10. RISK

The Potential Risks and Consequences of Implementing AI Agronomists for Smallholder Agriculture

Principal Author: Elliott Grant[1]
Contributors: Rhishi Pethe[2], Djouher Gaad[3], Tolessa Benti[4]

[1] Cambridge University Institute for Manufacturing
[2] Metal Dog Labs
[3] Biotechnology Research Center
[4] Ministry of Agriculture, Ethiopia

Abstract

AI holds tremendous potential for improving the livelihoods of smallholder producers – whether through AI agronomy advice, improved crop genetics, better remote sensing for state incentive programs, or automation. AI won't solve all the problems for smallholder farmers (e.g. unclear land rights, distorting government incentives, lack of access to fertilizer), but it can be part of the larger solution. However, innovations in complex systems pose risks and can create unintended consequences. Mixing a powerful technology like a non-deterministic, non-explainable AI agronomy tool into a complex system responsible for the health and livelihoods of hundreds of millions of smallholder farmers deserves a concerted effort in anticipating, preventing, and mitigating the worst potential outcomes. This white paper makes a first attempt to catalog the potential negative consequences and harms of an AI agronomist on smallholder agriculture and offers a framework to identify potential mitigating actions. The framework can be applied to any type of AI-driven tool.

Context

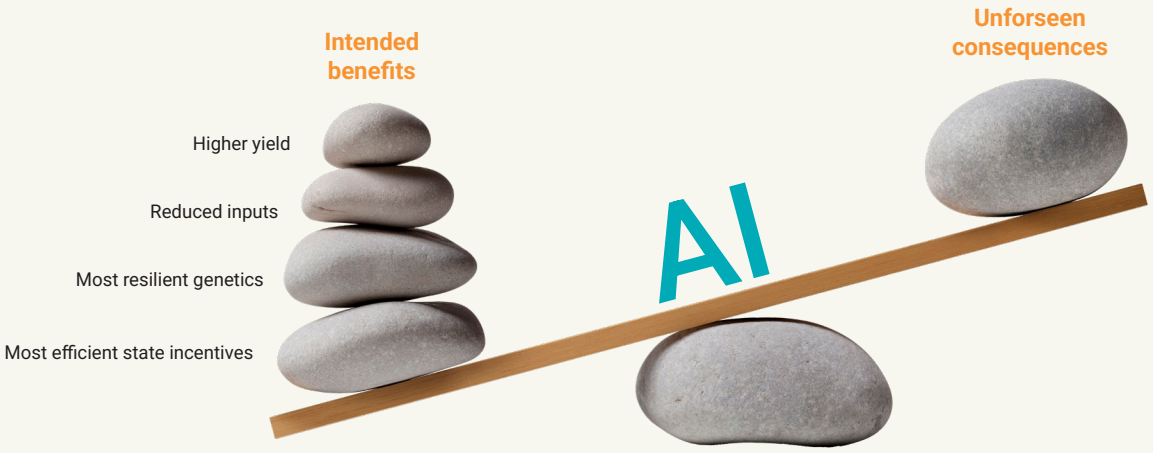
The history of agriculture is replete with stories of innovations driving dramatic improvements: fixed nitrogen, herbicides and pesticides, the Green Revolution, biotechnology and gene editing, food preservation, and global supply chains, to name a few. However, each of these interventions also turned out to have significant unintended consequences. These risks were unintended, but not unforeseeable.

