

Application of Reinforcement Learning from Human Feedback for Localizing Quality Agricultural Advice using Generative AI

V. Singh¹, S. Singh², L. Pedapudi³, W. Pasha⁴, E. Kumari⁵, R. Ghandi⁶,
A. Miller⁷, J. Repishti⁸, A. Karanam⁹, E. Firnhaber¹⁰

Abstract

Recent generative AI offers personalized, high-quality advice to smallholder farmers in resource-limited settings. Yet, most large language models (LLMs) lack training data for diverse agroecologies, often yielding generic, inaccurate, or locally misaligned advice. Digital Green adapted Reinforcement Learning from Human Feedback (RLHF) to agricultural advisory to deliver highly localized, relevant, information. This refined tool, called Farmer.Chat, is an AI assistant supporting over 670,000 farmers in India, Kenya, Ethiopia, and Nigeria with text, image, and voice-based content. This paper details Digital Green's RLHF approach: a web-based annotation tool, multi-phase implementation, and quality assurance. Over 25,000 expert-reviewed Q&A pairs yielded significant improvements in response quality, tone, context, and cultural fit, especially for region-specific agricultural queries. The work outlines key lessons, cost/equity, and replication guidance. It calls for researchers, governments, and NGOs to pool validated Q&A data, strengthening global AI systems. Future work explores multimodal RLHF (image, voice, video), aiming to foster a global, inclusive, evidence-based ecosystem for AI agricultural advice.

Article History

Received: May 30, 2025
Accepted: August 5, 2025
Published: January 26, 2026

Keywords

agricultural advisory; generative AI; reinforcement learning from human feedback (RLHF); smallholder farmers; localization; expert annotation; language models; digital extension; farmer-centered design; multimodal AI; SDG 2: Zero Hunger

1. Vineet Singh, Chief Technology Officer, Digital Green, Heritage Landmark, # 304, 1st & 2nd floors, 1st Main Rd, 1st block, Koramangala, Bengaluru – 560034, Karnataka, India, vineet@digitalgreen.org
2. Sanyam Singh, Senior Machine Learning Engineer, Digital Green, Heritage Landmark, # 304, 1st & 2nd floors, 1st Main Rd, 1st block, Koramangala, Bengaluru – 560034, Karnataka, India, sanyam@digitalgreen.org
3. Lakshmi Pedapudi, Director of AI, Digital Green, Heritage Landmark, # 304, 1st & 2nd floors, 1st Main Rd, 1st block, Koramangala, Bengaluru – 560034, Karnataka, India, lakshmi@digitalgreen.org
4. Waseem Pasha, Engineering Manager, Digital Green, Heritage Landmark, # 304, 1st & 2nd floors, 1st Main Rd, 1st block, Koramangala, Bengaluru – 560034, Karnataka, India, waseem@digitalgreen.org
5. Ekta Kumari, Software Engineer, Digital Green, Heritage Landmark, # 304, 1st & 2nd floors, 1st Main Rd, 1st block, Koramangala, Bengaluru – 560034, Karnataka, India, ekta@digitalgreen.org
6. Rikin Gandhi, CEO, Digital Green, 2443 Fillmore St., #380-6502, San Francisco, CA 94115, rikin@digitalgreen.org
7. Alesha Miller, Chief Strategy Officer, Digital Green, 2443 Fillmore St., #380-6502, San Francisco, CA 94115, alesha@digitalgreen.org
8. Jona Repishti, Assistant Director of Strategy, Digital Green, 2443 Fillmore St., #380-6502, San Francisco, CA 94115, Jona@digitalgreen.org
9. Archana Karanam, Head of Global Product Operations, Digital Green, 2443 Fillmore St., #380-6502, San Francisco, CA 94115, archana@digitalgreen.org
10. Eric Firnhaber, Senior Director of Communications, Digital Green, 2443 Fillmore St., #380-6502, San Francisco, CA 94115, eric@digitalgreen.org

