PROBES snapshot

An agricultural knowledge assistant using large language models

A probe submitted by Mukund Patil – ICRISAT – Digital Transformation Accelerator

Prepared by Grecia Benavente | DTA Program Manager & Daniel Jimenez R | Senior Scientist, DTA AoW3 lead - Alliance of Bioversity & CIAT

What is a probe?

PROBES are early-stage, safe-to-fail digital experiments launched under the Digital Transformation Accelerator to explore new tools and ideas across CGIAR. They are designed to test concepts in real contexts, learn what works and what does not, and generate insights that inform future design, prototyping, or scaling decisions.

Imagine it: Instead of searching through thousands of research articles, you ask a question, and your knowledge assistant locates the right scientific answer for you.

The promise

Agricultural scientists often spend significant time searching for accurate information buried in large – and growing – research repositories. This probe tests whether domain-specific large language models (LLMs) can cut through information overload and ensure scientists receive grounded, factual responses. Such tools can accelerate research cycles, strengthen evidence-based decision-making, and reduce risks associated with misinformation or unsupported AI-generated content.

What's innovative

This probe applies frontier AI approaches to agricultural research, such as retrieval-augmented generation (RAG), semantic search, domain-specific embeddings, and knowledge graph linking. The team grounded the assistant in more than 11,000 validated scientific documents from ICRISAT's Open Access Repository. It is one of CGIAR's first attempts to build a scientific LLM for agricultural research workflows.

What we did and learned

The team extracted, cleaned, chunked, and embedded content from PDFs, stored it in a vector database, and integrated a RAG system using an open-source LLM.

They implemented semantic search, built a retrieval pipeline, and began testing accuracy, relevance, and hallucination mitigation. The probe confirmed that this domain-specific LLM significantly reduced misinformation compared with general models, although it performed better with formal technical documents than non-technical inputs. Results also highlighted the importance of a human-in-the-loop to improve grounding and the need to balance accuracy with compute cost.

What's next

- Refining the model to answer reliably and handle poorly defined queries
- Improving data standardization and governance
- Piloting the assistant with scientists in India and sub-Saharan Africa

SDG alignment

Submit a probe

Are you a working with an emerging digital technology that might contribute to CGIAR's 2030 outcomes? Would you like to try this in a safe-to-fail environment?

Submit here