Advancements in Agricultural Development
Volume 7, Issue 2, 2026
agdevresearch.org

Asking the Right Question: Toward a Research Agenda for Responsible GAI in Agricultural Extension

E. Jones-Garcia[1], N. Singaraju[2], K. Davis[3], J. Koo[4]

Abstract

This study explores how generative AI (GAI) tools for agricultural extension can be designed and evaluated more responsibly. While current GAI systems offer scalable, personalized advice, they often ignore the lived realities of smallholder farmers—especially women—by relying on generic datasets and rigid evaluation metrics. We investigate three complementary methods: adversarial testing to expose gendered and contextual blind spots in model outputs; deliberative stakeholder engagement using the C-H-A-T framework, which focused on Collective knowledge, Human insight, Augmentation, and Trust, to surface value tensions and design trade-offs; and field-level insights from extension officers to uncover trust-building, diagnostic reasoning, and social intelligence absent from static GAI interactions. Together, these approaches reveal that responsible GAI requires more than technical accuracy. It demands participatory design processes that foreground user realities, surface stakeholder assumptions, and account for social and institutional context. We recommend developing gender-responsive benchmarks, embedding reflexive, participatory design methods, and modeling advisory reasoning based on real-world extension practice. The findings contribute to a growing agenda for responsible AI development—highlighting the importance of aligning GAI tools not only with technical goals, but with the social, cultural, and political contexts in which they operate.

Article History

Received: June 13, 2025
Accepted: August 5, 2025
Published: January 26, 2026

Keywords

SDG 5: Gender equality; participatory design; evaluation benchmarks; SDG 2: Zero Hunger

1. Eliot Jones-Garcia, Senior Research Analyst, International Food Policy Research Institute (IFPRI), 1201 Eye St., NW, Washington, DC 20005-3915 USA, e.jones-garcia@cgiar.org, https://orcid.org/0000-0003-0483-2368
2. Niyati Singaraju, Postdoctoral Fellow, International Rice Research Institute (IRRI), Pili Drive, Los Baños, Laguna 4031, Philippines, n.singaraju@cgiar.org, https://orcid.org/0009-0001-2521-086X
3. Kristin Davis, Senior Research Fellow, International Food Policy Research Institute (IFPRI), 1201 Eye St., NW, Washington, DC 20005-3915 USA, k.davis@cgiar.org, https://orcid.org/0000-0001-9604-921X
4. Jawoo Koo, Senior Research Fellow, International Food Policy Research Institute (IFPRI), 1201 Eye St., NW, Washington, DC 20005-3915 USA, j.koo@cgiar.org, https://orcid.org/0000-0003-3424-9229

Introduction and Problem Statement

As generative artificial intelligence (GAI) gains traction in agriculture, conversational agents—such as chatbots and voice assistants—are emerging as promising tools for digital extension. These systems allow farmers to interact with AI in natural language, offering timely, personalized, and scalable advice delivered via mobile phones and other accessible platforms (Sai et al., 2025; Steuck et al., 2025). By automating aspects of agricultural advisory services, GAI aims to address long-standing challenges associated with traditional, labor-intensive extension—particularly in reaching remote or underserved communities (Davis et al., 2025; Tzachor et al., 2023).

Yet despite rapid technical advances, current implementations of GAI fall short. Early evaluations show that outputs are often partially accurate but overly generic, and in some cases, outright dangerous—for example, recommending the misapplication of herbicides—as models are ungrounded in local data and lack expert oversight (Davis et al., 2025; Tzachor et al., 2023).

Theoretical and Conceptual Framework

The shortcomings of GAI reflect a disconnect between system design and the lived realities of users; models are often trained on assumed “average” farmers—literate, male farmers engaged in cash-crop production. As a result, the specific needs, knowledge systems, and constraints of women— who make up 71% of the agricultural workforce in Southern Asia and 66% in Sub-Saharan Africa —are frequently overlooked or misrepresented (FAO, 2023). Paradoxically, systems designed to promote equity can leave behind women when they fail to align with their priorities or deny access to tools that might otherwise support their autonomy and prosperity (Foster et al., 2023; Nelson et al., 2024; Singh et al., 2024).

