows that power relations and rural inequality are central to assessing responsible technology deployment [18].

Fairness certification in agri-food systems should therefore combine technical evaluation with institutional and contextual assessment. Technical tests should examine data representation, error rates, access conditions, affordability, interface usability, and performance across production contexts. Governance assessment should examine consultation, contestability, benefit sharing, contractual fairness, and whether affected users can understand, challenge, or opt out of consequential AI-supported decisions [16].

Sustainability Claims and Verification

Sustainability-claim verification is necessary because agri-food AI is frequently promoted as a tool for reducing emissions, saving water, improving biodiversity, reducing waste, optimising inputs, supporting climate adaptation, and strengthening sustainability reporting. These claims may be valuable, but they should not be certified unless the claimed outcome is specific, measurable, bounded, and linked to evidence. Digitalisation and AI can support sustainable food systems, but certification must connect such claims to demonstrable food-system outcomes rather than general innovation narratives [19].

Environmental claims require clear indicators, baselines, measurement boundaries, uncertainty ranges, and evidence of additionality. A precision-agriculture system that claims greenhouse-gas mitigation, input reduction, or productivity improvement should show how those outcomes were measured, compared, and attributed. Evidence on precision-agriculture contributions to emissions mitigation, productivity, and economics supports the need to connect sustainability claims to measurable performance rather than broad technological optimism [20].

Food-system certification must also distinguish verified sustainability claims from proliferating labels and vague product-level assertions. AI dashboards and automated reporting systems may create an appearance of precision while relying on incomplete data, uncertain proxies, or selective boundaries. Research on sustainability claims and labels for newly introduced food products demonstrates why certification should examine claim language, evidence quality, scope, and consumer-facing interpretation [21].

A trustworthy certification framework must therefore guard against greenwashing in AI-enabled sustainability claims. Certification bodies should reject claims that are vague, unverifiable, unsupported by suitable indicators, disconnected from baselines, or presented beyond the certified scope. Concepts and forms of greenwashing show why environmental communication can mislead when claims are selective, symbolic, exaggerated, or insufficiently substantiated [22]. Table 2 compares assurance requirements for safety, fairness, and sustainability claims in agri-food AI certification.

Table 2. Assurance Requirements for Safety, Fairness, and Sustainability Claims in Agri-Food AI Certification: Risk Evidence, Distributional Accountability, Environmental Verification, Failure Modes, and Governance Safeguards

Assurance dimension	Core certification question	Required evidence	Typical failure mode	Governance safeguard
Technical safety	Can the system operate reliably within its declared use conditions?	Validation results, robustness tests, uncertainty estimates, operational limits, cybersecurity review	High average performance hides failures under rare weather, unusual crops, weak connectivity, or sensor degradation	Require stress testing, declared operating boundaries, and incident-response procedures
Operational safety	Can the system be used safely by people and organisations in real deployment settings?	Human-oversight design, training records, user-interface review, escalation pathways, emergency procedures	Users over-rely on recommendations or cannot identify when automation should be overridden	Require human-in-the-loop controls, user training, contestability, and escalation duties
Environmental safety	Could AI recommendations create harmful ecological effects?	Input-use records, water and soil indicators, pesticide-risk assessment, biodiversity-relevant indicators, drift monitoring	Optimisation for yield or cost increases environmental pressure elsewhere	Require environmental risk screening and monitoring of indirect impacts
Food safety and quality	Could AI-supported decisions compromise product safety, quality, or traceability?	Inspection validation, false-negative analysis, traceability records, quality-control logs, recall procedures	Automated inspection misses contamination, defects, or unsafe storage conditions	Require validated inspection thresholds, audit trails, and recall-linked records
Fairness and access	Are benefits and risks distributed equitably across users and affected groups?	Subgroup performance, affordability analysis, access conditions, representation checks, user feedback	System performs best for large, data-rich farms while excluding or misclassifying lower-resource users	Require subgroup testing, access review, and contextual fairness assessment
Distributional accountability	Who gains, who loses, and who can challenge decisions?	Impact assessment, role allocation, complaint mechanisms, benefit-sharing records, stakeholder input	AI shifts risk, cost, surveillance, or decision authority onto weaker actors	Require accountability mapping and affected-user contestability
Sustainability claims	Are environmental or social claims specific, measurable, and verifiable?	Baselines, indicators, measurement methods, uncertainty ranges, life-cycle boundaries, third-party verification	Vague claims about sustainability are used without evidence or beyond the certified scope	Require claim-by-claim certification and prohibit unsupported marketing language
Monitoring and re-certification	Does assurance remain valid after deployment?	Drift logs, incident reports, update records, periodic audits,	Certification becomes stale after model updates, climate shifts, or changed farm practices	Require continuing monitoring, re-certification triggers,