An AI agronomy advisory tool, delivered by smartphone, voice, SMS, WhatsApp, or through an intermediary, holds tremendous promise to improve the productivity and sustainability of hundreds of millions of smallholder producers (Fabregas, R. et al. 2019). Working systems of this kind already exist; for example, Farmer.Chat deploys a retrieval-augmented LLM advisory for smallholders (Singh, N. et al. 2024) and LLMs can be fine-tuned on curated data to improve accuracy (Singh, S. et al. 2026). Agronomy AI won't solve all the problems facing these producers – they often face challenges caused by unclear land ownership, distorting state incentives, price fixing, lack of availability of inputs, consolidated buyers, and so on – but it can be part of the larger solution. In particular, AI

A Chinese farmer accidentally killed 25 acres of his sesame seedlings after an AI Chatbot he was relying on recommended spraying the field with flufenacet

can bridge the information and knowledge gap. Such technologies should be embraced and invested in enthusiastically. But as we prepare to unleash AI into the agriculture system, what lessons can be learned from history? For smallholders operating at subsistence margins, a single bad recommendation (wrong planting date, inappropriate input application) can be catastrophic. In August 2026, the Independent reported how a Chinese farmer accidentally killed 25 acres of his sesame seedlings after an AI Chatbot he was relying on recommended spraying the field with flufenacet (Cuthbertson 2026). Misguided promotion of AI Agronomy tools could lead to crop failure, financial loss, or environmental degradation.

Figure 1. Weighting obvious benefits against the unknown and unforeseen risks



Consequences of AI Agronomy Tools

There is a deep and growing body of research on benchmarking AI agronomists to ensure advice is accurate — see, for example, work on evaluating and fine-tuning conversational AI for agricultural advisory (Singh, S. et al. 2026) and IFPRI's Generative AI for Agriculture (GAIA) initiative. The World Bank's World Development Report 2026 identifies retrieval-augmented grounding (RAG) – anchoring models to trusted, region-specific data – as a central method for keeping automated advice accurate and verifiable (World Bank 2026). Minkoff, et al. identified five key risks facing LLM-powered advisory tools that can be trapped with benchmarks: linguistic and cultural mismatch, contextual inaccuracy/ irrelevant, unverified recommendations, data governance and privacy, and lack of consumer protection and liability mechanisms (Minkoff, et al. 2025). The purpose of this paper is to extend this analysis, and attempt to comprehensively anticipate the second- and third-order negative consequences that can be less obvious. We applied the Failure Mode and Effects Analysis (FMEA) to identify and address potential risks in advance. By articulating them, we can then rank their likelihood and plan actions to avoid or mitigate the worst risks.

CASE STUDY

The World Bank's World Development Report 2026 documents this failure mode. Its agricultural-advisory spotlight gives an example where guidance that is sound in one region can be actively harmful in another: advice suited to Kenya's Great Rift Valley may be wrong for the alluvial plains of Bihar, India. Researchers call this contextual hallucination – a general-purpose model, deployed without grounding in local soil, climate, and cropping conditions, producing fluent, confident advice that does not fit the farmer who receives it. In agriculture the report treats this as a safety problem, not merely a quality one: an incorrect pesticide concentration, an incorrect fertilizer dosage, or a mistimed planting recommendation can cause crop losses, financial harm, chemical contamination, or health risks – and for households that depend on their own production for food, a failed harvest can threaten family nutrition and erode hard-won trust in digital services. The case illustrates this paper's central thesis: an AI system that optimizes for fluent, generic efficiency, without regard for local context, can unintentionally increase the vulnerability of the smallholders it is meant to serve (World Bank 2026; Fanuel et al. 2025).



The analysis generates a long list of potential risks and consequences, then groups the types of risk into eight clusters: model, ethics, linguistic, governance, security, user, economic and environmental.

