REGENERATIVE AGRICULTURE RESEARCH ENTERPRISE (RARE) FEASIBILITY STUDY

Illustration generated by Leonardo.AI

Assessing the potential for leveraging Artificial Intelligence (AI) to support the advancement and scaling of smallholder regenerative agriculture in the Global South, starting in South America.

Made possible through a grant from the VF Foundation.

Comments on the Regenerative Agriculture Research Enterprise (RARE):

"What excites me most about the RARE project is its commitment to valuing and elevating both scientific innovation and traditional agricultural knowledge. Together, these efforts support the creation of a global agricultural knowledge commons—one where everyone, everywhere, can access the best possible agricultural knowledge in their own language. This empowers communities to improve their environments and livelihoods while strengthening both food and data sovereignty."

—Dorn Cox, Founder, OpenTEAM / *USA*

"Having seen firsthand how smallholder farmers in the Global South drive regenerative agriculture R&D through both scientific innovation and traditional knowledge, I believe RARE has the potential to unlock a vast, untapped wealth of field research. These insights are too often missing from global discourse, and without a more holistic, well-communicated database, regenerative agriculture innovations—whether in science or within farming communities—struggle to scale and attract the funding and policy support they need to thrive."

—John Mundy, Director of Global Partnerships,
One Acre Fund / *Africa*

"The RARE project holds significant potential for advancing the scaling of regenerative agriculture among smallholder farmers in the Global South."

—Bita Corera, Executive Director, Foundation For The Application
And Teaching Of Science (FUNDAEC) / *Colombia*

"For the past 28 years, Sustainable Harvest International has run a regenerative agriculture extension program in Central America, often finding it difficult to access much of the relevant research needed to best serve the thousands of farmers in our network. The RARE project could be a tremendously valuable tool for our work as we move towards our scaling goal of a million farms."

—Florence Reed, Founder + Director of Strategic Growth,
Sustainable Harvest International / *Central America*

"In relation to the RARE project, it is in our interest to contribute with our expertise to the expansion of forms of agricultural production on a regenerative basis for family farmers in Brazil, and an expanded way in the Global South."

—Carlos Armenio Khatounian, Professor at Luiz de Queiroz College of Agriculture, University of São Paulo / *Brazil*

"We drew heavily on research conducted in other countries when working with our agronomists and farm leaders to design our first smallholder regenerative program here in Haiti. I think the RARE project has tremendous potential to unlock more of this kind of research that will be helpful as we continue to scale regenerative agriculture here, and an invaluable tool to others throughout the Global South."

—Timote George, Co-founder and Executive Director, Smallholder Farmers Alliance / *Haiti*

"I believe that RARE has a great potential to positively influence regenerative agriculture, especially among smallholder farmers in Colombia and other countries of the Global South."

—Jairo Leonardo Cuervo Andrade, Associate Professor at National University of Colombia / *Colombia*

"Applying artificial intelligence in support of smallholder regenerative research, as the RARE project sets out to accomplish, represents an important step forward in scaling the overall transition to regenerative agriculture in the Global South."

—Rob Johnson, CEO, Acceso / *Colombia, El Salvador, Haiti*

"It is our hope that the Federal University of Mato Grosso can be part of RARE, which we think has potential here in Brazil because its central theme is the scaling up of regenerative agriculture by smallholder farmers."

—Daniela Campos, Professor at Federal University of Mato rosso do Sul / *Brazil*

Regenerative Agriculture Research Enterprise (RARE) Feasibility Study

by Hugh Locke, Tim Tensen, Charles Darling, Chris Kaput, Nayla Almeida

March 3, 2025

The RARE Feasibility Study examines the role of Artificial Intelligence (AI) in unlocking, translating, disseminating and supporting smallholder regenerative agriculture research from the Global South, with an initial focus on South America. It evaluates current research availability, reviews AI translation capabilities, and explores partnerships with universities, government agencies, and NGOs to create an accessible and open-source knowledge-sharing platform for scaling regenerative agricultural practices.

Ultimately, RARE aims to become the world's largest multilingual database on smallholder regenerative agriculture, bridging scientific and traditional knowledge to support and accelerate the scaling of regenerative practices across the Global South.

This study was made possible through a grant from the VF Foundation.

Contact information: Hugh Locke, hugh@futurra.org

Contents

Executive Summary

The Regenerative Agriculture Research Enterprise (RARE) Feasibility Study, conducted by [Futurra] in partnership with the [Impact Farming Foundation] and funded by the [VF Foundation], explores how Artificial Intelligence (AI) can enhance the accessibility and scalability of smallholder regenerative agriculture research in the Global South.

The study reveals a critical lack of accessible research, finding that less than one tenth of one percent of scientific and academic library collections cover smallholder regenerative agriculture, while an estimated 30,000 relevant research documents remain inaccessible on restricted university, government, and NGO servers across South America. To address this challenge, RARE proposes creating an open-source, AI-powered library that would host currently unavailable research and incorporate AI translation tools to make it multilingual and fully searchable.

RARE's AI translation approach involves a hybrid model that combines large language models (LLMs) and small language models (SLMs), along with a retrieval augmented generation (RAG)-based multilingual glossary to ensure high translation accuracy, particularly for technical agricultural and culturally specific terms. While LLMs achieve slightly higher text translation accuracy than SLMs, they struggle to maintain document structure and visuals, whereas SLMs preserve images, graphs, and layouts far more effectively. This makes SLMs the preferred choice for RARE's platform, with LLMs and their more extensive reach helping to develop and continually update the glossary.

To further support knowledge generation, the study recommends establishing a RARE Regional Network of academics, agronomists, and researchers to source and curate research, help create and maintain the RAG-based glossary, coordinate and implement new research through grants, and uphold peer-review standards. Additionally, while RARE's primary mission is to be a free and open-source platform), the study suggests exploring self-financing opportunities through data-driven services such as trend mapping, agricultural forecasting, and consulting for institutions, corporations, governments, and NGOs.

The next step, RARE Phase 1, will focus on implementing a pilot program in South America, where the initiative will begin by developing a minimum viable product (MVP) of the online library and hybrid AI translation system. The project will establish partnerships with universities, NGOs, and government institutions to expand research accessibility while conducting user testing and feedback to refine platform functionality. Small grants will initiate the process for generating new research, while long-term financial sustainability models will also be explored.

Theory of Change

RARE aims to contribute to the global adoption of smallholder regenerative agriculture by creating an accessible and open-source knowledge-sharing platform for scaling regenerative agricultural practices.

A quiet revolution is taking shape in rural communities throughout the Global South as hundreds of thousands of smallholder farmers are switching to the practice of regenerative agriculture and agroforestry. This new holistic system of farming combines scientific and applied research with indigenous and ancestral traditions, and the result is increased food production along with measurable and positive environmental, social and economic impacts.

Although not yet widely known or understood by the public at large, regenerative agriculture is placing long overlooked and marginalized smallholder farmers at the forefront in addressing the interconnected global challenges of climate change, declining food sovereignty, widespread biodiversity loss, and large-scale migration.

The Regenerative Agriculture Research Enterprise (RARE) is based on the premise that scaling regenerative agriculture will require exponentially more scientific and applied research than has been conducted to date. Our theory of change is that building the world's largest multilingual database of research on smallholder regenerative agriculture can be a key factor in its continued expansion throughout the Global South.

The RARE Feasibility Study quantifies the existing body of research on this topic. The study examines what portion of this research is currently freely available online, while also estimating the volume of documentation that remains essentially invisible due to being stored on restricted-access local servers in the Global South. The study sets out a methodology for securing this currently inaccessible research and analyzes the AI tools available for accurately translating these documents in order to render them a valuable public resource.

For the purposes of this Feasibility Study, the geographic focus has been limited to South America. The results and recommendations set forth in the following are intended to inform the design and implementation of a pilot operation—RARE Phase 1—that will also focus on South America before eventually scaling to include Africa, Asia and Oceana.

Background

RARE was conceived in response to the need for research-driven solutions to food security, aiming to build an open-source, multilingual database to unlock and expand access to smallholder regenerative agriculture research.

In January of this year, over 150 Nobel and World Food prize laureates signed an open letter calling on world leaders to engage in “moonshot” efforts to ramp up global food production before an impending world hunger catastrophe. Central to their message is the need for “basic and applied research” to improve agriculture. “We must be prepared to pursue high risk, high reward, scientific research,” states the letter, “with the goal of transforming our food systems to meet the nutritional needs of everyone sustainably.”