		corrective-action reports	and suspension mechanisms
--	--	---------------------------	---------------------------

Proposed Certification Framework

The proposed trustworthy AI certification framework treats certification as a recurring governance mechanism rather than a one-time approval. Its first dimension, auditability and evidence requirements, establishes whether the system can be examined through documentation, data traceability, model records, decision logs, and verification responsibilities. Ethics-based auditing for trustworthy AI supports this approach because audit processes can help translate trustworthiness into examinable evidence, provided that audit scope and institutional independence are clearly defined [23].

The second dimension, safety and risk assurance, evaluates whether the system has been tested against foreseeable harms and whether post-deployment safeguards exist. The third dimension, fairness and distributional accountability, evaluates whether benefits, risks, access, and decision impacts are distributed justifiably across affected actors. The fourth dimension, sustainability-claim verification, evaluates whether environmental or social claims are measurable, bounded, and independently supportable. Certification should connect these dimensions through conformity assessment before deployment and post-market monitoring after deployment [24].

The framework should operate across multiple governance layers. At the technical layer, certification assesses data, model design, robustness, explainability, monitoring, and cybersecurity. At the organisational layer, it assesses role allocation, documentation routines, incident response, procurement, user training, and accountability. At the institutional layer, it assesses independent audit rules, sector-specific standards, regulatory oversight, claim verification, complaint systems, and enforcement, reflecting the need for layered AI governance [25]. Figure 2 presents the lifecycle assurance pathway through which certification evidence is scoped, independently audited, translated into claim-specific approval, monitored after deployment, and revised when risks, data, models, operating contexts, or sustainability claims change.

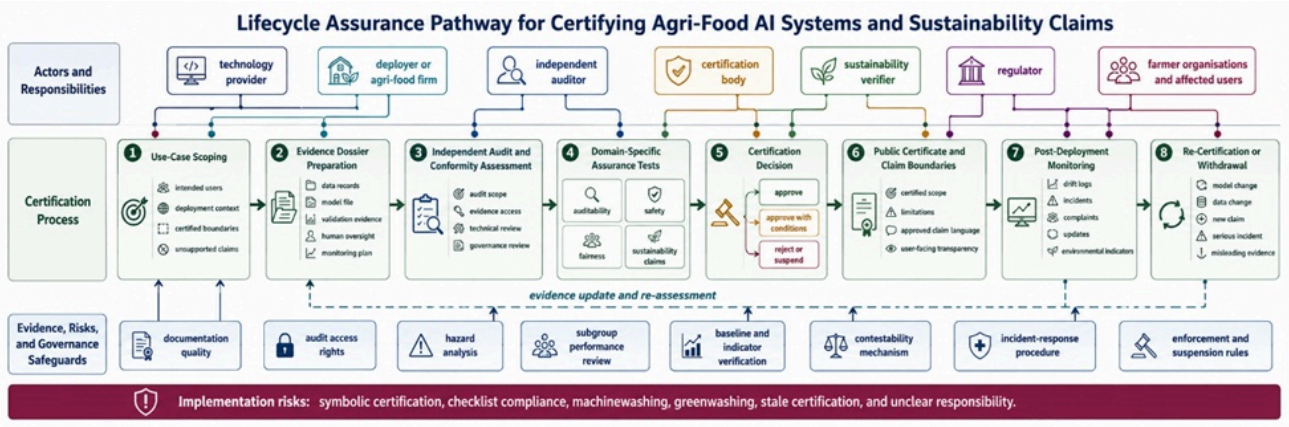


Figure 2. Lifecycle Assurance Pathway for Certifying Agri-Food AI Systems and Sustainability Claims.