Table 1. Possible Risks and Consequences

Nature Of Risk	Possible Negative Consequences
Model	• Errors and hallucinations lead to negative outcomes. • Bias in training data — e.g. models trained on Global North agriculture misapplied to smallholder systems, or recommendations that don't account for intercropping, mixed farming systems, or the biological complexity of regenerative practices lead to negative outcomes • Poor data quality, algorithmic bias, or inappropriate model design for local contexts lead to negative outcomes • Unbalanced training data causes the model to overemphasize some practices (e.g. conventional vs. regenerative). • Difficulty maintaining quality of advice as more entrants compete and prices drop. • Loss of diversity of advice: potential to precipitate a local 'flash crash', where synchronized production decisions lead to market flooding, price collapse, or reduced agrobiodiversity. • Misalignment: optimization for yield maximization may drive negative ecological outcomes: soil health, over-extraction, excess chemical inputs/runoff, reduced resilience. • Temporal / distribution drift: a model trained on historical climate or agronomic conditions degrades as those conditions shift, giving advice tuned to conditions that no longer exist. • Feedback-loop contamination ('model autophagy'): future models trained partly on AI-generated advice converge toward self-reinforcing, homogenized practices, narrowing the space of options over time. • Socio-political incoherence: Western-trained models tend to have an underlying socio-political bias that is more secular and liberal than the prevailing worldview in the countries targeted by these AI Advisors. While the impact of a political bias on agronomy advice is unclear – the misalignment with worldview is not (Economist 2026).
Ethical	• Loss of sovereignty: farmer data can be extracted, controlled, or monetized by external actors without consent or benefit-sharing. No compensation for training data provided by indigenous sources. • Exploitation: ultimate business model becomes 'monetize the farmer' (similar to social media platforms). • Exacerbates digital and economic divide: farmers with technology capture value at the expense of the unconnected; women farmers, pastoralists, and those in remote areas with limited connectivity or digital literacy are further excluded. • Concentration of power by app provider, model trainer, and/or data owner. • Benefit asymmetry and surveillance: value and visibility flow to the platform while risk stays with the farmer; continuous data capture becomes a surveillance channel. • World Bank 2026 similarly flags demographic bias in advice quality to women and marginalized farmers, and warns that advisory AI depends on the presence and quality of digital public infrastructure (DPI), such as farmer and land registries, soil databases, and digital identity.
Linguistic	• "Low-resource" languages are poorly represented by LLMs (Claus, H. 2024) and many concepts are hard to translate (e.g. a crop or practice with no direct lexical equivalent), creating confusion and errors. • Cultural bias is embedded within language[1.1] (e.g. from English). • Misinterpretation across dialects, and over time erosion of minority languages if advice standardizes on dominant ones.
Governance	• Moral hazard I: over-focus on advisory draws attention away from more transformational investments such as crop climate resilience. • Moral hazard II (state as funder): governments stop investing in local agricultural research and extension. • Over-regulation (state as regulator): governments mandate explainability in ways that inadvertently prohibit useful non-deterministic models. Non-deterministic, personalized models become hard to certify in an overly-strict regulatory regime. • Agentic systems collude or discriminate in hidden interactions. • Uneven enforcement across jurisdictions creates regulatory arbitrage (certification-shopping) and a race to the bottom.

Continued on next page

Nature Of Risk	Possible Negative Consequences
Security	• Capture of the AI model by input companies or off-takers; data poisoning or manipulation of training data to steer recommendations (Brundage et al. 2018). • Extensive use of AI creates new vulnerabilities to cyberattacks on agriculture (e.g. deliberate bad advice). • Coordinated disinformation campaigns (e.g. fake advisories or scams distributed through the channels farmers come to trust).
User	• Erosion of human role: letting algorithms dictate choices reduces farmers ability to cultivate tacit knowledge, 'wide' knowledge, taste, and agency. (Mollick, 2026) • Farmers treat AI advice as infallible. Confident, fluent output drives decisions beyond what the underlying evidence supports. • Dependency and lock-in to a system farmers don't understand or control, with exposure to pricing changes, service withdrawal, or vendor lock-in. • Farmers use the most convenient general-purpose LLM for agricultural advice and don't realize the advice may be poor or unscientific.
Economic	• Reinforcing power concentration: network effects and two-sided-market dynamics concentrate value in platform owners and data managers; consolidated buyers or input suppliers gain visibility into, or control over, what advice farmers receive. • Labor and livelihood displacement (e.g. for extension agents and agro-dealers) • Infrastructure dependence: reliance on compute, energy, connectivity, and cloud supply chains that smallholders and their states do not control. • Finance; if recommendations lead to crop failure, then collateral such as farm title is at risk, and credit score is compromised for future loans.
Environmental	• Systemic drift toward monoculture or a narrow set of 'optimal' varieties, reducing biodiversity and on-farm resilience. • Soil-health decline, water over-extraction, and nutrient runoff where advice optimizes short-term yield over long-term system health. • Accelerated resistance (pests, weeds, pathogens) if synchronized advice drives uniform input use across a landscape.