RARE has been conceived by Futurra in partnership with the Impact Farming Foundation and is a direct response to this clarion call. The ultimate goal is to deploy Artificial Intelligence (AI) in building the world’s largest open-source, multilingual and fully searchable database of academic, scientific and field research on regenerative and related sustainable agricultural and agroforestry methodologies being implemented by smallholder farmers and ranchers in the Global South.

RARE started with this Feasibility Study, funded by the VF Foundation, in order to assess the need, design and eventual implementation of this undertaking. It follows on from the groundbreaking series of seven reports, produced by the Futurra team and sponsored by The Rockefeller Foundation, that explore the definition, implementation, verification and scaling of smallholder regenerative agriculture globally.

In responding to the challenge put forward by the 150 laureates in their moonshot appeal, RARE takes its cues from the benefits seen by Green Revolution’s use of extensive research and scientific advances that led the successful scaling of industrial agriculture during the 1950s and ‘60s, albeit with significant and ongoing environmental impact. Historically, smallholder agriculture has been largely excluded from this same category of research despite estimates by the Food and Agriculture Organization of the United Nations (FAO) that smallholder farmers currently produce around a third of the world’s food and up to 80 percent in some countries of the Global South.

Exclusion from research has begun to change recently with the emergence of the smallholder regenerative movement in the Global South. This development has generated an entirely new level of interest, particularly because regenerative methodology emphasizes the alignment of science and traditional knowledge systems.

This growing interest has, to date, yielded a somewhat limited corpus of research due to the following factors, each of which has been explored in depth as part of this Feasibility Study:

- The major library platforms specializing in scientific and academic research have huge online repositories that include agriculture as a subject area, but only a tiny fraction of their collections focus on smallholder farming and an even smaller fraction can be found on the topic of smallholder regenerative farming;

- These library platforms can locate documents and make them available (some free and some charge for access), but none are able to offer in-house translation services and none are able to do full-text scans of the documents in their collections;

- Major search engines that specialize in scientific and academic research can reach further afield than library platforms to conduct literature searches, but their translation is more geared to general meaning and keyword recognition for information retrieval rather than the technically accurate translations possible with AI-powered translation tools

- Major search engines specializing in scientific and academic research can access a broader range of sources beyond traditional library platforms. However, their translation capabilities are primarily designed for information retrieval, focusing on general meaning and keyword recognition rather than ensuring technical precision. AI-powered translation tools offer more context-aware and accurate translations, particularly for specialized and culturally-specific terminology;

- Whether via library platforms or search engines, much of the smallholder regenerative research that does exist is in one local language and currently inaccessible on local restricted servers of universities, government agriculture departments, and smallholder-focused NGOs throughout the Global South; and

- If that research was somehow to be made available on a library platform, it would remain in its original language without translation. If it was made open to search engines, the translation technology used to retrieve information would fail to retain the original meaning of many of the technical and cultural terms used to present research on smallholder farming.

These factors led to the concept of RARE as a specialized open-source library platform that can host research documents on smallholder regenerative agriculture from sources that are currently inaccessible. The recommendation is to have RARE break the norm for online libraries by incorporating the best AI translation tools available to make the entire collection multilingual, and further break the norms of library platforms by making the contents of all documents fully searchable via the site. Having the RARE platform host the highest quality multilingual translations will also overcome the problem of search engines having to rely on their translation capacity when conducting literature searches.

Key Findings & Recommendations

The RARE Feasibility Study identified 30,000 inaccessible smallholder regenerative agriculture documents in South America, highlighting the inadequacy of existing online libraries and AI search engines for multilingual research access. It recommends a hybrid AI translation approach using small language models (SLMs) and a RAG-based glossary to ensure culturally accurate translations. To support research expansion, RARE proposes a regional network to maintain the glossary, coordinate grants, and uphold peer-review standards. Lastly, it suggests that RARE remains open-source while exploring self-financing through data-driven services for institutions, corporations, and NGOs.

Note that A1, A2, etc. refer to specific appendices.

The Regenerative Agriculture Research Enterprise (RARE) Feasibility Study set out to examine the need and practical considerations for developing a specialized online library platform designed to share and translate smallholder regenerative agricultural research.

For this study we contacted universities in Argentina, Brazil, Colombia and Peru to determine the region's existing body of relevant research and how to access it. We later added government agriculture departments in Brazil and two smallholder-focused NGOs—one in Colombia and one in Haiti (see list on page 18).

Following are the study's five main categories of findings and recommendations.

1

The RARE study found that research on smallholder regenerative agriculture is scarce in online scientific libraries, difficult to access via AI-powered search engines, and hindered by inadequate translation of technical and cultural terms.

The first step in our year-long exploration was to determine the volume of relevant research that is currently available online and searchable in multiple languages. Here is what we found:

- **Online scientific libraries provide a very limited volume of topic-specific research.**

  We found that research and general information on the subject of smallholder regenerative agriculture represents less than one tenth of one percent of the more than 110 million documents located on the most widely used scientific and academic library platforms (A2).

  These platforms allow you to find and download documents from their respective collections, but they do not offer an in-house option to translate those documents. In addition, they do not offer full-text search options and are limited to meta-based searches (title, author, abstract and keywords).

- **AI-powered scientific search engines have wide reach but significant limitations.**

  Two of the most widely used specialized scientific and academic search engines do not house documents on their sites, but instead scour the internet and search the contents of relevant documents found on publicly accessible websites.

  We found that while they may be able to access more documents than scientific libraries, these search engines have limitations when accessing the contents of documents that are in multiple languages because the technology they use has limited capacity to accurately translate technical and culturally-specific terms and phrases that are not part of standardized universal scientific terminology (A2).

2 **The RARE study suggests at least 30,000 inaccessible research documents on smallholder regenerative agriculture exist in South America and recommends creating an open-source platform to make them freely available, emphasizing the importance of field research and its role in advancing regenerative agriculture.**

The above findings regarding the volume and accessibility of existing research led the RARE Feasibility Study team to look elsewhere for research docyments, noting that at this point we narrowed the scope of our investigation to South America in the interests of practicality. This is what we learned:

- **A significant amount of research exists but is currently inaccessible because it is on restricted sites.**

  We reached out to universities, government agencies and NGOs in South America to learn if there are research documents on smallholder regenerative agricultural practices (which includes individual practices that are also found within agroecology, organic and other sustainable methodologies) that are on local or restricted websites that are not accessible by search engines. The answer to our admittedly non-scientific and incomplete search was that there are most likely at least 30,000 such documents relating to smallholder regenerative and sustainable agriculture in the South American region (A2).

  We asked this same group if the research they were responsible for could be made available free of charge, and the answer was resoundingly positive. This led us to the recommendation to create a dedicated RARE platform to house such a collection and make it free and open-source.

- **Institutions in the Global South produce more field research, which is important when it comes to smallholder farmers.**

  The various academics, agronomists and researchers we spoke with pointed out that much of what they were involved in was 'field research.' Research in a broad sense encompasses all systematic methods of gathering and analyzing information on a topic. This can include studying existing literature, analyzing data, conducting experiments in controlled environments, or reviewing historical documents.

  Field research involves collecting data and making observations in real-world settings rather than in a controlled environment or through secondary sources. This clarification placed even greater importance on finding research conducted by the institutions in closest proximity to smallholder farmers, which in this case is the universities, government agencies and NGOs in countries with resident smallholder farmer populations.

- **There is a direct connection between research and advancing regenerative agriculture.**

  Research is about discovering new knowledge and understanding fundamental principles. Research and development (R&D) involves taking the results of scientific investigation, in this case focused on smallholder regenerative agriculture, and translating that into tangible advances in the methodology, technology, financing and policies that continue to evolve and shape regenerative agriculture as practiced by smallholders.

3 **The RARE study recommends a hybrid AI translation approach using SLMs and a RAG-based glossary to create a multilingual library of smallholder regenerative agriculture research, ensuring high translation accuracy while preserving indigenous and technical terminology.**

Having determined that there is a considerable volume of field research that could be part of a new online RARE library platform, the next issue was to explore the best way to incorporate the highest quality AI-powered translation service to make the entire collection multilingual and full-text searchable, including by external search engines. Here is a summary of the results of this exploration:

- **Large language models (LLMs) averaged 90% accuracy in translations, but struggled to maintain visuals and layout.**

  We began by looking into two of the most widely used LLMs: GPT-4 (the underlying language model behind ChatGPT) and Claude 3.5. These and other LLMs learn by drawing on text collections, known as datasets or corpora, that include vast internet data. LLMs excel at general language tasks but struggle with the nuances of technical writing, including jargon, complex sentence structures, and contextual understanding.