Introduction and Problem Statement

For nearly two decades, Digital Green has worked at the intersection of agriculture, community media, and technology to help smallholder farmers access the information they need to make informed decisions. Our earliest efforts focused on participatory video, supporting farmers—especially women—in co-creating short, locally relevant videos that promoted climate-smart and regenerative practices. These videos were distributed through community networks and shown in facilitated group settings, often with the support of frontline extension workers. Over time, this model produced more than 10,000 unique videos in over 40 local languages and reached over 3 million farmers, primarily across India, Ethiopia, and parts of West Africa.

While the impact of participatory video was strong in terms of adoption and behavior change, it required consistent investment in training, equipment, and frontline facilitation. It was not easily scalable without extensive partner infrastructure, such as government extension systems or donor-funded networks. Additionally, this approach struggled to keep pace with changing climatic conditions, pest outbreaks, or shifts in input prices—all of which require fast, responsive advisory systems. These constraints led us to explore digital alternatives that could complement or replace face-to-face approaches.

The first generation of digital tools we piloted were deterministic chatbots—rule-based systems designed to mimic conversation by routing users through a predefined tree of prompts and replies. While these worked in constrained settings, they quickly failed under real-world use. Farmers often asked questions that the chatbot couldn't recognize or respond to. The diversity of cropping systems, agroecologies, dialects, and advisory needs in smallholder agriculture made it nearly impossible to pre-code every possible interaction. Furthermore, deterministic bots couldn't understand photos or audio—key modalities for farmers with limited literacy.

In 2022, the public release of ChatGPT by OpenAI marked a turning point in the capabilities of natural language systems. For the first time, it was possible to engage with a model that could generate long-form, context-aware text in a conversational tone. Digital Green began testing early prototypes using GPT-based models in late 2022 and launched a WhatsApp-integrated assistant called Farmer.Chat in mid-2023. Initial deployments were conducted in India and Kenya, with Ethiopia and Nigeria joining soon after. Uptake was rapid: by May 2025, over 220,000 farmers and extension agents had used the platform, submitting more than 2.5 million questions.

However, our initial evaluations showed that general-purpose models—even when tuned with agricultural prompts—frequently returned answers that were too generic, overly academic, or poorly aligned with local conditions. For example, a farmer in Bihar might ask what to do about yellowing leaves on a wheat crop. The model might correctly identify a disease, but then recommend a chemical not registered in India, or fail to mention low-cost, culturally familiar alternatives like neem-based sprays. Sometimes, it missed the mark entirely, offering advice more suited to North American agribusiness.

These gaps were not due to a lack of sophistication in the model, but rather a misalignment in training data and contextual understanding. Large language models are typically trained on broad internet corpora dominated by content from high-income, English-speaking countries. They contain little to no information about minor crops, indigenous practices, or region-specific disease cycles. More critically, they do not reflect the voices or needs of the smallholder farmers who make up the majority of the world's agricultural labor force.

In response, we sought methods that could bring the model closer to the ground—closer to the everyday realities of farmers. This led us to explore Reinforcement Learning from Human Feedback (RLHF). RLHF is a technique initially developed by AI researchers to improve the performance of general-purpose LLMs. OpenAI, for example, used RLHF to align ChatGPT's responses with human preferences across dimensions such as helpfulness, safety, and clarity. In the health domain, OpenAI's (2025) HealthBench project engaged over 260 physicians from 60 countries to score thousands of question-answer pairs for realism, accuracy, and tone. The resulting dataset was used to fine-tune models to perform better in clinical settings.

Despite its success in other sectors, RLHF had not yet been widely applied in agriculture—particularly not in low-resource or smallholder contexts. We saw this as both a challenge and an opportunity. Agriculture is uniquely complex because it involves biological variability, seasonality, climate risk, local labor patterns, and market fluctuations. What is “correct” advice is often context-dependent, and different stakeholders may have legitimate, diverging views on what constitutes best practice. For example, a conservation agriculture purist may prioritize minimal soil disturbance, while a farmer managing livestock and seeking fodder may prefer partial tillage and residue removal.