Not all risks are equal. When evaluating an AI Agronomy tool or deployment or deciding on mitigation effort we can prioritize the risks using a systematic approach. First, whether the risk impacts an individual or the system. One bad pesticide recommendation can ruin a single farmer's season – whereas a 'flash crash' correlated failure across a village or region will have a much larger impact. Second, is whether harm is detectable and reversible. Hidden collusion within a model may be very difficult to detect and loss of indigenous knowledge over several years may be irreversible. Undetectable and irreversible harms will rank higher for concern. Finally, the risk may vary over time horizon. Some risks are present immediately (e.g. language errors), some risks accumulate with adoption (e.g. market concentration, flash crash, deskilling), and some will reduce over time (e.g. low-resource language errors as models improve).

The relative importance of each risk and harm will also be a function of geography and user-type – making it impractical to offer a definitive ranking of risks here.

Cultural bias is embedded within language

10. RISK The Potential Risks and Consequences of Implementing AI Agronomists for Smallholder Agriculture //

The Counterfactual: Inaction Is Also a Risk

A comprehensive catalogue of risks can create the impression that the safest course is to do nothing or wait. This would be a mistake. The Food and Agriculture Organization (FAO) suggests roughly one extension agent per 500–2,000 rural inhabitants, but the reality is advice is far more scarce: in Nigeria a single public agent may serve 5,000–10,000 farmers, and in India the ratio is about one to 5,000. The advice that does reach farmers is typically generic, one-directional, and tied to fixed seasonal calendars rather than the conditions in a particular field. Hundreds of millions of smallholders farm today have no reliable advisory support at all (World Bank 2026).

The measured upside is large. The World Bank documents early deployments that moved from generic to plot-level advice: CGIAR's NextGen Agroadvisory raised wheat yields by 38% for more than 50,000 Ethiopian smallholders; iSDA's Virtual Agronomist, operating across seven African countries, reported yields 1.4–1.9 times those of neighboring non-users, with Tanzanian sunflower yields rising up to 70%; and India's Ama Krushi (now Krushi Samruddhi) scaled past seven million farmers with benefit–cost ratios of 9:1 to 15:1 at scale, as the cost of delivering advice fell from roughly $30 per farmer toward under $1 (World Bank 2026). These gains come from early, voluntary adopters and should be read as indicative rather than guaranteed, but the direction is unambiguous.

The correct framing is therefore not 'deploy AI' versus 'do nothing', but an argument for intentional, risk-mitigated, well-governed AI against guardrail-free AI. The risks catalogued above are an argument for finding mitigation - designing rules of the road and guardrails

CGIAR's NextGen Agroadvisory raised wheat yields by 38% for more than 50,000 Ethiopian smallholders

In Nigeria a single public agent may serve 5,000 to 10,000 farmers, and in India the ratio is about 1 to 5,000



Mitigations

Despite the seemingly long list of potential harms, the negative consequences of introducing AI agronomy advisory tools are not necessarily inevitable. Actors in the system can take steps in advance of deployment to minimize or detect the risks. These mitigations can be implemented alongside enthusiastic efforts to develop and promote the technology. We are not the only authors arguing for this 'slow down and fix things' approach: the World Bank's recently-published WDR 2026 independently converges on the same safeguards: retrieval-augmented grounding, red teaming (its term for "white-hat attacks"), human-in-the-loop validation, and the need for human intermediaries to ensure farmer trust (World Bank 2026).