While coordination among donors, scientists, engineers, and extension agents is often cited as key to developing effective GAI systems (Tzachor et al., 2023), this collaborative ambition has not translated into socially responsive design. GAI continues to be framed as a neutral, technical solution—masking the social and institutional dynamics that shape trust, access, and relevance in already unequal digital landscapes (Foster et al., 2023). Much of the existing research remains focused on data-intensive domains like farm optimization, and the prevailing evaluation paradigm privileges model performance and accuracy—rather than attributes that reflect the lived experiences and situated needs of smallholder farmers (Kpodo & Nejadhashemi, 2025). Approaches such as value-sensitive design offer ways to foreground these lived experiences and social dynamics, but remain largely absent from current agricultural design practices (Gil et al., 2025).

Despite decades of field-level advisory experience, we are yet to meaningfully integrate its lessons into the design and evaluation of GAI systems (Adve et al., 2024; Gauba et al., 2025). Extension officers do not simply deliver answers; they diagnose problems, navigate ambiguity, and adapt to unpredictable conditions. Their role is as much about clarification and trust-building as it is about technical advice (Birner et al., 2009; Davis et al., 2025). In contrast, GAI models are often treated as precision tools—designed to produce fixed recommendations from structured inputs—and are typically trained on question–answer datasets that assume clearly defined problems (Sai et al., 2025; Steuck et al., 2025). This limits their capacity for multi-turn, diagnostic reasoning. If GAI is to complement rather than flatten complex extension networks, its development must be guided by a deeper understanding of real-world advisory practice (Kpodo & Nejadhashemi, 2025)

Recent proposals have called for more responsible development pathways, including human-in-the-loop design approaches and digital sandboxes—controlled environments where systems can be tested with stakeholders before deployment (Singh et al., 2024; Tzachor et al., 2023). Others advocate for co-design processes that embed extension experts throughout the AI lifecycle—from data labeling to output interpretation—to improve transparency, trust, and relevance (Kpodo & Nejadhashemi, 2025). These approaches highlight a shift away from purely technical fixes toward participatory, iterative, and socially responsible AI development (Steuck et al., 2025).

Purpose

This article contributes to that shift by exploring three novel methods through which food systems researchers can inform the design of GAI tools. Together, these approaches aim to build a research agenda for responsible AI in agricultural extension. In light of the critiques outlined above, we ask:

1. How might GAI reinforce digital divides or inadvertently exclude certain users? We explore this through adversarial testing, identifying ways GAI can reflect gendered blind spots even when outputs appear neutral.
2. What are the social, organizational, and policy trade-offs to GAI in agricultural extension? We probe these through deliberative stakeholder engagement, bringing together funders, developers, and researchers to surface hidden value trade-offs.
3. How do real-world advisors navigate ambiguity and surface actionable insights with minimal context? We investigate this through workshops with extension officers, examining their strategies for interpreting farmer needs and building trust.

Together, these perspectives offer a multi-layered view of what responsible AI in extension might look like—one that centers not only accuracy, but on inclusivity, equity, and contextual responsiveness.

Methods

Study 1: Adversarial Evaluation for Risk Discovery

We conducted structured adversarial testing of five leading GAI models—ChatGPT 4o, Claude 3.5 Sonnet, Llama 3.3 70B Instruct, Jamba 1.5 Large, and Nova Pro 1.0—to examine how models respond to ambiguity-laden prompts about gender in agriculture, with a specific focus on women farmers in India. Each model was tested under identical conditions (temperature 0.1, Top P 0.1) using 12 prompts. These settings were chosen to ensure that the responses from all models were more focused and consistent, rather than unpredictable, making it easier to compare how each model responded to the same question. All questions were posed from the perspective of a woman farmer in India and grouped under three categories: gender equality, gender responsiveness, and informal norms (see Appendix document).

While red teaming has been used to probe gender bias in general-purpose AI systems—particularly in consumer-facing domains like online safety, hiring, and customer service (McDonnell & Baxter, 2019; Su et al., 2023)—its application in agricultural advisory remains underexplored. Between general-purpose GAI and domain-specific retrieval systems lies a range of refinement techniques that, while often overlooked by major AI developers, remain accessible and valuable for nonprofits, governments, and agri-businesses seeking to strengthen agricultural AI. One example is Singh et al. (2024), who evaluated Farmer.Chat's gender responsiveness within a customized GAI trained on curated agricultural content.

In contrast, our study tested general-purpose models without any domain-specific tuning. Rather than focusing on inference or factual accuracy alone, we examined how well these models could interpret structural constraints and deliver advice that is both contextually relevant and gender-responsive—specifically, whether they could recognize the real challenges faced by Indian women farmers, including unequal access to labor, limited institutional support, and restrictive cultural norms. While some models demonstrated awareness of systemic barriers, others defaulted to vague or generic encouragement. This variation underscores persistent interpretive and contextual blind spots that conventional evaluation metrics often fail to capture.