This tension makes agriculture a prime candidate for an RLHF-based approach. Human reviewers—especially those with extension or farmer training backgrounds—can bring not just technical knowledge, but empathy, cultural nuance, and practical experience. By scoring and annotating model responses, they can steer AI toward advice that is not only agronomically sound but locally actionable, safe, and persuasive. Moreover, this approach enables a feedback loop: as farmers submit new types of queries or experience new climate shocks, reviewers can flag these edge cases, ensuring the model evolves in step with lived reality.

In summary, this manuscript outlines Digital Green's work to build an RLHF-driven pipeline for agricultural advisory content. We hope it contributes to the growing body of methods that aim to make AI more inclusive, context-aware, and grounded in the needs of those often left out of digital innovation.

Methods

The RLHF pipeline developed by Digital Green is designed to improve the contextual relevance, clarity, and trustworthiness of AI-generated responses to farmer queries. Our approach

integrates expert validation, feedback loops, and human-in-the-loop reinforcement in a replicable and modular framework. At the core of this pipeline is Farmer.Chat Evaluate, a web-based evaluation platform that facilitates the review and scoring of AI-generated answers by domain experts.

Overview of the RLHF Process in Agriculture

Reinforcement Learning from Human Feedback typically involves three components: (1) the generation of responses by an AI model, (2) human evaluation of those responses, and (3) the use of that evaluation to fine-tune the model. In our agricultural application, we adapted this to account for domain-specific needs and the multilingual, multimodal nature of real-world farmer queries.

Rather than working solely from synthetic prompts, we prioritized real questions submitted by actual farmers and extension workers using the Farmer.Chat assistant via WhatsApp, Telegram, and other digital interfaces. These included voice notes, typed text, and occasional images of diseased crops or livestock. By anchoring our dataset in user-generated content, we ensured that the evaluation process would reflect real use cases, not idealized scenarios.

Each query is linked to the model's original response, as well as any previous interactions in that conversation. These Q&A pairs are then presented to trained reviewers within Farmer.Chat Evaluate. Reviewers assess each response in two modes:

- *Absolute Evaluation*: The reviewer scores the AI-generated response on a 5-point scale based on accuracy, clarity, tone, safety, and contextual relevance. They may also suggest edits or corrections directly.
- *Comparative Evaluation*: In cases where two or more model responses are generated (e.g., from prompt variations or updated model versions), the reviewer is asked to indicate which one they prefer and why.

To ensure robustness and reduce individual reviewer bias, we implemented a system of partial redundancy: approximately 20% of responses are reviewed by more than one expert. An additional 10% of the dataset consists of "control" pairs—previously evaluated responses included to track reviewer consistency and flag drift in scoring standards.

Reviewer Recruitment and Training

Our initial team consisted of four expert reviewers, selected for their diverse agricultural backgrounds and practical field experience. Each reviewer had at least 5–10 years of experience as either a government extension officer, NGO-based agronomist, or lead farmer trainer. Two reviewers were based in Bihar, India, while others supported the project remotely. All were fluent in at least two working languages, such as Hindi and English, and familiar with regional crop cycles and pest management practices.

Digital Green conducted onboarding workshops to align reviewers on evaluation rubrics, scoring expectations, and platform functionality. Reviewers were paid for their time and provided with real examples from our production dataset for training. Over time, we

introduced structured feedback loops, including group review sessions and inter-rater reliability assessments, to ensure consistency.

As we scale the process, we plan to expand the reviewer pool to include more women, younger agricultural professionals, and experts from additional geographies (e.g., western Kenya, northern Nigeria). This is important not only for representation but for capturing the diversity of cropping systems and linguistic variation in smallholder contexts.