  We undertook a comparison test by taking four research papers from Brazil on smallholder agriculture—dealing with liquid biofertilizers, the effect of soybeans on soil microbiology, and two papers on green manure—and having them translated from Portuguese into English. The detailed breakdown is in appendix A4, but the overall percentage of accuracy for technical and cultural terms and phrases was 1) professional translator: 100%, 2) GPT-4: 90.13%, and 3) Claude 3.5: 88.9%.

  The LLMs did a terrible job of maintaining photos, graphic images and graphs, most of which were lost in their translations. This reduces the usefulness of the end product because visuals and document structure are often critical in giving the complete context for a full understanding of the text translations. There were also several instances where complete paragraphs were dropped by the LLMs.

- **Small language models (SLMs) averaged 80% accuracy in translations, but did a very good job of maintaining visuals and layout.**

  These are essentially smaller and purpose-built versions of their LLM counterparts. While they learn from a less robust corpora and lack the breadth of LLMs, they are easier to fine-tune for specific tasks such as translating. They also use significantly less computational power than LLMs.

  We looked at two of the most popular SLM-supported translation services—Google Translate and DeepL Pro—and did a test of the four Portuguese documents translated into English. The detailed breakdown is in appendix A3, but the overall percentage of accuracy for technical and cultural terms and phrases was 1) professional translator: 100%, 2) DeepL Pro: 82.5%, and 3) Google Translate: 79%. While these two SLMs scored lower than the LLMs, both DeepL Pro and Google Translate were excellent at maintaining photos, graphic images and graphs.

  We conducted a second test by using both SLMs to translate a French language regenerative cotton training manual from Haiti into English and comparing the results to a professional translation of the same document. DeepL Pro had an accuracy rating of 87% and Google Translate was 78%. However, this 55-page document had a very high level of images (37) throughout, and both SLMs maintained the visuals and document structure with a high degree of accuracy in their respective translations.

- **Glossaries are the key to improving translation accuracy.**

The performance of both LLMs and SLMs can be improved by the use of multilingual translation glossaries. There are two basic approaches when it comes to creating these glossaries.

The first involves manually creating an Excel or CSV format table with 1) the original word or phrase, 2) the translation of that word or phrase, and 3) various linguistic and contextual details such as part of speech, example sentences, domain or category, synonyms, definitions, priority levels, and source references to ensure accuracy and consistency in translation.

The second approach is to use a retrieval augmented generation (RAG) system, which is a tool for creating and continually updating a more comprehensive multilingual translation glossary. A key feature of RAG systems is to act as an intermediary, retrieving definitions and contextual information from up-to-date technical glossaries, incorporating input from subject matter experts, and using the resulting glossary to improve the accuracy of the LLM or SLM during the translation process. This external knowledge injection allows the language model to generate more accurate, contextually appropriate, and professional translations of technical documents.

- **The study recommends a hybrid AI translation solution using a RAG-based glossary.**

We found that online libraries can house large numbers of documents, but cannot translate or scan their contents. By contrast, search engines can find documents and do multilingual searching, but their translation tools have a very low level of accuracy when it comes to technical and culturally-specific terms and phrases. LLMs and SLMs, on the other hand, can translate to a much higher standard of text accuracy and particularly when enhanced by a translation glossary, but SLMs alone are able to maintain visuals and layout.

Based on this analysis, we recommend a hybrid approach that overcomes the restrictions of both online libraries and search engines. To do this we propose creating RARE as an online library that, unlike the current norm is 1) capable of translating a resident document collection to the highest levels of accuracy, and 2) capable of incorporating a feature for full-text searching the contents of those multilingual resident documents.

SLMs represent the better choice of translation model for RARE because, while roughly on par with LLMs for text translation, they significantly outperform when it comes to maintaining visuals and layout.

As all AI translation tools rely on glossaries, we recommend bypassing the Excel of CSV manual versions to create an external RAG-based multilingual translation glossary that can be updated using, among other sources, the extensive reach of LLMs. The resulting multilingual library collection will be open to search engines that will be able to scan highly accurate multilingual translations that overcomes the handicap of the low accuracy of in-house search engine translation.

- **Glossaries are particularly important for accurately translating indigenous and cultural terminology.**

Achieving accuracy ratings between 80 to 90% when translating the agricultural and cultural terms and phrases in research documents is a problem. This is particularly true because the adaptation of regenerative practices in response to local agricultural, cultural and climatic conditions is key to the overall regenerative methodology. When you then factor in translating tens of thousands of research documents into multiple languages, the 10 to 20% gap in accuracy is considerably exacerbated.

It is important to note that the results in the translation tests for this study were between more common languages (e.g. Portuguese, English and French) and does yet factor in the potential additional challenges that come with translating terminology from indigenous languages. To this point, one group that caught our attention was the Translation Glossary Project. They produced a Chichewa-English translation glossary of development terminology with an indigenous community in Malawi. It included dozens of Chichewa language terms that are not in published dictionaries.

Here are just a few of the agricultural terms and expressions in Brazilian Portuguese that are used by smallholder farmers throughout the country, and for which there are no widely accepted translations into other languages:

- **amontoa:** agricultural practice of piling soil around the base of plants to strengthen the stem and protect the roots.
- **roçado:** area of land where vegetation has been cut and cleared for farming or pasture.
- **enxadão:** agricultural tool similar to a hoe, but larger and heavier, used for tilling soil and weeding.
- **roça:** traditional cultivation system practiced by indigenous peoples, river-adjacent communities and family farmers in Brazil.
- **raleio:** technique of removing excess plants or fruits to improve the development of the remaining ones.
- **barreado:** soil management practice involving the construction of barriers to control erosion and retain water.
- **sementes crioulas:** traditional seed varieties, preserved and reproduced by traditional communities over generations.
- **terra preta de índio:** fertile, dark soil enriched with organic matter and charcoal, created by indigenous peoples of the Amazon.
- **roça de toco:** traditional farming system involving cutting and burning vegetation (coivara) to prepare the soil for planting.
- **capoeira:** similar to the indigenous roça system, based on the natural regeneration of the soil through secondary vegetation.

4 **The RARE study proposes establishing a regional network of academics, agronomists, and researchers to uncover inaccessible research, coordinate new funding, and uphold peer-reviewed standards for a high-quality multilingual research library.**

Having located a source of existing but inaccessible research on smallholder regenerative and sustainable agriculture, and having determined the best way to translate this research and make it accessible and full-text searchable for free via a RARE library platform, we then considered how to generate new research moving forward. Here is what we found:

- **There is an existing network of academics, agronomists and researchers who, with proper funding, are ready to take field research to a new level.**

  In the course of reaching out to universities, government departments of agriculture, and smallholder-focused NGOs in connection with this Feasibility Study, it became clear that those we spoke with represent an invaluable and largely untapped resource when it comes to scaling smallholder regenerative agriculture.

  This led to our recommendation to set up a RARE Regional Network of academics, agronomists and researchers to both find the existing documents as well as helping to coordinate and allocate new funding for research by members of the network along with setting guidelines for peer review to ensure the highest standards for research documents housed by the RARE library platform.

- **This network will be ideal for developing the translation glossary.**

  Another important role will be to have this network contribute to the RAG-based glossary by collaboratively curating, refining, and expanding its content, ensuring that it remains a dynamic and authoritative resource for smallholder agriculture. Drawing inspiration from Wikipedia's model, the glossary can be maintained through continuous input from academics, agronomists, and researchers who validate terms, add contextual insights, and integrate emerging regenerative agriculture practices.

  Members can also engage by submitting new culturally specific terms, providing translations, and sharing real-world case studies to enrich the glossary's practical relevance. Regular discussions and workshops foster knowledge exchange and adaptation to evolving agricultural challenges. By actively shaping and sustaining this resource, the network will ensure that smallholder farmers, policymakers, and extension workers have access to a continually evolving, context-sensitive, and linguistically inclusive agricultural knowledge base.