Study 2: Deliberative Stakeholder Engagement

We facilitated a two-hour in-person workshop with 20 participants—including funders, developers, extension experts, and researchers—who joined voluntarily from stakeholder organizations in the project consortium, reflecting diverse ages, backgrounds, and perspectives. Drawing on principles of value-sensitive design (Friedman et al., 2013)—an approach still nascent in agricultural research (Gil et al., 2025)—the session aimed to surface the implicit values and trade-offs embedded in the design of GAI systems. In contrast to technical or performance-based evaluations, the intervention focused on how GAI tools might reinforce, reconfigure, or disrupt existing advisory dynamics.

To structure the conversation, the authors introduced the C-H-A-T framework, which invites reflection across four dimensions of responsible advisory design: **Collective knowledge**, **Human insight**, **Augmentation**, and **Trust** (see Appendix document). Due to time constraints, the session did not attempt to elicit stakeholder values from scratch. Instead, each theme was pre-framed with a normative dilemma developed by the authors based on ongoing research and literature on digital extension. Participants engaged in these dilemmas through guided prompts and structured group discussion. Reflections were documented in real time using Miro, a collaborative whiteboard platform that supported clustering and synthesis of emerging insights.

Study 3: Extension officer discussions

We conducted discussions with field extension officers attending the 7th Africa-wide Extension Week conference in Lilongwe, Malawi. The aim was to document how human advisors diagnose farmer needs in low-context, trust-sensitive environments, to inform the training of GAI. Specifically, we sought to understand the conversational strategies, reasoning processes, and adaptive questioning techniques that underpin effective advisory interactions.

Although lay user involvement is widely cited as a principle of responsible AI design, it remains largely absent in practice—particularly in agricultural contexts (Vincenzi et al., 2024). The workshop was designed to address this gap. Questions were structured to enable extension officers to actively contribute to the design of GAI systems. The exercise positioned their expertise not as background data, but as critical to defining the capabilities and limits of what a conversational advisory agent should do (Katell et al., 2020).

Participants were a self-selecting convenience sample of 30 extension experts—men and women of varying ages from West, East, and Southern Africa. The session began with an introduction by the lead researcher, after which participants were guided through a set of discussion questions based on findings from the two preceding studies (see Appendix document). They then broke into six self-facilitated groups of 3–6 people, with two facilitators circulating to observe and assist as needed. Conversations were recorded, transcribed verbatim, cleaned, and analyzed using NVivo (version 15). Due to ambient noise from the conference venue, some audio segments were incomplete.

Findings

Adversarial Evaluation

Gender Equality: Affirmation Without Depth

All GAIs affirmed gender equality, stating clearly that gender should not restrict farming. However, their responses lacked depth. Jamba's reply— "Yes, women can do anything men can do. Be confident!"—was motivational but ignored systemic barriers such as unequal access to land, inputs, and extension services. Nova claimed that men are "better at physically demanding work" while women are suited to "tasks involving care and nurturing"—a framing that reinforces, rather than challenges, gender stereotypes.

When asked to cite sources, several models referenced literature from the 1990s, reflecting a broader issue: GAIs are not trained on current, region-specific data and default to outdated narratives. In contrast, field-based studies have documented the evolving role of women in Indian agriculture, especially in regions with high male outmigration, where women are increasingly managing farms, adopting technology, and leading collectives (Leder, 2022; Niyati, 2020; Raj et al., 2025).

Gender Responsiveness: Limited Specificity and Outdated Content
Model responses showed weak contextual alignment with real-world constraints. While ChatGPT referred to Mahila Kisan Sashaktikaran Pariyojana (MKSP) and Self-Help Groups (SHGs), others offered vague or irrelevant suggestions. Nova advised “try contacting local banks,” and Llama suggested “investing in beauty salons”—misaligned with agricultural needs.

When asked about labor-saving technologies, most models gave generic suggestions such as “use threshers,” without mentioning low-cost tools, leasing mechanisms, or collective access models used by women to overcome capital constraints (Agarwal, 2018). On low-labor crops, answers like “try pulses or millets” failed to address deeper constraints, such as limited access to improved seed varieties and exclusion from extension networks—issues well-documented in fieldwork (Puskur et al., 2021; Tenneti et al., 2024).