Data Sources and Selection Criteria

The data evaluated in this RLHF process comes from live queries submitted through the Farmer.Chat interface. Since the assistant is used by a mix of farmers, farmer producer organizations (FPOs), and extension workers, the dataset spans a wide range of topics, including:

- Crop-specific questions (e.g., “What should I spray for fall armyworm in maize?”)
- Livestock care (e.g., “My goat has a bloated stomach, what could be the cause?”)
- Natural input preparation (e.g., “How do I make jeevamrit?”)
- Climate adaptation (e.g., “Which paddy variety is drought-tolerant?”)
- Market and seed access questions (e.g., “Where can I buy PBW-343 in my area?”)
 - We prioritize Q&A pairs for evaluation based on:
 - Query frequency (high-volume questions)
 - Novelty or ambiguity (new pest or weather-related patterns)
 - Model confidence (low-confidence output, as flagged by internal quality scores)
 - Reviewer feedback (responses previously flagged as problematic)

By focusing on these high-impact queries, we aim to improve model performance where it matters most: where farmers are making time-sensitive decisions about inputs, disease control, and harvest timing.

Platform Features and Workflow

The Farmer.Chat Evaluate platform allows reviewers to:

- View the original farmer query, model response, and metadata (e.g., crop, language, region)
- Score the response across five dimensions using a Likert-type interface
- Provide comments, suggested edits, or links to local resources
- Compare multiple responses side-by-side and select a preferred option
- Track progress and consistency via automated dashboards

In the backend, Digital Green’s technical team aggregates these scores and annotations into a training dataset. High-quality, corrected Q&A pairs—referred to as “Golden Q&A”—are stored in a supervised fine-tuning dataset. Preference data (comparisons) are used to train a reward model, which helps guide the main model’s responses to better align with expert judgment.

Quality Assurance and Ethical Considerations

Throughout the process, we adhered to principles of fairness, transparency, and community ownership. All reviewers were informed of how their feedback would be used, and our system maintains full audit trails of scoring changes. To avoid reinforcing any single “truth,” reviewers are encouraged to flag situations where multiple legitimate answers exist or where local customs diverge from formal recommendations.

We are also developing protocols to flag and avoid unsafe or unethical content—such as use of banned pesticides, gender-insensitive advice, or misinformation. These are marked as “red line” responses and removed from future training sets.

Implementation Guide

To make RLHF methods actionable and replicable within the agriculture sector, Digital Green developed a five-phase implementation framework. This model can be adapted by governments, NGOs, agritech startups, and research institutions that want to localize LLM performance and continuously improve advisory quality over time. Each phase is interdependent and collectively forms a cycle of data-informed model refinement. Below, we describe each phase as a narrative to illustrate both process and considerations.

Phase 1: Data Collection and Preparation

The first step in applying RLHF to agriculture involves curating a dataset of real-world farmer queries and model responses. These queries typically originate from chat-based interfaces (like WhatsApp or Telegram) and include a mix of text, voice notes, and images. Voice and image inputs are transcribed or classified to enable consistent review. Each question is matched to the corresponding AI-generated response, along with metadata such as crop type, language, region, and time of year. This metadata is essential to support disaggregation and targeted performance improvements. Care must be taken to protect user privacy, de-identify sensitive information, and maintain auditability. At this stage, the goal is to gather a representative, diverse, and sufficiently large dataset to serve as the foundation for expert review.

Phase 2: Expert Review and Annotation

In the second phase, this raw data is reviewed and scored by trained agricultural experts using a standardized interface. Reviewers assess the quality of model-generated answers based on a 5-point rubric that incorporates accuracy, specificity, tone, cultural appropriateness, and safety. For example, a response that recommends a banned pesticide or omits key safety instructions would be scored poorly, even if the diagnosis is correct. To minimize bias, a portion of responses is double-reviewed and compared for consistency. Reviewers are recruited for both technical expertise and contextual understanding—many have backgrounds as extension officers, lead farmers, or agronomy consultants. In Digital Green’s implementation, reviewers were compensated fairly and received training in the use of the evaluation platform. The output of this phase is twofold: a corpus of “Golden Q&A pairs” with validated answers and a comparative dataset indicating which of two or more responses is preferred and why.