The RARE study recommends that the proposed online platform be set up as a free, open-source library supporting research, program development, and outcome frameworks while exploring self-financing options through data-driven services for institutions, corporations, and NGOs 5

The final issues we explored were about who would use RARE, what should it cost them to use it, how should it be financed to become operational, and who will use it. Here is what we recommend:

- **RARE is poised to meet a real need that is currently not being addressed.**

  We heard back from university professors that they would value having access to the RARE online library as a learning tool as well as for developing new field research programs. NGO agronomists and project developers expressed interest in accessing the library as they design and refine smallholder regenerative farm and agroforestry programs. Others raised the potential for RARE to be a source of data for governments, corporations, philanthropies and NGOs involved in planning and financing sustainable development either for the common good or for product procurement.

- **RARE can be an important tool for contributing to regenerative outcome frameworks.**

A recurring theme in our investigation was the idea of RARE providing critical data on local adaptation indicators and metrics in support of the various regenerative outcome frameworks currently being developed.

Three such frameworks were reviewed as part of our study: 1) RegenScore is U.S.-focused with no Global South presence, but their framework for on-boarding, scoring, and verification is excellent, 2) Regen10 is international and their framework is very comprehensive, and 3) Textile Exchange has a Regenerative Agriculture Outcomes Framework that, although specific to textiles, is nonetheless well regarded for being very detailed and targeted to smallholder producers.

- **While RARE should be a free basic service, it should also have secondary services that are fee based in order to become self-financing.**

Our team was unanimous in recommending RARE be free and open-source in order to fulfill its mission “to support the advancement and scaling of smallholder regenerative agriculture in the Global South.” For this reason we will go after new philanthropic funding to implement the operational pilot RARE Phase 1 in South America.

We also want this pilot to include research into the kind of services that could eventually be offered on a fee basis in order to make the service self-financing. While keeping the document collection open to all and full-text searchable for free, it would seem entirely possible to draw on the collective data in the library and be able to charge fees to map trends, engage in agricultural and economic forecasting, and/or provide consulting services for institutions, corporations, governments and NGOs that goes beyond information retrieval. Built into this system could be a portion of the income going back to the universities and NGOs that have helped to assemble the RARE collection.

Next Step: Making RARE Operational

Conducting this RARE Feasibility Study, and taking into account its focus on South America, has clearly shown the need, role and potential impact that this project can have in helping to advance and scale smallholder regenerative agriculture in the Global South.

RARE Phase 1 will take the recommendations from the study as the basis for implementing a RARE pilot program in South America. The insights and outcomes from Phase 1 will subsequently inform the later expansion of RARE throughout the Global South in Phase 2.

RARE Phase 1 will develop a minimum viable product (MVP) of a specialized online library with a translation service. This includes:

- **Establishing a robust RARE Regional Network by:**

- Forming partnerships with universities, government ministries of agriculture, and smallholder-focused NGOs throughout South America;
- Creating protocols for network participation and collaboration;
- Developing systems for coordinating and peer-reviewing existing and new research; and
- Setting up mechanisms for network members to contribute to knowledge sharing.

- **Building the technical infrastructure to:**

- Create an open-source library platform capable of processing, translating and storing an estimated 30,000 documents from the region;
- Translate all documents into Spanish, Portuguese, French and English;

Develop streamlined systems for receiving, cataloging, and managing research documents; and

Implement comprehensive full-text search functionality across the entire content of all documents.

- **Implementing a hybrid AI translation system as a key element of the technical infrastructure and featuring:**

- Integration of small language models (SLMs) for high-accuracy technical translation;
- Support from large language models (LLMs) for training and improvement of the SLMs;

- Development of a RAG-based multilingual translation glossary;

- Collaboration with the Regional Network to continually update and improve the glossary;

- Special focus on preserving terms specific to local agricultural and cultural traditions; and

- Developing the capacity of translations to retain the formatting and layout of the original documents as well as ensuring the retention and integrity of photos, graphs, charts and other digital graphic elements.

- **Conducting user testing and feedback to improve platform features and functionality through:**

- Early product testing with network members;

- Regular feedback sessions for product development road mapping; and

- User experience evaluation from different stakeholder groups.

- **Establishing a pilot grant program including:**

- Distribution of small grants to participating institutions, particularly to facilitate graduate student-level searching and evaluation of relevant research documents for including in the platform;

- Testing of new funding procedures and mechanisms;

- Evaluation of grant effectiveness and impact; and

- Extensive research into future sources of philanthropic, institutional, governmental and corporate funding for future expansion of the grant program.

- **Creating analysis and reporting capabilities as part of exploring future self-funding possibilities for RARE by:**

- Identifying the needs of the institutions, corporations, governments and NGOs involved in financing, supporting and implementing smallholder regenerative agriculture programs;

- Exploring the technical and structural requirements needed to draw on data in the RARE library collection to map trends in regenerative agriculture practices; and

- Exploring best practices in agricultural and economic forecasting.

Meet the RARE Team

The authors of the RARE Feasibility Study are:

Hugh Locke is engaged in smallholder regenerative agriculture and agroforestry as president and co-founder of the Smallholder Farmers Alliance in Haiti, co-founder of both Futurra and VeriFi Data Corp. Hugh is President of the Impact Farming Foundation, a board member of One Tree Planted, and author of The Haiti Experiment (2012). His TEDx talk is “Trees are Financing a Farming Revolution in Haiti” (2017) and he is featured in the documentary “Kombit: The Cooperative” (2017).

Tim Tensen is a designer and executive working to advance data sovereignty for farmers, true-cost accounting in our materials systems, and develop a more cohesive thesis for an agricultural data economy. He is a co-founder of Futurra, co-founder and President of VeriFi Data Corp, and CEO of Terra Genesis where he works with food and beverage, fashion and apparel, and cosmetics and home goods brands, farmers across the planet, academics, and government to test and implement methods for advancing regeneration.

Charles Darling is a founder and investor with extensive experience across multiple sectors including climate tech, B2B SaaS, and medtech in both the US and internationally. He currently serves as the CEO of VeriFi Data Corp., a data contracting platform for sustainable agriculture, and is a co-founder of Futurra. Additionally, he lends his time as a Strategic Advisor to Motif Neurotech and as a Venture Partner to Elite Capital Advisors. Charles holds a Bachelor of Arts from Columbia University.

Chris Kaput is a regenerative agriculture and development specialist with extensive experience across Latin America and the Caribbean. As the co-founder of Regenerates, he collaborates with companies, organizations, and smallholder farmers using a whole-systems approach to develop equitable agricultural supply chains that foster lasting environmental and social resilience.

Nayla Almeida is an agronomist and MSc in Rural Development, specializing in Southern Agriculture markets, policies, and sustainable production systems. Experienced in working with smallholder farmers in the global south, focusing on agroecology, market access, and food security. Currently engaged in designing and managing rural development projects in regenerative agriculture, sustainable production, and public food policies.

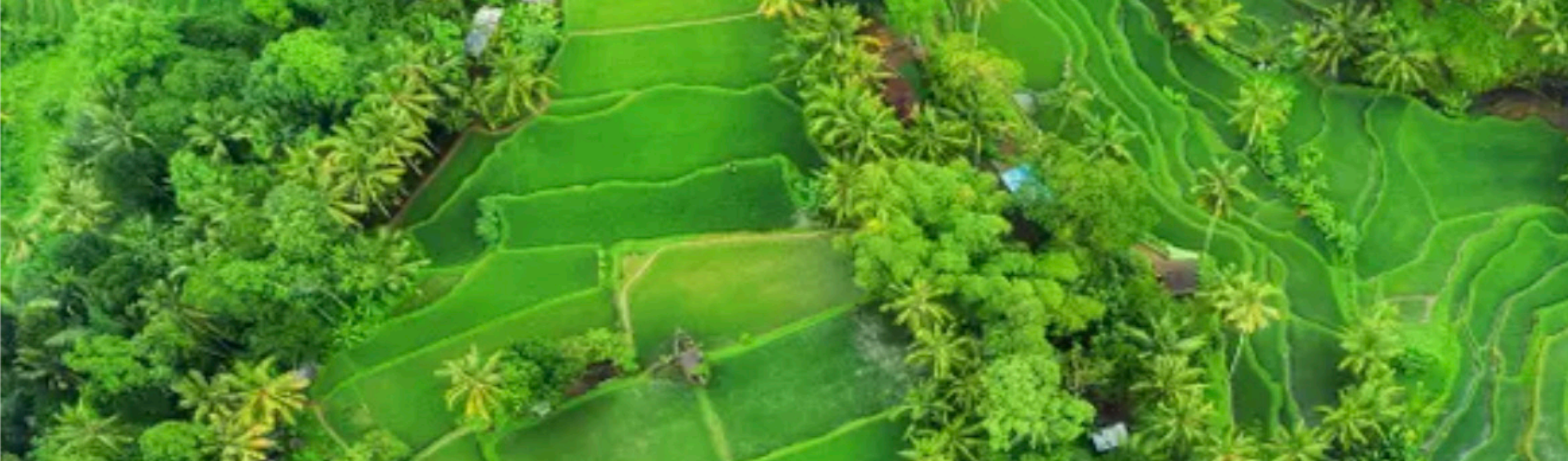
The RARE Feasibility Study is a joint project of:

Futurra is a U.S.-based 501c3 charitable organization committed to a resilient future where people and nature thrive together. Our work ensures that economic prosperity and environmental health reinforce, not compete with each other. We redesign supply systems to eliminate inefficiencies, maximize value, and build profitable, regenerative models. By aligning stakeholders—from investors and corporations to producers and consumers—we ensure that sustainability is an economic advantage, not a compromise.

www.Futurra.org

Impact Farming Foundation is a U.S.-based 501c3 charitable organization that works with smallholder farmers internationally to advance small-scale business solutions that integrate sustainable food production with increased tree cover and self-financed community development.

www.ImpactFarming.org



Sources of Input to the Feasibility Study

The following individuals were contacted and provided input for the RARE Feasibility Study:

- **Anelise Dias**, Adjunct Professor, Federal University of Rio de Janeiro, Brazil
- **Antonio Carlos de Souza Abboud**, Professor, Federal University of Rio de Janeiro, Brazil
- **Alvaro Acevedo Osorio**, National University of Colombia - Faculty of Agricultural Sciences, Colombia
- **Bita Corera**, Executive Director, Fundacion Para la Aplicación y Ensen (FUNDAEC), Colombia
- **Carlos Armenio Khatounian**, Professor, Luiz de Queiroz College of Agriculture, University of São Paulo, Brazil
- **Carlos German Muñoz**, National University of Colombia - Faculty of Agricultural Sciences, Colombia
- **Daniela Tiago da Silva Campos**, Professor, Federal University of Mato Grosso, Brazil
- **Gustavo Franco de Castro**, Adjunct Procesor, Federal University of Viçosa, Brazil
- **Jairo Leonardo Cuervo Andrade**, Professor, National University of Colombia, Colombia
- **Julio Cesar Alegre Orihuela**, Professor, Universidad Nacional Agraria La Molina, Peru
- **Laura Figueiras Marina**, First Assistant, University of Buenos Aires, Argentina
- **Maristela Watthier**, Adjunct Professor, Federal University of Viçosa, Brazil
- **Saray Siura Céspedes**, Professor, Universidad Nacional Agraria La Molina, Peru
- **Taís Brandão**, Masters Student, University of São Paulo, Brazil
- **Thomas Ludewigs**, Professor, Federal University of Para, Brazil
- **Timote Georges**, Executive Director, Smallholder Farmers Alliance, Haiti
- An early consultation with **Noble Scripts** provided valuable insights into large language models (LLMs) and their possibilities and limitations.
- **Ashish Pawar,** Business Development Manager with HipHype Tech, undertook web research.

Appendices

REGENERATIVE AGRICULTURE AND SMALLHOLDER FARMERS: A PRIMER

APPENDIX

A1

Compiled and edited by Chris Kaput

For smallholder farmers, the health of the land is the foundation of their livelihoods. Soil fertility, water availability, and biodiversity are essential for sustaining crops, livestock, and rural communities. However, climate change, soil degradation, and economic pressures have made farming increasingly difficult.

Regenerative agriculture offers a way forward—a system of farming that restores soil health, conserves water, improves biodiversity, and strengthens rural resilience. Unlike conventional approaches that deplete natural resources over time, regenerative agriculture works in harmony with nature, making farms more productive and resilient.

Importantly, many of the practices now labeled as “regenerative” are not new. They are rooted in Indigenous and traditional farming methods that have sustained communities for generations. Recognizing and building upon this knowledge is key to creating a sustainable and just agricultural future.

What is Regenerative Agriculture

Regenerative agriculture is a holistic, adaptable approach to farming that focuses on improving soil, water, and biodiversity while ensuring stable and productive yields. Some key principles include:

- **Soil Health:** Keeping soil covered with crops or mulch, minimizing tillage, and using compost or natural fertilizers to increase organic matter.
- **Biodiversity:** Growing a variety of crops, integrating trees and shrubs, and maintaining natural habitats for pollinators and beneficial insects.
- **Water Conservation:** Using techniques like rainwater harvesting, mulching, and agroforestry to improve water retention and reduce erosion.
- **Livestock Integration:** Practicing rotational grazing and using manure to enrich soil health.
- **Reducing External Inputs:** Using natural fertilizers and pest control methods to cut costs and maintain long-term soil fertility.

These principles help regenerate farm ecosystems—leading to more productive, drought-resistant, and profitable farming systems over time.

Why Regenerative Agriculture Matters for Smallholder Farmers

1. **Improved Soil Fertility and Higher Yields**

   Soil degradation is one of the biggest threats to smallholder farming. Healthy soil holds more nutrients, absorbs more water, and supports stronger plant growth. Practices like composting, crop rotation, and agroforestry help rebuild soil fertility, leading to better yields without relying on expensive synthetic fertilizers.

2. **Greater Resilience to Climate Change**

   Extreme weather events such as droughts, floods, and unpredictable rainfall are becoming more common. Regenerative practices, like maintaining ground cover and improving soil organic matter, help farms retain moisture during droughts and reduce erosion during heavy rains. This makes farming more stable and less dependent on costly external inputs.

3. **Reduced Dependence on Costly Inputs**

   Many smallholder farmers struggle with the rising costs of chemical fertilizers and pesticides. By improving soil health and using natural methods to manage pests and diseases, farmers can reduce their reliance on expensive external inputs, making their farms more self-sufficient and profitable.

4. **Strengthening Local Economies and Communities**

   Regenerative agriculture creates opportunities for local seed saving, composting initiatives, and farmer cooperatives, which all strengthen community resilience. It also helps ensure that farming remains a viable and dignified livelihood for future generations.

5. **Access to New Markets and Financial Support**

   Consumers and businesses are increasingly seeking products grown using sustainable and regenerative methods. Farmers who adopt regenerative practices may gain access to premium markets, certifications, and financial support from organizations investing in sustainable farming.

A Shift Towards Outcome-based Approaches

In the past, many agricultural programs focused on specific farming techniques—such as no-till or cover cropping—without measuring their actual impact. However, there is now a shift toward outcome-based approaches, where success is measured by improvements in soil health, biodiversity, and water retention, rather than by following a rigid set of prescribed practices.

For smallholder farmers, this means greater flexibility to adapt regenerative techniques to local conditions. Instead of being required to follow external rules, farmers can experiment and apply what works best for their land, climate, and resources—as long as it leads to measurable benefits.

The Role of Traditional Knowledge and Social Justice

Many Indigenous and smallholder farming communities have been practicing forms of regenerative agriculture for centuries. However, these contributions are often overlooked in modern discussions. Recognizing the value of traditional ecological knowledge is not only a matter of fairness, it is also practical, as these time-tested farming methods offer some of the best solutions for regenerating land today.

Additionally, ensuring that smallholder farmers have access to land, fair prices, financial support, and decision making power is just as important as implementing good agricultural practices. True regeneration must include social and economic justice alongside ecological restoration.

Financing and Support for Smallholder Farmers

Transitioning to regenerative agriculture takes time, and farmers often need financial and technical support to make the shift. Some potential support mechanisms include:

- **Farmer-to-Farmer Knowledge Sharing:** Learning from other farmers who have successfully implemented regenerative practices.
- **Microfinance and Low Interest Loans:** Access to small loans for composting, tree planting, or water management projects.
- **Community Led Initiatives:** Cooperatives and local partnerships that provide training, tools, and resources.
- **Outcome Based Payments and Incentives:** Financial rewards for farmers who improve soil health, water retention, and biodiversity.
- **Direct Market Access:** Partnerships with buyers who value regenerative practices and are willing to pay fair prices.

Investing in regenerative agriculture is not just an expense, it is an investment in long-term farm productivity, community resilience, and economic security.

What About Carbon Credits?

Many companies and organizations are interested in carbon credits, which pay farmers to sequester carbon in soils. While this can be an opportunity, smallholder farmers should approach carbon markets with caution. The science of soil carbon storage is still evolving, and some carbon credit schemes have complex requirements that may not always benefit farmers fairly.