Some responses cited outdated programs or general global content, reflecting a reliance on static, non-local data. As Pava et al., (2025) note, GAIs trained on broad internet resources often default to globally averaged outputs, which are of little use in localized advisory settings.

Informal Norms: Cultural Contexts Ignored or Misread
This category revealed major shortcomings. Claude stated that “informal laws are not real,” dismissing cultural norms altogether. ChatGPT identified stigma, family pressure, and community expectations, offering a relatively more grounded response.

However, none of the models recognized caste dynamics, land inheritance practices, or region-specific restrictions on women’s autonomy. As Sloane and Moss (2019) suggest, this blind spot reflects a lack of training on qualitative, ethnographic, and field-level data. Without sociological studies, regional language inputs, or field notes, GAIs misinterpret—or completely miss—the informal systems shaping agricultural access.

Deliberative Stakeholder Engagement

Collective Knowledge
Participants emphasized that farmers often validate advice socially—through peer exchange, observation, and communal discussion. One-to-one chatbot interactions, they noted, may isolate users and undermine these established forms of learning. At the same time, some acknowledged the value in reaching farmers who are excluded from group-based support. To navigate this tension, participants proposed design alternatives such as peer-to-peer Q&A features, tools that treat learning as iterative and communal, and systems that accommodate shared phone use—common in contexts where individuals, especially women or youth, lack personal devices. WhatsApp groups were cited as successful models for collaborative problem-solving, prompting suggestions that GAI could be trained on similar dialogue patterns. Others envisioned leaderboard systems, community-contributed content, or mechanisms that allow farmers to take ownership of their learning pathways. Participants also emphasized the need for further research to understand what forms of knowledge are personal versus collective, and how GAI systems might support both without displacing existing networks.

Human Insight

Participants contrasted the strengths of human advisors with the limitations of GAI in handling ambiguity and interpreting context. They described how farmers often express needs through shorthand or embodied knowledge—for instance, saying “tomato” with the expectation that an advisor will infer a disease or pest issue, or describing symptoms without knowing their cause. While human advisors typically respond with follow-up questions and iterative probing, GAI tools—especially those based on static Q&A frameworks—tend to treat such queries as complete, limiting their diagnostic capacity. Some participants acknowledged the appeal of GAI’s consistency and availability, particularly in under-resourced settings. However, they also highlighted the risk of misalignment when systems fail to adapt to fluid, culturally embedded language or respond appropriately to contextual cues. These reflections underscored a broader trade-off: between the efficiency of automated advice and the situated, relational knowledge that human advisors bring to complex problem-solving.

Augmentation

While many acknowledged that GAI could help ease pressure on overstretched extension systems, they emphasized its limitations—particularly in high-stakes or context-specific scenarios such as pest outbreaks, loan applications, or disputes requiring discretion. In such cases, human advisors were seen as essential. Participants warned that without clear handoff mechanisms, GAI systems risk becoming bottlenecks rather than bridges, failing to escalate cases when automated advice reaches its limits.

At the same time, participants recognized the practical constraints in many low-resource settings, where substitution may be the only viable option. In these contexts, they argued, GAI should not be measured against the gold standard of in-person extension, but against the real alternatives farmers face: radio, SMS, or no access at all.

The conversation also surfaced broader questions about the evolving role of extension. As digital tools increasingly bundle advisory functions with marketplaces and financial services, participants highlighted trade-offs in how responsibility and trust are distributed. Farmers with prior exposure to in-person extension may view chatbots through a comparative lens, while for others, GAI may be their first and only advisory experience—shaping how advice is received, acted upon, and trusted.

To manage these transitions responsibly, participants stressed the importance of planning not only for scale, but for exit—including clear handoff protocols and pathways for transitioning control to community-based institutions.

Trust

Trust emerged as both relational and fragile. Some farmers associate digital tools with scams—especially when tone, language, or delays feel misaligned with expectations. Participants called for clearer onboarding, personalization, **and** explanation prompts to clarify why recommendations are made and how they can be adapted. Trust can also be undermined when advice clashes with expectations, such as omitting fertilizer recommendations in high-stakes situations. A lack of accountability—no clear recourse to challenge or clarify advice—weakens user confidence. While bundling services (e.g., input, weather, credit) may offer convenience, it risks overwhelming users and blurring responsibility. Participants additionally noted that identity-based benefits may be strategically misused, allowing dominant actors to game the system and obscure marginalization. These dynamics challenge the assumption that participation is inherently equitable and underscore the need for intersectional design, inclusive representation, and thoughtful governance.