Certification should also assign responsibilities across developers, deployers, auditors, users, procurement bodies, regulators, sustainability verifiers, and affected communities. Algorithmic accountability requires that obligations be defined in relation to system design, decision authority, evidence access, monitoring, remediation, and public justification [26]. In agri-food systems, those obligations must also address farmer data rights, data-sharing reluctance, legal uncertainty, and the conditions under which farm data are collected, reused, or commercialised [27]. Table 3 presents the proposed trustworthy AI certification framework for agri-food systems.

Table 3. Proposed Trustworthy AI Certification Framework for Agri-Food Systems: Auditability, Safety, Fairness, Sustainability Claims, Certification Evidence, Oversight Mechanisms, and Policy Implementation Logic

Framework component	Certification purpose	Minimum evidence requirement	Oversight mechanism	Re-certification trigger	Policy implementation logic
Auditability and evidence requirements	Make AI systems inspectable, traceable, and accountable	Dataset records, model documentation, validation evidence, decision logs, monitoring records, responsibility map	Independent technical audit and documentation review	Major data, model, platform, deployment, or claim change	Certification is valid only within documented use boundaries
Safety and risk assurance	Prevent foreseeable harm and support safe operation	Hazard analysis, robustness testing, cybersecurity review, human-oversight design, incident reporting plan	Safety assurance review and operational audit	Serious incident, model drift, new automation function, changed operating environment	Safety evidence must cover real deployment conditions, not only laboratory performance
Fairness and distributional accountability	Identify unequal impacts, access barriers, and unjust risk allocation	Bias assessment, subgroup performance, access analysis, affordability review, stakeholder input, contestability mechanism	Fairness audit and affected-user consultation	Evidence of unequal performance, exclusion, complaints, or changed user population	Certification must consider agricultural context and affected groups
Sustainability-claim verification	Ensure environmental and social claims are credible and bounded	Baselines, indicators, measurement methods, uncertainty ranges, claim scope, third-party verification records	Sustainability audit and claim-specific approval	New claim, changed metric, new reporting boundary, contradictory evidence	Claims are certified individually and cannot exceed verified evidence
Lifecycle monitoring	Keep certification current after deployment	Drift logs, update records, incidents, complaints, environmental indicators, periodic performance reports	Post-deployment surveillance and periodic audit	Scheduled review or significant system change	Certification is a continuing assurance relationship, not a permanent label
Accountability allocation	Clarify who is responsible for evidence, harm,	Role matrix, contractual duties, audit-access rights, escalation	Governance audit and regulator-access provision	Responsibility gaps, blocked audit access, unresolved	Certification should be suspended when

	remediation, and communication	pathways, remediation plan		harm, ownership transfer	responsibilities are unclear
Stakeholder and farmer participation	Ensure certification reflects real use conditions and affected interests	Consultation records, user feedback, farmer-organisation input, complaint records, accessibility review	Participatory review within certification process	New affected group, repeated complaints, deployment in different region	Certification credibility depends on context-sensitive and inclusive governance
Enforcement and public credibility	Prevent symbolic certification, machinewashing, and misleading trust claims	Public certificate scope, limitations, audit summary, enforcement rules, withdrawal procedure	Certification body oversight, regulator coordination, market surveillance	Misleading claim, unverifiable evidence, audit failure, unresolved incident	Certification must include consequences for non-compliance

Implementation Risks and Policy Implications

The main implementation risk is that certification becomes a symbolic trust label rather than a demanding assurance process. This can occur when certification bodies rely on checklist compliance, when auditors lack independence, when firms control evidence access, or when ethical principles are adopted without organisational change. Evaluations of AI ethics guidelines show that ethical language may remain weak when it is not connected to enforcement, incentives, institutional uptake, and practical accountability [28].