Rather than focusing only on carbon markets, smallholder farmers should prioritize regenerative practices for their many other benefits, such as soil fertility, water conservation, and improved yields. If engaging in carbon programs, farmers should ensure they understand the terms, receive fair compensation, and retain control over their land and farming decisions.

Conclusion

Regenerative agriculture offers smallholder farmers a practical and proven path to healthier soils, stronger crops, and greater economic stability. By working with nature rather than against it, farmers can increase productivity, reduce costs, and build resilience to climate change.

However, regenerative agriculture is not just about farming techniques. It is about empowering farmers, recognizing traditional knowledge, and ensuring fair access to resources and markets. For smallholder farmers, the future of agriculture does not have to mean struggling with depleted soil, rising costs, and unpredictable weather. By adopting regenerative practices, farmers can restore their land, strengthen their communities, and build a more secure and thriving future.

VOLUME, LOCATION AND ACCESSIBILITY OF CURRENT RESEARCH

APPENDIX

A2

Contents

1. Summary

The RARE Feasibility Study looked into the accessibility of research on smallholder regenerative agriculture through major scientific library platforms and AI-powered search engines. It found that despite over 110 million documents in key digital libraries, only 0.002% specifically addressed this topic. The libraries analyzed lacked translation features and the ability to do full-text searching within the documents in their collections. The two main AI-powered search engines studied did not store original research documents, but had further reach to find documents on other sites. However, these two examples, plus search engines in general, have limitations in translating technical and culturally specific terms in the documents they located. Additionally, an estimated 30,000 relevant research documents exist in South America (the focus of this study) but remain inaccessible to search engines due to their presence on restricted or local institutional repositories. These findings highlight a significant gap in the availability and accessibility of research crucial for regenerative agriculture development by smallholder farmers.

2. Introduction

When setting out to determine the need for the Regenerative Agriculture Research Enterprise (RARE) as part of this Feasibility Study, a key question was whether there is a significant body of smallholder regenerative agriculture research that is already online and easily accessible. We set out to answer this question by looking at two existing, but very different, approaches:

- **Scientific library platforms** that function as online repositories of documents that are housed on their site.

- **AI-powered scientific search engines** that conduct literature searches, but do not have documents on their sites.

Key Scientific Library Platforms

We conducted a comprehensive analysis of eight of the largest scientific digital library platforms, examining their document collections, search capabilities, language support, and technical features. They were selected based on the size of their collection and whether that collection included agricultural research as subject area. Two of the platforms had additional relevance because they focus specifically on South America. Here are the platforms that were analyzed:

- International System for Agricultural Science and Technology (AGRIS)
- JSTOR
- MDPI
- National Center for Biotechnology Information (NCBI)
- Redalyc Scientific Information System
- Science Direct
- Scientific Electronic Library Online (SciELO)
- Springer Nature Link

Secondary Scientific Library Platforms

In order to be comprehensive in our investigation, we looked closely at eight smaller platforms in order to have a complete overview of available online research, and in particular as it pertains to South America. Our review of these platforms confirmed they did not add significantly to the RARE Feasibility Study, but we have listed them here for reference:

- Acesso Livre à Informação Cientifica da Embrapa (ALICE)
- Center for International Cooperation in Agricultural Research for Development (CIRAD)
- Consortium of University Libraries of Ecuador (Cabuec/Bibec)
- INTA Digital - Instituto Nacional de Tecnología Agropecuaria Institutional Repository

- Latin American Open Science Network (La Referencia)
- Research Gate
- Sistema Regional de Información en línea para Revistas Científicas de América Latina, el Caribe, España y Portugal (Latin Index)
- UNAM Institutional Repository

AI-powered Scientific Search Engines

We examined the two main AI-powered search engines that specialize in literature searches of scientific and academic research papers. They focus on search and retrieval of these documents rather than serving as a traditional library. We reviewed their capabilities to determine if they are already providing the services that RARE has identified, or if there are gaps that RARE is uniquely suited to addressing:

- Consensus
- SCISPACE

3. Analysis of Key Scientific Library Platforms

Step one was to determine the total number of documents available across all eight platforms. The answer: over 110 million (see Table 1, Column 2 below), noting that the documents are in multiple languages and duplicates were removed when platforms showed the same document in multiple languages.

We then did a search to find out how many documents were in the general category of agriculture in order to have a baseline for comparison. Given that search terms in English do not automatically find documents in other languages, we ran our search using “agriculture” and its translations below. The results are shown in Table 1, Column 3:

- English: "agriculture"
- Spanish: "agricultura"
- French: "agriculture"
- Portuguese: "agricultura"
- German: "Landwirtschaft"
- Chinese: "农业"

Next, we wanted to determine how many of these documents focused on regenerative agriculture. The results are shown in Table 1, Column 4:

- English: "regenerative agriculture"
- Spanish: "agricultura regenerativa"
- French: "agriculture régénérative"
- Portuguese: "agricultura regenerativa"
- German: "regenerative Landwirtschaft"
- Chinese: "再生农业"

Finally we broke it down to regenerative agriculture and smallholder farmer keywords. The results are shown in Table 1, Column 5:

- English: "regenerative agriculture" + "smallholder farmer"
- Spanish: "agricultura regenerativa" + "pequeño agricultor"
- French: "agriculture régénérative" + "petit exploitant"
- Portuguese: "agricultura regenerativa" + "pequeno agricultor"
- German: "regenerative Landwirtschaft" + "Kleinbauer"
- Chinese: "再生农业" + "小农"

Table 1 – Statistics for Key Platforms				
1	**2**	**3**	**4**	**5**
Platform	**Documents in all languages**	**Search word "agriculture"**	**Search words "regenerative agriculture"**	**Search words "regenerative agriculture" + smallholder farmer"**
AGRIS	13,845,762	3,284,567	4,256	623
JSTOR	14,782,932	1,987,654	3,892	528
MDPI	742,856	287,654	892	156
NCBI	41,267,843	1,876,543	967	98
Redalyc	942,224	198,765	289	76
Science Direct	21,984,567	2,987,654	5,234	682
SciELO	1,042,873	246,892	342	89
Springer Nature	15,842,967	2,456,789	3,245	156
Total documents:	110,452,024	13,326,518	19,117	2,677
Percent of total documents:		12.1%	0.017%	0.002%

As the results above clearly indicate, documents that address "regenerative agriculture," and "smallholder regenerative agriculture," are extremely low compared to those focused on "agriculture" alone. In fact they are practically a statistical zero in comparison to the total number of documents on these platforms, despite the fact that all sites include agriculture as a subject area.

Having determined these statistics, we explored other features of the various platforms and found that:

- none of the platforms offered an on-site function to translate documents found on their respective sites into other languages; and

- none of the platforms can search the contents of individual documents.

4. Analysis of AI-powered Scientific Search Engines

There are a number of AI-powered search engines at present, but we chose the two that have the broadest search capacity and relevant technical features.

Consensus

Consensus deploys AI, including natural language processing (NLP) models (computer programs designed to understand and work with human language), in order to generate answers to specific questions by extracting and summarizing insights from scientific research papers. What we learned is that Consensus:

- does not require users to upload documents; it focuses on sourcing from publicly available research databases;
- is good for generating broad, evidence-based insights across multiple studies;
- supports queries in over 100 languages and retrieves academic insights regardless of the original language of the study; and
- is optimized for summarizing content but does not allow deep interaction with specific documents.

SciSpace

SciSpace deploys AI, including natural language processing (NLP) models, in order to help researchers and academics easily create, manage, and collaborate on scientific documents. It offers tools for formatting, translating, summarizing, and analyzing research papers to streamline the academic workflow.

What we learned is that SciSpace:

- includes an AI assistant for uploaded PDFs or specific research papers, enabling users to highlight sections and ask detailed questions.
- supports interactions in over 75 languages but focuses more on enabling users to read, summarize, and interact with individual papers in their chosen language.
- is ideal for users needing a deeper understanding of one paper rather than broad analytic summaries across many papers.

Both Consensus and SciSpace were found to have similar limitations regarding translation:

- both have limited accuracy when translating technical and culturally-specific terms and phrases that are not part of standardized universal scientific terminology, noting that they use NLP technology that is similar to that used by GPT and Claude (see appendix A4), and
- both are limited to translating the text of documents, and are unable to maintain photos, graphs and other digital images that are often key to the complete explanation and understanding of the text in research papers. This limitation is the same for GPT and Claude (see appendix A4).