Extension Officer Workshops

We identified strategies for building trust and rapport, interacting with different types of clients, and for structuring interactions with farmers.

Trust

Most groups talked about the importance of trust and building rapport. One respondent said "... building rapport. That's what's most important in extension...You need to build that connection to build trust." "...if you're introducing a new technology, you must first build the farmer's trust. Because if the farmer doesn't believe in you, it's hard to influence him [or her]."

How do you build rapport and increase trust? Through social interaction, respect for local customs and norms, and being part of the community.

Social Interaction

Human extension interaction typically includes "small talk" involving greetings and discussion of social issues, rather than diving straight into a technical problem. Group interactions may even include dancing and other social activities. Extension officers may be offered a cup of tea by the farmer. One respondent summarized their group discussion: "So we're going first greet the farmers and then we see whether there are like social issues or family dynamic problems if there are, like, family dynamic problems or social issues, we don't continue."

"Culture" in Agriculture

Local customs and norms should drive how extension officers interact with farmers. Conversational content and formality are adapted according to whom the agent is interacting with. Interaction with a leader may involve a more formal approach than with a young person. Dress is also adapted. One respondent said that in Malawi, "if you are a woman, you need to put on your wrapper... and if you are to meet young people sometimes, you might wear your hat, but if you are meeting the elders, you need to take off your hat when you are talking to them because that is a sign of respect."

Community Integration

Respondents noted that extension workers must be part of the local community. "They should live in the area where they work, so they can feel both the people and the place." This is linked to trust: "for that rapport to happen, the extension worker must not come from outside, the key is [s]he has to be part of the society."

Interactions and Structuring
The discussions helped identify tacit practices of questioning, listening, and adaptive reasoning, and to surface interactional practices, guiding conversational design and model prompting.

As opposed to farmer-AI interaction, extension officer to farmer interaction is often group-based. Thus there is often a mobilization process, getting the community members ready for an interaction, followed by the interaction process. As explained above, this starts out with greetings and social interactions, followed by an introduction to the interaction and setting expectations.

One respondent said "You are trying to feel the situation the farmer is in. You immerse yourself in it. And from there, you begin to analyze how the farmer feels. That's when you find a solution."

In terms of structure of an individual engagement, extension officers normally build from the known to the unknown and move from simple to complex. The flow of questioning or advice is adjusted based on how complex the problem is. There is a process of continual probing to get to practical answers that are co-created with farmers.

A final point is that "...as an extension agent, you go with the mindset that the farmer is the best expert on his [or her] farm. The best knowledge on his [or her] farm. So it's important for you to listen so you can have detailed information to feed back as research."

Conclusions, Discussion, and Recommendations

This study demonstrates the value of rethinking how GAI tools are evaluated, designed, and embedded within agricultural extension. By combining adversarial evaluation, deliberative engagement, and field-level insights, we highlight not only technical shortcomings, but also the social tensions and design trade-offs that GAI systems must navigate.

Adversarial testing offered a diagnostic lens to probe whether GAI reflects the social and institutional realities of its users (Singh et al., 2024). Rather than assessing correctness alone, gender-responsive prompts—especially those reflecting the lived constraints of Indian women farmers—triggered vague or misaligned responses, exposing the absence of regional language data, qualitative insight, and interdisciplinary grounding in current training pipelines.

The C-H-A-T framework extended this lens by prompting stakeholders to reflect not only on what GAI can do, but on what it should do—and for whom. Research on agricultural digitalization has shown that new technologies often have uneven impacts, reinforcing existing power asymmetries by prioritizing efficiency and yield maximization over broader concerns such as equity, rural livelihoods, and food justice (Bronson, 2018). These dynamics have tended to benefit large agribusinesses while marginalizing small and medium-sized farmers. Without deliberate efforts to surface and negotiate values, similar patterns are likely to emerge in the design and deployment of GAI. In the workshop, participants reconsidered default assumptions such as accuracy, efficiency, or personalization, and surfaced value tensions between personalization and social learning, scale and nuance, automation and care. This process encouraged design actors to anticipate risk, contextualize goals, and imagine more inclusive and grounded development pathways (Vincenzi et al., 2024).