A second risk is that certification narrows fairness into a technical metric while ignoring structural inequality and distributional power. In agri-food systems, this would mean certifying systems that satisfy limited bias tests while still reinforcing unequal access, platform dependence, surveillance burdens, or market exclusion. Critiques of algorithmic fairness show that antidiscrimination discourse can fail when it overlooks the broader social conditions under which data and automated decisions operate [29].

A third risk is that certification enables machinewashing, where firms use AI trustworthiness claims to enhance reputation without providing meaningful evidence or accountability. This risk is especially acute when AI systems are marketed as sustainable, responsible, or intelligent while their evidence base remains proprietary, unverifiable, or weakly audited. A credible policy response should require independent auditing, transparent evidence standards, sector-specific assurance criteria, farmer and stakeholder input, post-deployment monitoring, re-certification triggers, sustainability-claim verification, and enforcement mechanisms that can suspend or withdraw certification when claims become misleading [30].

Conclusion

This article has developed a trustworthy AI certification framework for agri-food systems centred on auditability, safety, fairness, and sustainability-claim verification. The framework treats certification as a governance mechanism that depends on evidence quality, audit trails, risk assurance, distributional accountability, environmental verification, monitoring, and institutional credibility. It is designed for AI systems that influence farm decisions, automated operations, food-quality processes, supply-chain coordination, sustainability reporting, and public trust.

The central argument is that certification should not function as a symbolic trust label. A certified agri-food AI system should be able to show what data it used, how it was validated, where it may fail, who is responsible for oversight, how risks are monitored, whether impacts are fair, and whether sustainability claims are verifiable. Without those

requirements, certification can become a weak compliance signal that obscures uncertainty, inequality, environmental ambiguity, or accountability gaps.

Trustworthy AI certification can support agri-food innovation only when it is demanding, transparent, context-sensitive, and enforceable. Future certification systems should therefore combine technical auditability with safety assurance, fairness assessment, sustainability verification, participatory governance, and post-deployment accountability. Used in this way, certification can help build agri-food AI systems that are auditable, safe, fair, environmentally credible, and publicly accountable.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 18 Sep 2025

Revised: 21 Oct 2025

Accepted: 06 Dec 2025

Published online: 15 March 2026

Rights and permissions

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

References

Wolfert J, Ge L, Verdouw C, Bogaardt MJ. Big data in smart farming – a review. *Agric Syst.* 2017;153:69-80.

<https://doi.org/10.1016/j.agry.2017.01.023>.

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 Wageningen J Life Sci.* 2019;90-91:100315.

<https://doi.org/10.1016/j.njas.2019.100315>.

Rose DC, Wheeler R, Winter M, Lobley M, Chivers C-A. Agriculture 4.0: Making it work for people, production, and the planet. *Land Use Policy.* 2021;100:104933.

<https://doi.org/10.1016/j.landusepol.2020.104933>.

Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell*. 2019;1(9):389-99.
<https://doi.org/10.1038/s42256-019-0088-2>.

Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach*. 2018;28(4):689-707.
<https://doi.org/10.1007/s11023-018-9482-5>.

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

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

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>.

Mittelstadt B. Principles alone cannot guarantee ethical AI. *Nat Mach Intell*. 2019;1(11):501-7.
<https://doi.org/10.1038/s42256-019-0114-4>.

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, Morley J, Taddeo M, Floridi L. Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. *Sci Eng Ethics*. 2021;27(4):44.
<https://doi.org/10.1007/s11948-021-00319-4>.

Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumé H 3rd, et al. Datasheets for datasets. *Commun ACM*. 2021;64(12):86-92.
<https://doi.org/10.1145/3458723>.