5. Locating and Estimating the Volume of Missing Research

Given the findings above, we wanted to determine if there were more research papers available than those found via the digital library platforms and search engines that we analyzed.

To answer this question for the South American region, we asked five universities across Brazil, Peru and Colombia, plus one government agency in Brazil, to roughly estimate the number of relevant documents on the restricted servers of their institutions that would potentially be available for translation and sharing. The request was specifically for documents related to "smallholder sustainable and/or regenerative agriculture" and that were published after 2014 (i.e. the last 10 years).

While this is by no means a thorough analysis of these specific sources, the table below does represent an overview for the purposes of this RARE Feasibility Study. The rough estimates of relevant research papers, technical guides, reports, theses, and studies at each institution is as follows:

Institution	Number of Documents	Country	Website
EMBRAPA – Alice repository	1,169	Brazil	https://www.alice.cnptia.embrapa.br/
University of Sao Paulo – Digital Library for theses and dissertations	5,273*	Brazil	https://www.teses.usp.br/
University of Campinas – Institutional Repository	250**	Brazil	https://repositorio.unicamp.br/
Federal University of Viçosa – theses and dissertations, technical repository	2,410	Brazil	https://locus.ufv.br/communities
Institutional Repository - La Molina National Agrarian University	1,033	Peru	https://repositorio.lamolina.edu.pe/
National University of Colombia – Institutional Repository	1,050	Colombia	https://repositorio.unal.edu.co/
	11,185		

* Not possible to filter the year and other features
** Keyword used: family farming

This non-scientific and incomplete search identified 11,185 documents that would appear relevant to RARE and which are potentially available for translation and sharing.

From this figure we extrapolated a rough estimated total volume of relevant materials located in South America. Given that the largest sample portion is from Brazil, and taking into account that Brazil represents roughly 49% of South America's total population, we estimate that there are a minimum of 30,000 documents available in the region for translation and dissemination by RARE.

6. Conclusions

Our analysis found that the number of research papers on smallholder regenerative agriculture in the eight of key online scientific library platforms that include agriculture as a subject area is extremely low, amounting to 0.002% of a combined total of more than 110 million documents in all subject categories.

While these library platforms house documents in various languages, they are not able to translate these documents and they do not have the capacity to conduct searches of the contents of any documents.

By contrast the two AI-powered scientific search engines analyzed have no documents housed on their sites. Instead, they both specialize in answering user questions by retrieving information from scientific research papers found on other platforms.

While it is not possible to generate the same statistics as with library platforms, it is fair to say that both search engines already include the scientific libraries referred to above in their searches. Of course, the search engines cast a broader net than just these libraries, but even if the total volume of available documents is three or four times higher, the percentage of documents specific to smallholder regenerative agriculture is unlikely to exceed 0.01 % of the total volume.

The two search engines reviewed allow users to pose specific questions that are answered because the engines have the capacity to search the contents of documents in order to generate answers. And while both sites allow for text translations of documents into other languages, neither is able to achieve the highest standards of accuracy when translating technical and culturally-specific terms and phrases, nor are they able to maintain digital images. In this regard, both search engines have the same restrictions in translating documents that was encountered with GPT and Claude (see appendix A4).

Based on initial research, we estimate there are at least 30,000 relevant research documents that currently exist in South America but which are not on scientific library platforms and which are on local or restricted websites that make them inaccessible to search engines.

ORIGINAL RESEARCH DOCUMENTS + PROFESSIONAL AND AI TRANSLATIONS

APPENDIX

A3

Contents

COMPARISON OF AI vs PROFESSIONAL TRANSLATIONS

APPENDIX

A4

Contents

1. Summary

We undertook an exhaustive comparative analysis of AI-powered and professional translations of agricultural research papers and a training manual, focusing on accuracy, terminology consistency, and the preservation of visuals and layout. Four AI models—GPT-4, Claude 3.5, DeepL Pro, and Google Translate—were evaluated against professional translations, with an emphasis on technical agricultural terms and country-specific phrases.

Results showed that AI models varied in performance when compared to professional translations. LLMs (GPT-4 and Claude 3.5) performed well with general text but struggled with technical precision. SLMs (DeepL Pro and Google Translate) performed slightly behind the LLMs, but preserved document formatting and visuals.

Given the role of glossaries to increase the accuracy of text translations (see appendix A5), the recommendation is to use SLMs as the primary AI translation tool for the RARE platform.

2. Introduction

Note: a complete set of all original and translated documents is available in appendix A3.

One of the questions we set out to answer in the course of this Feasibility Study is the degree to which currently available AI-powered translation models are able to accurately translate research papers.

There are two primary considerations when it comes to AI and translation. The most important is translating the basic text, which represents the bulk of our investigation. The second has to do with how photos, graphs, illustrations, layout and format are maintained in the process of translation. We address the latter in section 9 following.

For translating text there is a further differentiation between the accuracy of translating technical and cultural words or phrases and the related issues of semantics, fluency and sentence structure. Our study focused on the former, while the set of related issues are addressed in a detailed discussion of glossaries in appendix A5. In summary, the glossary addresses semantics, fluency and sentence structure by including a range of conditions for the placement of translated words and phrases in specific contexts.

To answer the question we set for ourselves about AI translation quality, we selected four of the most widely used AI translation models for a comparison test:

1. GPT-4 (the underlying language model behind ChatGPT)

2. Claude 3.5

3. DeepL Pro

4. Google Translate

The first two (GPT-4 and Claude 3.5) are large language models (LLMs) that learn by drawing on text collections, known as datasets or corpora, that include vast internet data. LLMs excel at general language tasks but struggle with the nuances of technical writing, including jargon, complex sentence structures, and contextual understanding.

The other two (DeepL Pro and Google Translate) are small language models (SLMs) that are essentially smaller and purpose-built versions of their LLM counterparts. While they learn from a less robust corpora and lack the breadth of LLMs, they are easier to fine tune for specific tasks such as translating. They also use significantly less computational power than LLMs.

We started with four research papers in PDF format (see appendix A3 for the originals) on smallholder sustainable agriculture that dealt with liquid biofertilizers, the effect of soybeans on soil microbiology, and two papers on green manure; they were all provided by the Brazilian Agricultural Research Corporation (EMBRAPA). These papers were only available in Portuguese and we first had them translated into English by a professional translator with a background in smallholder agriculture and who is a native Portuguese speaker. He is also qualified to translate into English. We then used the four AI models to translate the same documents into English, and what follows is a detailed comparison of the text translations. The comparison of the translations regarding the maintenance of visuals and layout is addressed separately in section 9.

Prior to translation we added highlights to the documents: green indicates a technical agricultural term or phrase and blue indicates a country-specific word or phrase (see an example in section 8). When the papers were translated the green and blue highlights were maintained, but errors were highlighted in pink.

Having determined that the two SLMs outperformed the LLMs by a considerable degree in maintaining visuals and layout in their translations, we conducted a comparison test involving only DeepL and Google Translate by using them to translate a French language regenerative cotton training manual from Haiti into English and comparing the results to a professional translation of the same document. The results are presented in section 7.

3. Translation Analysis for Document #1

Original document: Appendix A3

Title: *Acúmulo de fitomassa e nutrientes e estádio mais adequado de manejo do feijão-de-porco para fins de adubação verde*

Authors: Padovan, M. P., Motta, I. D. S., Carneiro, L. F., Moitinho, M. R., & Fernandes, S. S. L.

Year of publication: 2011

Source: Revista Brasileira de Agroecologia, 6(3),182-190

Document type: Paper (PDF format)

Note: Bibliography not included in the comparison

Number of pages: 7

Total word count: 4,343

Table 1: **Total Terms and Expressions Reviewed for Document #1**

Type of translation	Correct vs total technical terms & expressions reviewed		Correct vs total country-specific terms & expressions reviewed	
	number	percent correct	number	percent correct
Professional	67/67	100	13/13	100
GPT-4	54/67	81	12/13	92
Claude 3.5	63/67	94	12/13	92
DeepL Pro	64/67	96	9/13	69
Google Translate	61/67	91	10/13	77

In Document #1, as shown in Table 1, 67 words/technical expressions and 13 country-specific words/expressions were identified, for a total of 80 words and expressions to be analyzed.