Field-level insights from the extension officer discussions further emphasized the limitations of viewing advice as simple information delivery. Effective advisors diagnose through dialogue—probing, listening, and adapting in real time (Birner et al., 2009). Their work builds relational trust, draws on accumulated field knowledge, and reflects adaptive reasoning—qualities rarely captured by static Q&A-driven GAI systems. These findings underscore the need for GAI to emulate not just expert content, but the interactional dynamics of meaningful human advisory practice (Steuck et al., 2025).

Each method offers distinct strengths. Adversarial testing is replicable, low-cost, and effective at stress-testing models under real-world conditions, making hidden exclusions—particularly around gender and context—more visible. However, it captures only surface-level interactions and cannot assess how models evolve over multi-turn dialogue (Su et al., 2023). The deliberative engagement created space for value-based reflection, enabling stakeholders to challenge assumptions and explore tensions that often remain implicit in design. Its main limitation is its reliance on facilitation and subjective interpretation, which may limit comparability across contexts (Sadek & Mougenot, 2024). The extension officer discussions brought uniquely grounded insights into real advisory practice—offering a relational and iterative model of interaction rarely represented in current GAI systems (Katell et al., 2020). Yet its depth comes at the cost of scale and standardization, making it harder to generalize (Adve et al., 2024; Gauba et al., 2025). In short, while each method is partial and imperfect, their combined use balances diagnostic scrutiny, normative reflection, and experiential learning to support socially responsive GAI development.

Our study reinforces growing calls to move beyond conventional benchmarks toward value-sensitive and justice-oriented approaches to AI (Kpodo & Nejadhashemi, 2025). As Leeuw (2025) argues, responsible design must account for informal norms, power asymmetries, and institutional dynamics. Without this shift, GAI risks reinforcing exclusion, particularly for women and marginalized farmers whose needs are not captured by default user assumptions or data pipelines (Shrestha & Das, 2022).

Insights from extension officers offer a compelling alternative: a conversational, iterative model of advisory work grounded in trust, clarification, and local intelligence. For GAI to meaningfully support rather than displace this role, designers must embed multi-turn reasoning, user history, and context-aware prompts into system architecture (Adve et al., 2024; Gauba et al., 2025).

Ultimately, this demands a reorientation in both research and practice—from prioritizing reach and scale to enabling relevance and responsiveness. Institutions must invest in participatory methods, regionally grounded data, and evaluation criteria that reflect not only what GAI can do in theory, but how it performs in practice—across diverse users, settings, and forms of knowledge.

Moving forward, we identify three key areas for future research and development:

1. Develop Benchmarks for Gender Responsiveness
Building on our adversarial testing, we are mapping existing benchmarks and developing new ones that reflect real-world relevance. This includes using regional and qualitative data to test whether GAI tools reinforce or challenge exclusionary patterns. Planned next steps include participatory testing with women farmers, co-designing advisory prompts, and establishing shared indicators for gender responsiveness. These benchmarks aim to ensure GAI systems are not only equitable in theory but inclusive in practice.

2. Expand Participatory Design Across Users and Contexts
To build more inclusive systems, we plan to broaden our participatory design efforts beyond extension officers. Future workshops will directly engage farmers to better understand how different groups conceptualize advice, evaluate trust, and navigate trade-offs. We also aim to make cross-country comparisons to examine how extension logic varies across cultural and institutional settings. This will deepen understanding of user diversity and help tailor GAI systems accordingly.

3. Advance Diagnostic Reasoning
Findings from the extension officer discussions highlight the need for GAI to emulate human diagnostic reasoning. we are developing a prototype dialogue model based on the conversational structure outlined in the Appendix Document. The goal is to move beyond static question–answer exchanges and enable systems to:

- Engage in multi-turn dialogue, asking clarifying questions when farmer inputs are vague or ambiguous.
- Identify and respond to hidden constraints—such as time, labor, or gender-specific challenges—through sensitive, contextual probing.
- Sequence interactions in ways that reflect how skilled advisors move from symptoms to causes, rather than jumping to fixed recommendations.

Together, these efforts aim not just to improve GAI functionality, but to reimagine what responsiveness, inclusion, and reasoning should look like in AI-mediated agricultural advisory.

Acknowledgments

Funding Information: This research was supported by the Gates Foundation (Investment ID INV-069415).

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Previous Dissemination: No portion of this article has been previously published, included in a thesis or dissertation, or presented as a conference paper or abstract.

Artificial Intelligence: ChatGPT was used iteratively during the planning, drafting, and revision of this article. The author provided the content and the text was carefully reviewed and edited for publication.