Actors and their mitigating actions:



Developer

- Provide proof of balanced datasets for the target audience (e.g. embedding coverage).
- Make claims about who the AI is for (e.g. geography / crop) explicit.
- Communicate the uncertainty of advice in easy-to-understand language, and hand off to a human when confidence is low.
- Mandatory Reinforcement Learning from Human Feedback (RLHF).
- "White-hat attacks" by an authorized agent to find weaknesses before a malicious actor does.
- Expose chain of reasoning on demand.
- Apply established frameworks such as the CARE Principles for Indigenous Data Governance (Carroll et al. 2020)
- Participatory field testing with farmers, extension agents, and local institutions before deployment; transparency about training data.



State

- Require more than one AI advisor to be available, and/or maintain a public-option advisory (i.e. no monopoly).
- Implement a liability policy for developers – creating an accountability mechanism for errors.
- Implement an insurance program for a flash crash.
- Publish performance benchmarks on all AI advisory tools; develop a 'seal of approval' for AI agronomy apps, with annual renewals / a list of approved apps.
- Create user-involved governance (e.g. an advisory group).
- Sustain funding for independent local agricultural research.
- Fund independent oversight bodies and align on cross-border norms to reduce regulatory arbitrage.
- Collective data trusts so farmers negotiate over their data as a group.



End User

- Educate farmers on AI.
- Require user feedback on responses (👍/👎).
- Retention of face-to-face channels to limit the atrophy of local knowledge.
- Community-based validation of recommendations before wide adoption
- Clear opt-out mechanisms for data usage.



Researcher

- Methods for approving non-deterministic, stochastic, personalized models.
- Bias detection.
- Develop guidelines for a 'seal of approval'.



Intermediaries (extension agents, co-ops, dealers, NGOs)

- Act as a human check on AI advice and as a channel for feedback and redress; preserve the face-to-face relationships that have historically buffered bad advice.
- Build and maintain trust with farmers through socially connected intermediaries.
- Localize and validate recommendations against on-the-ground conditions
- Human intermediaries to establish farmer trust and drive uptake; the World Bank cites Rwanda's Digital Ambassadors Program as evidence (World Bank 2026).



Platform & connectivity providers

- As gatekeepers where 'monetize the farmer' can occur, adopt data-use commitments, interoperability, and non-discriminatory access.
- Avoid exclusive lock-in that would make one advisory the only reachable option.



Funders & development-finance organizations

- Condition funding on risk management practices.
- Fund public goods such as open benchmarks, shared datasets, and sustained local research, rather than exclusively private deployments.

Conclusions

Very often, promoters of innovation fail to imagine, ignore, or downplay its potential risks. For a system as critical and precarious as smallholder production, this approach is not acceptable. This paper lays out a process and framework for conceptualizing and prioritizing potential unintended consequences and harms, identifying mitigating actions that stakeholders can take, and spotting unhedged risks.

On balance - and given the demonstrated gains from digital advice already in the literature, the application of AI to the challenges facing smallholder farmers should be net positive. However, this will only happen if the effort of avoiding the unintended consequences is taken proactively and if the set of actors identified here, from intermediaries to platform providers to funders, each take up their part. The question is no longer whether AI can deliver good advice, but whether institutional investment, data infrastructure, and governance arrive quickly enough to serve the hundreds of millions who still farm without reliable support (World Bank 2026).

AI Disclosure Statement

The central concept of this paper, the specific risks it identifies, and the mitigations it proposes were developed by the authors at a workshop held at Mohamed bi Zayed University of Artificial Intelligence, Abu Dhabi, UAE on 10-11 December 2025, and refined through subsequent discussions among the co-authors and a literature review.

The authors used Claude (Anthropic) in three limited capacities during preparation of the manuscript: to check the analysis for completeness, by identifying topics or counter-arguments the authors may have overlooked; to confirm the existence and accuracy of cited references; and to provide editorial suggestions on structure, clarity and style. Any points raised by the tool were assessed by the authors, who decided whether and how to address them. All references were verified against their original sources by the authors.

The authors take full responsibility for the content of this paper.