Phase 3: Reward Model Training

Once enough comparative annotations are collected, the third phase begins: training a reward model. This model is designed to predict which AI-generated responses are more likely to be rated highly by human reviewers. It does so by analyzing preferences from the comparative evaluations conducted in Phase 2 and learning the patterns associated with preferred outputs—such as use of farmer-centered language, practical detail, and region-specific examples. This reward model can then be used to guide the base LLM’s behavior in a reinforcement learning loop, nudging it toward producing responses that align more closely with human judgment. Training the reward model requires moderate machine learning infrastructure but can be done using widely available open-source tools. Care should be taken to avoid overfitting to a small group of reviewers or to unintended stylistic cues. A well-constructed reward model becomes a critical asset for continuous fine-tuning and helps the system improve in subtle, hard-to-code dimensions such as tone and empathy.

Phase 4: Model Fine-Tuning and Testing

In Phase 4, the base language model is fine-tuned using the Golden Q&A pairs from Phase 2 and, optionally, via reinforcement learning using the reward model from Phase 3. Fine-tuning can be accomplished through supervised learning or by adapting smaller parameter-efficient methods such as LoRA or PEFT when computing resources are limited. Testing should include both automated metrics (like BLEU, ROUGE, or model perplexity) and manual review of a held-out validation set. This testing is particularly important in agriculture, where even small deviations in language—for instance, misrepresenting a dosage or omitting a crop stage—can have serious consequences. We recommend including reviewers or agronomy leads in the testing process to verify that the model’s updated behavior reflects the intent of the feedback it received. At this point, the refined model can be prepared for deployment to end-users via apps, messaging platforms, or integration with other digital extension services.

Phase 5: Deployment, Monitoring, and Continuous Feedback

The final phase involves deployment of the improved model and the establishment of systems to collect ongoing feedback. This closes the RLHF loop and allows the process to remain dynamic and adaptive. Deployed models should include user-facing tools to capture simple feedback signals—such as thumbs-up/down ratings, follow-up questions, or self-reported satisfaction. Additionally, new queries received post-deployment should be periodically sampled and routed back into the review system to identify new edge cases or emerging issues. Monitoring dashboards that track performance by crop, language, geography, and season can help surface patterns, reveal bias, and guide future prioritization. This phase also includes continued engagement with reviewers—many of whom can serve as “community validators” to ensure that the model evolves alongside local practice and farmer experience. Over time, the goal is to establish a virtuous cycle: farmer queries → AI response → expert review → model improvement → more relevant farmer queries.

Results and Validation

The implementation of RLHF within the Farmer.Chat ecosystem has yielded early, promising results that point to significant improvements in both model performance and the relevance of AI-generated agricultural advice. In this section, we summarize the key findings from our initial rounds of expert review and model tuning. These include quantitative ratings of response quality, qualitative insights from reviewers, and early signals of improved alignment with real-world farmer needs.

Dataset Summary

As of May 2025, more than 25,000 Q&A pairs had been reviewed by trained experts through the Farmer.Chat Evaluate platform. These Q&A pairs were drawn from a larger pool of over 2.5 million user queries submitted across India, Kenya, Ethiopia, and Nigeria. Queries were selected based on frequency, novelty, or low initial confidence scores. Approximately 55% of the reviewed pairs were scored highly (4 or 5 out of 5), with a subset of these designated as “Golden Q&A” pairs for supervised fine-tuning. In total, 3,300 Golden Q&A pairs were validated by an internal review team and used in Phase 4 of the RLHF process.