Burton S, Habli I, Lawton T, McDermid JA, Morgan PDJ, Porter ZLM. Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. *Artif Intell*. 2020;279:103201.
<https://doi.org/10.1016/j.artint.2019.103201>.

Habli I, Lawton T, Porter Z. Artificial intelligence in health care: Accountability and safety. *Bull World Health Organ*. 2020;98(4):251-6.
<https://doi.org/10.2471/BLT.19.237487>.

Sparrow R, Howard M. Robots in agriculture: Prospects, impacts, ethics, and policy. *Precis Agric*. 2021;22(3):818-33.
<https://doi.org/10.1007/s11119-020-09757-9>.

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):1-35.
<https://doi.org/10.1145/3457607>.

Birhane A. Algorithmic injustice: A relational ethics approach. *Patterns (N Y)*. 2021;2(2):100205.
<https://doi.org/10.1016/j.patter.2021.100205>.

Rotz S, Duncan E, Small M, Botschner J, Dara R, Mosby I, et al. The politics of digital agricultural technologies: A preliminary review. *Sociol Ruralis*. 2019;59(2):203-29.
<https://doi.org/10.1111/soru.12233>.

Marvin HJP, Bouzembrak Y, van der Fels-Klerx HJ, Kempenaar C, Veerkamp R, Chauhan A, et al. Digitalisation and artificial intelligence for sustainable food systems. *Trends Food Sci Technol*. 2022;120:344-8.
<https://doi.org/10.1016/j.tifs.2022.01.020>.

Balafoutis A, Beck B, Fountas S, Vangeyte J, van der Wal T, Soto I, et al. Precision agriculture technologies positively contributing to GHG emissions mitigation, farm productivity and economics. *Sustainability (Basel)*. 2017;9(8):1339.
<https://doi.org/10.3390/su9081339>.

Nes K, Antonioli F, Ciaian P. Trends in sustainability claims and labels for newly introduced food products across selected European countries. *Agribusiness*. 2024;40(2):371-90.
<https://doi.org/10.1002/agr.21894>.

de Freitas Netto SV, Sobral MFF, Ribeiro ARB, Soares GRL. Concepts and forms of greenwashing: A systematic review. *Environ Sci Eur*. 2020;32(1):19.
<https://doi.org/10.1186/s12302-020-0300-3>.

Mökander J, Floridi L. Ethics-based auditing to develop trustworthy AI. *Minds Mach*. 2021;31(2):323-7.
<https://doi.org/10.1007/s11023-021-09557-8>.

Mökander J, Axente M, Casolari F, Floridi L. Conformity assessments and post-market monitoring: A guide to the role of auditing in the proposed European AI regulation. *Minds Mach*. 2022;32(2):241-68.
<https://doi.org/10.1007/s11023-021-09577-4>.

Gasser U, Almeida VAF. A layered model for AI governance. *IEEE Internet Comput*. 2017;21(6):58-62.
<https://doi.org/10.1109/MIC.2017.4180835>.

Horneber D, Laumer S. Algorithmic accountability. *Bus Inf Syst Eng*. 2023;65(6):723-30.
<https://doi.org/10.1007/s12599-023-00817-8>.

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 Wageningen J Life Sci*. 2019;90-91:100301.
<https://doi.org/10.1016/j.njas.2019.04.007>.

Hagendorff T. The ethics of AI ethics: An evaluation of guidelines. *Minds Mach*. 2020;30(1):99-120.
<https://doi.org/10.1007/s11023-020-09517-8>.

Hoffmann AL. Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Inf Commun Soc*. 2019;22(7):900-15.
<https://doi.org/10.1080/1369118X.2019.1573912>.

Seele P, Schultz MD. From greenwashing to machinewashing: A model and future directions derived from reasoning by analogy. *J Bus Ethics*. 2022;178(4):1063-89.
<https://doi.org/10.1007/s10551-022-05054-9>.