Author Contribution Statement: **E. Jones-Garcia** – Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft, Writing – Review & Editing; **N. Singaraju** – Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft, Writing – Review & Editing; **K. Davis** – Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft, Writing – Review & Editing; **J. Koo** – Conceptualization, Methodology, Supervision, Project Administration, Funding acquisition.

References

Adve, V. S., Wedow, J. M., Ainsworth, E. A., Chowdhary, G., Green-Miller, A., & Tucker, C. (2024). AIFARMS: Artificial intelligence for future agricultural resilience, management, and sustainability. *AI Magazine*, *45*(1), 83–88. https://doi.org/10.1002/aaai.12152

Agarwal, B. (2018). Gender equality, food security and the sustainable development goals. *Current Opinion in Environmental Sustainability*, *34*, 26–32. https://doi.org/10.1016/J.COSUST.2018.07.002

Birner, R., Davis, K., Pender, J., Nkonya, E., Anandajayasekeram, P., Ekboir, J., Mbabu, A., Spielman, D. J., Horna, D., Benin, S., & Cohen, M. (2009). From best practice to best fit: A framework for designing and analyzing pluralistic agricultural advisory services worldwide. *The Journal of Agricultural Education and Extension, 15*(4), 341–355. *https://doi.org/10.1080/13892240903309595*

Bronson, K. (2018). Smart farming: Including rights holders for responsible agricultural innovation. *Technology Innovation Management Review, 8*(2), 7–14. *http://doi.org/10.22215/timreview/1135*

Davis, K., Gandhi, R., Koo, J., Kramer, B., Miller, A., Repishti, J., Spielman, D. J., & Sulaiman, V., R. (2025). From transfer of technology to facilitation for innovation. *Global Food Policy Report (IFPRI)*, 195–220. https://cgspace.cgiar.org/items/bdf4515e-8f9f-479d-862e-19b05b4a513c

FAO. (2023). *The status of women in agrifood systems.* FAO. https://doi.org/10.4060/CC5343EN

Foster, L., Szilagyi, K., Wairegi, A., Oguamanam, C., & de Beer, J. (2023). Smart farming and artificial intelligence in East Africa: Addressing indigeneity, plants, and gender. *Smart Agricultural Technology*, *3*. https://doi.org/10.1016/j.atech.2022.100132

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value sensitive design and information systems. In *Philosophy of engineering and technology* (Vol. 16, pp. 55–95). Springer Nature. https://doi.org/10.1007/978-94-007-7844-3_4

Gauba, A., Pi, I., Man, Y., Pang, Z., Adve, V. S., & Wang, Y.-X. (2025). *AgMMU: A comprehensive agricultural multimodal understanding and reasoning benchmark*. arXiv preprint https://arxiv.org/pdf/2504.10568

Gil, R. M., Ryan, M., & García, R. (2025). Sovereignty by design and human values in agriculture data spaces. *Agriculture and Human Values, 42*, 1413–1438. https://doi.org/10.1007/s10460-024-10674-0

Katell, M., Young, M., Dailey, D., Herman, B., Guetler, V., Tam, A., Binz, C., Raz, D., & Krafft, P. M. (2020). Toward situated interventions for algorithmic equity: Lessons from the field. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 45–55. https://doi.org/10.1145/3351095.3372874

Kpodo, J., & Nejadhashemi, A. P. (2025). Navigating challenges/opportunities in developing smart agricultural extension platforms: Multi-media data mining techniques. *Artificial Intelligence in Agriculture*, *15*(3), 426–448. https://doi.org/10.1016/J.AIIA.2025.04.001

Leder, S. (2022). Beyond the 'feminization of agriculture': Rural out-migration, shifting gender relations and emerging spaces in natural resource management. *Journal of Rural Studies*, *91*, 157–169. https://doi.org/10.1016/J.JRURSTUD.2022.02.009

Leeuw, F. (2025). Evaluating the effects of using AI and big data in interventions: Examples and the need for a realist agenda in: Artificial Intelligence and Big Data. In F. Leeuw & M. Bamberger (Eds.), *Artificial Intelligence and Big Data* (pp. 126–145). Edward Eglar Publishing. https://doi.org/10.4337/9781803925677.00017

McDonnell, M., & Baxter, D. (2019). Chatbots and gender stereotyping. *Interacting with Computers*, *31*(2), 116–121. https://doi.org/10.1093/IWC/IWZ007