The crop coverage of the dataset spanned more than 20 crops and livestock systems. These included cereals (e.g., wheat, rice), pulses (e.g., lentils, chickpeas), oilseeds (e.g., mustard), vegetables (e.g., tomato, brinjal), and dairy-related queries. Pest management, seed selection, organic treatments, stage-specific crop advice, and climate-resilient practices were among the most frequently asked topics.

Q&A Quality Improvement by Crop Category

To assess performance improvements before and after RLHF fine-tuning, we analyzed the percentage of responses rated 4 or 5 (on a 5-point scale) across key crop categories. Reviewers were blinded to whether a given response was pre- or post-tuning to avoid bias. The Table 1 below summarizes these results.

Table 1

Difference in & Rating of Crop Category Responses Before and After RLHF

Crop Category	% Rated 4–5 Before RLHF	% Rated 4–5 After RLHF
Cereals (e.g., wheat, rice)	48%	73%
Vegetables (e.g., tomato, brinjal)	32%	65%
Pulses (e.g., lentil, chickpea)	41%	70%
Oilseeds (e.g., mustard)	39%	68%
Livestock (e.g., dairy)	25%	50%

High-scoring responses often included location-specific examples, cost-saving options, and farmer success stories—features that helped build trust. In contrast, low-rated responses were often too generic or lacked regional nuance. Vegetables showed the largest relative gain in quality, while livestock remained underrepresented in the sample.

BLEU-Style Metrics and Human Evaluation

To supplement expert scoring, we computed BLEU-style similarity scores using reviewer-edited “ideal” responses as ground truth. While BLEU is not a perfect fit for open-ended advisory content, it served as a rough proxy for content alignment. Preliminary findings showed an average BLEU score increase of 0.18 after RLHF fine-tuning. BLEU scores were particularly helpful in flagging overly generic or irrelevant responses, though human reviewers remained the gold standard for detecting issues like tone, trustworthiness, or unsafe recommendations.

Discussion

Our implementation of Reinforcement Learning from Human Feedback (RLHF) demonstrates that it is both feasible and effective for improving the quality of AI-generated agricultural advisory. However, its value depends on more than just model accuracy—it must also be equitable, cost-conscious, and embedded in systems that prioritize local knowledge.

Equity and Representation

RLHF helps surface gaps in representation—especially in underrepresented crop systems, gendered labor roles, and indigenous knowledge. Our reviews revealed limited advisory for women’s crops, livestock systems, and agroecological practices. Including diverse reviewers is critical to closing these gaps and ensuring AI reflects real-world agricultural diversity.

Sustainability and Governance

To remain relevant, RLHF must be ongoing. Crop systems evolve, and climate shocks shift advisory needs. Governance models—such as reviewer networks embedded in extension systems or led by consortia—will be needed to maintain transparency, avoid drift, and ensure advisory integrity. Golden Q&A pairs should be openly licensed and clearly attributed to build public trust.

RLHF in a Broader AI Toolkit

RLHF is not a stand-alone solution. In some cases, retrieval-augmented systems or hybrid models that blend structured data and fine-tuned text generation will be more effective. RLHF plays a key role in shaping tone, usability, and local relevance—even when retrieval systems supply the base content.

A Call to Collaborate

We encourage global collaboration to pool validated Q&A pairs and create shared, open-source datasets for agricultural AI. Governments, researchers, and farmer networks can all contribute to a federated ecosystem of regionally grounded, locally reviewed, and continuously improving training data.

Next Steps: Toward Multimodal RLHF

Expanding RLHF into image and voice modalities is a natural next step. With Farmer.Chat already receiving thousands of voice notes and images, there is urgent need to build reviewer protocols for visual symptom detection, tone assessment, and multimodal safety checks. This will improve inclusion for farmers with limited literacy and help tune AI to real-world communication styles.

References

OpenAI. (2025). *HealthBench: Evaluating LLMs on medical QA*.
<https://openai.com/research/healthbench>

© 2026 by authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).