References

1. Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, et al. 2018. The *Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.* https://doi.org/10.48550/arXiv.1802.07228.
2. Carroll, Stephanie Russo, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, Mark Parsons, et al. 2020. "The CARE Principles for Indigenous Data Governance." *Data Science Journal* 19 (1): 43. https://doi.org/10.5334/dsj-2020-043.
3. Claus, Hannah. 2024. "Now You Are Speaking My Language: Why Minoritised LLMs Matter." *Ada Lovelace Institute.* https://www.adalovelaceinstitute.org/blog/why-minoritised-llms-matter/.
4. Cuthbertson, Anthony. 2026. "Farmer Accidentally Kills 25 Acres of Crops after Trusting AI Advice." *The Independent.* https://www.independent.co.uk/tech/farmer-crops-killed-ai-advice-china-b3030985.html.
5. Fabregas, Raissa, Michael Kremer, and Frank Schilbach. 2019. "Realizing the Potential of Digital Development: The Case of Agricultural Advice." *Science* 366 (6471): eaay3038. https://doi.org/10.1126/science.aay3038.
6. Fanuel, Mesafint, Mahmoud Nabil Mahmoud, Crystal Cook Marshal, Vishal Lakhotia, Biswanath Dari, Kaushik Roy, and Shaohu Zhang. 2025. "AgriRegion: Region-Aware Retrieval for High-Fidelity Agricultural Advice." *arXiv* preprint arXiv:2512.10114. https://doi.org/10.48550/arXiv.2512.10114.
7. Krishnakumari, Parvathy, Sunil Madan, and Parmesh Shah. 2026. "Spotlight 2: The Power of AI to Enhance Agricultural Advisory and Extension Services." In *World Development Report 2026: The Promise of Artificial Intelligence,* 147–153. Washington, DC: World Bank. https://www.worldbank.org/en/publication/wdr2026.
8. Minkoff, Michael, Sivaramakrishnan Balasubramanian, Nupur Mishra, and Kaustav Ghosal. 2025. "A Look at Benchmarking Initiatives for GenAI-Powered Advisory in Agriculture." *AGX AI Discussion Papers.* https://agxai.notion.site/A-Look-at-Benchmarking-Initiatives-for-GenAI-Powered-Advisory-in-Agriculture-234b4002b2a88013ab57db75c05d62f0.
9. Mollick, Ethan. 2026. Co-Existence: The Next Phase of AI. New York: Penguin Random House.
10. Singh, Namita, Jacqueline Wang'ombe, Nereah Okanga, Tetyana Zelenska, Jona Repishti, Jayasankar G K, Sanjeev Mishra, et al. 2024. "Farmer.Chat: Scaling AI-Powered Agricultural Services for Smallholder Farmers." *arXiv* preprint arXiv:2409.08916. https://doi.org/10.48550/arXiv.2409.08916.

11. Singh, Sanyam, Naga Ganesh, Vineet Singh, Lakshmi Pedapudi, Ritesh Kumar, SSP Jyothi, Archana Karanam, et al. 2026. "Fine-Tuning and Evaluating Conversational AI for Agricultural Advisory." *arXiv* preprint arXiv:2603.03294. https://doi.org/10.48550/arXiv.2603.03294.
12. The Economist. 2026. "AI Models' Values Are Very Different from Most People's." *The Economist*. https://www.economist.com/briefing/2026/06/25/ai-models-values-are-very-different-from-most-peoples.
13. World Bank. 2026. "Spotlight 2: The Power of AI to Enhance Agricultural Advisory and Extension Services." In *World Development Report 2026: The Promise of Artificial Intelligence*, 147–53. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-2331-2.



AGX AI is a collaborative initiative led by DevGlobal, Dev-Afrique and Athena Infonomics, to accelerate the responsible design, development, and delivery of generative AI tools and services for small-scale producers (SSPs) in Africa and Asia. Funded by the Gates Foundation, the project builds a vibrant, aligned community to ensure that AI serves the people who grow the world's food.

agx.community/agx-ai