Nelson, I. L., Faxon, H. O., & Ehlers, M. H. (2024). Feminist political ecologies of agrarian technologies: Knowing the digital differently. *Journal of Peasant Studies*, *51*(6), 1303–1330. https://doi.org/10.1080/03066150.2024.2308637

Niyati, S. (2020). Women in the rice economy of India: Evidence from village studies. In M. Swaminathan, S. Nagbhushan, & V. K. Ramachandran (Eds.), *Women and in rural India* (pp. 109–133). Tulika Books. https://cup.columbia.edu/book/women-and-work-in-rural-india/9788193926963/

Pava, J., Meinhardt, C., Badi Uz Zaman, H., Friedman, T., Truong, S., Zhang, D., Cryst, E., Marivate, V., & Koyejo, S. (2025). *Mind the (language) gap: Mapping the challenges of LLM development in low-resource language contexts*. Stanford Institute for Human-Centered Artificial Intelligence (HAI). https://hai.stanford.edu/policy/mind-the-language-gap-mapping-the-challenges-of-llm-development-in-low-resource-language-contexts

Puskur, R., Mudege, N. N., Njuguna-Mungai, E., Nchanji, E., Vernooy, R., Galiè, A., & Najjar, D. (2021). *Moving beyond reaching women in seed systems development*. International Food Policy Research Institute (IFPRI). https://doi.org/10.2499/9780896293915_03

Raj, R., Ravula, P., C. M, P., Bhanjdeo, A., Sogani, R., & Rao, N. (2025). Male migration and the transformation of gendered agriculture work: A comparative exploration of heterogeneity across selected Indian states. *Gender, Place and Culture*, 1–27. https://doi.org/10.1080/0966369X.2025.2468178

Sadek, M., & Mougenot, C. (2024). Challenges in value-sensitive AI design: Insights from AI practitioner interviews. *International Journal of Human-Computer Interaction*, 1–18. https://doi.org/10.1080/10447318.2024.2439021

Sai, S., Kumar, S., Gaur, A., Goyal, S., Chamola, V., & Hussain, A. (2025). Unleashing the power of generative AI in agriculture 4.0 for smart and sustainable farming. *Cognitive Computation*, *17*(1). https://doi.org/10.1007/s12559-025-10420-6

Shrestha, S., & Das, S. (2022). Exploring gender biases in ML and AI academic research through systematic literature review. *Frontiers in Artificial Intelligence*, *5*. https://doi.org/10.3389/frai.2022.976838

Singh, N., Wang'ombe, J., Okanga, N., Zelenska, T., Repishti, J., K, J. G., Mishra, S., Manokaran, R., Singh, V., Rafiq, M. I., Gandhi, R., & Nambi, A. (2024). *Farmer.Chat: Scaling AI-powered agricultural services for smallholder farmers*. arXiv preprint. http://arxiv.org/abs/2409.08916

Sloane, M., & Moss, E. (2019). AI's social sciences deficit. *Nature Machine Intelligence, 1*, 330–331. https://doi.org/10.1038/s42256-019-0084-6

Steuck, P.-F., Di Maria, M., Bierschwale, D., Gottschewski-Meyer, P. O., & Knackstedt, R. (2025). Designing anthropomorphic conversational agents to enhance laypeople's acceptance of generative AI. In H. Degen & S. Ntioa (Eds.), Artificial Intelligence in *HCII 2025* (Vol. 15820, pp. 282–300). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-93415-5_17

Su, H., Cheng, C-C., Farn, H., Kumar, S. H., Sahay, S., Chen, S.-T., & Lee, H.-Y. (2023). *Learning from red teaming: Gender bias provocation and mitigation in large language models*. arXiv preprint. https://arxiv.org/pdf/2310.11079

Tenneti, S., Kori, P., & Puskur, R. (2024). *Understanding the information gender gap when talking to women farmers about seeds.* Rice Today. https://ricetoday.irri.org/understanding-the-information-gender-gap-when-talking-to-women-farmers-about-seeds/

Tzachor, A., Devare, M., Richards, C., Pypers, P., Ghosh, A., Koo, J., Johal, S., & King, B. (2023). Large language models and agricultural extension services. *Nature Food*, *4*(11), 941–948. https://doi.org/10.1038/s43016-023-00867-x

Vincenzi, B., Taylor, A. S., & Stumpf, S. (2024). Lay user involvement in developing human-centric responsible AI systems: When and how? *ACM Journal on Responsible Computing*, *1*(2), Article 14. https://doi.org/10.1145/3652592