Incorrect translations, drawn from all four models, included the following:

- “Cerrado do Brasil Central” was literally translated, compromising the semantics and the cultural consistency.
- “Fitomassa” in Portuguese was translated as “biomass” and “high biomass production” instead of “phytomass”, which would be the correct English word.
- “Pragas” in Portuguese should be “plagues” in English, but instead showed up as “insects.”
- “soil” was used instead of “edaphic,” noting that edaphic conditions take into account the properties (fertility, structure) of the soil regarding plant production.
- "sistema radicular ramificado" was translated as “branched roots,” instead of "branched root system"
- "culturas perenes" was translated to “crops perennials” instead of “perennial crops.”

4. Translation Analysis for Document #2

Original document: Appendix A3

Title: *Decomposição da biomassa de adubos verdes no sudoeste de Mato Grosso e sua estimativa pelo modelo NDICEA*

Authors: Nascimento.; Mattos, L. S. Jorge; Mendonça, Eduardo S.

Year of publication: 2016

Source: Revista Brasileira de Agroecologia, 11 (4), 319-327

Document type: Paper (PDF format)

Note: Bibliography not included in the comparison

Number of pages: 5

Total word count: 5,119

Table 2: **Total Terms and Expressions Reviewed for Document #2**				
Type of translation	**Correct vs total technical terms & expressions reviewed**		**Correct vs total country-specific terms & expressions reviewed**	
	number	**percent correct**	**number**	**percent correct**
Professional	62/62	100	11/11	100
GPT-4	58/62	94	10/11	91
Claude 3.5	47/62	76	10/11	91
DeepL Pro	55/62	89	7/11	64
Google Translate	46/62	74	5/11	45

In Document #2, as shown in Table 2, 62 words/technical expressions and 11 country-specific words/expressions were identified, for a total of 73 words and expressions to be analyzed.

Incorrect translations, drawn from all four models, included the following:

- "guandu anão" in Portuguese was translated as "pigeon pea" in English. It should have been "dwarf pigeon pea."
- "consórcio" was translated as "mixtures" instead of "intercropping."
- "biomassa vegetal" was translated as "biomass", instead of "plant biomass" or "vegetal biomass."
- "family labor-run practices" is an English translation that damages the fluency of the text and could have a better grammatical construction.
- the expression "boas práticas de conservação e ciclagem de nutrientes do solo" encompasses two technical expressions ("boas práticas" and "conservação e ciclagem de nutrientes do solo") in Portuguese. The English translation of this phrase was "good soil conservation and nutrient cycling practices", but the original sense would be better preserved with another construction such as "good practices for the conservation and cycling of soil nutrients."
- "organic waste" was used instead of "organic residue."
- "soil classification" was not translated to English, remaining in Portuguese.

5. Translation Analysis for Document #3

Original document: Appendix A3

Title: *Efeito de sistemas diversificados de produção de soja em atributos microbiológicos do solo*

Authors: Fior, B. B.; Charnobay, A. C.; Balbinot Junior, A. A.; Hungria, M.; Nogueira, M. A.

Year of publication: 2023

Source: Jornada Acadêmica da Embrapa Soja, 59-64.

Document type: Paper (PDF format)

Note: Bibliography not included in the comparison

Number of pages: 5

Total word count: 2,141

Table 3: **Total Terms and Expressions Reviewed for Document #3**				
Type of translation	**Correct vs total technical terms & expressions reviewed**		**Correct vs total country-specific terms & expressions reviewed**	
	number	**percent correct**	**number**	**percent correct**
Professional	31/31	100	4/4	100
GPT-4	31/31	100	4/4	100
Claude 3.5	31/31	100	4/4	100
DeepL Pro	31/31	100	3/4	75
Google Translate	31/31	100	3/4	75

In Document #3, as shown in Table 3, 31 words/technical expressions and 4 country-specific words/expressions were identified, for a total of 35 words and expressions to be analyzed. This document stands out because it had a 100 percent score, with two exceptions. One explanation is that the document is relatively short and has the lowest number of technical terms and country-specific terms of all the documents analyzed.

6. Translation Analysis for Document #4

Original document: Appendix A3

Title: *Preparo de biofertilizantes líquidos, Embrapa Semi-Árido*

Authors: Silva, A. F.,Pinto, J. M., França, C. R. R. S., Fernandes, S. C., Gomes, T. C. A., Silva, M. S. L., Matos. A. N. B.

Year of publication: 2007

Source: Preparo de biofertilizantes líquidos, Embrapa Semi-Árido. Comunicado técnico.

Document type: Technical guide (PDF format)

Note: Bibliography not included in the comparison

Number of pages: 4

Total word count: 2,510

Table 4: **Total Terms and Expressions Reviewed for Document #4**				
Type of translation	**Correct vs total technical terms & expressions reviewed**		**Correct vs total country-specific terms & expressions reviewed**	
	number	**percent correct**	**number**	**percent correct**
Professional	89/89	100	10/10	100
GPT-4	83/89	93	7/10	70
Claude 3.5	87/89	98	6/10	60
DeepL Pro	86/89	97	7/10	70
Google Translate	80/89	90	8/10	80

In Document #4, as shown in Table 4, 89 words/technical expressions and 10 country-specific words/expressions were identified, for a total of 99 words and expressions to be analyzed.

Incorrect translations, drawn from all four models, included the following:

- “Plantações”, in Portuguese, was translated to “yields”, which refers to crop productivity. In this case “crop” would be more accurate.
- “Massa plástica” was literally translated as “plastic mass”, which should be “plastic putty.”
- for “copa da planta”, in Portuguese, the translation was “plant canopy”, but this means "dossel", the upper layer of a forest formed by treetops. A more accurate translation would be “top of the plant”, because the fertilizer is to be applied to each plant.
- proper names are generally not translated, but “Embrapa Semi-Árido”, in Portuguese, was translated to “Embrapa Semi-Arid.”
- similarly “Fermentado biológico”, in Portuguese, was translated to “Biological Ferment” (in this case at the beginning of the text).
- “property” is better translated to “farm.”
- “carboy” (“bombona”, in Portuguese), would be correctly translated to “drum”, considering that a carboy is a plastic bottle for corrosive liquids and plastic drums are for non-corrosive liquid.

7. Translation Comparison of Document #1 to #4

Table 5: **Total Terms and Expressions Reviewed – Documents 1 to 4**					
	Professional	**GPT-4**	**Claude 3.5**	**DeepL Pro**	**Google Translate**
% = Correct vs total technical terms & expressions					
Doc 1	100	81	94	96	91
Doc 2	100	94	76	89	74
Doc 3	100	100	100	100	100
Doc 4	100	93	98	97	90
% = Correct vs total country-specific terms & expressions					
Doc 1	100	92	92	69	71
Doc 2	100	91	91	64	45
Doc 3	100	100	100	75	75
Doc 4	100	70	60	70	80
Average >	**100**	**90.13**	**88.9**	**82.5**	**79**

Note: Document #5 is not included in this table because the process for comparison was different. See the next section 7 for details.

8. Translation Analysis for Document #5

Original document: Appendix A3

Document type: Training Manual (PDF format)

Title: *Coton Régénérative: Manuel d'Estension*

Authors: Terra Genesis and Smallholder Farmers Alliance

Year of publication: 2022

Source: Smallholder Farmers Alliance

Document type: Technical guide (PDF format)

Number of pages: 55

Total word count: 4,325

Image count: 37 unique images (not counting repeated use of logos and icons)

For the purpose of this study we did an analysis of a fifth document, but this time using GPT-4 to conduct the test and using only two translations that were done using only the SLMs: DeepL Pro and Google Translate. This decision was made because of the superior capacity of SLMs over LLMs to maintain visuals and layout, and in this case the 55 page document had 37 images.

Comparison A analyzes the DeepL Pro translation of the original French document into English, using a professional translation as the standard for evaluation.

Comparison B analyzes the Google Translate translation for a similar evaluation process.

Comparison A: DeepL Pro vs. Professional Translation

DeepL's translation from the French original into English is generally well-structured and grammatically accurate, but there are several notable discrepancies when compared to the professional translation. Below are key observations:

1. **Terminology Accuracy:**
   - DeepL often translates "agriculture régénérative" as "agriculture régénératrice" instead of "regenerative agriculture", which is an incorrect choice of words in English.
   - "Petit exploitant" is translated inconsistently. In some cases, it translates correctly as "smallholder", while in others, it is inaccurately translated as "small farmer", which slightly alters the intended meaning.
   - "Résultats régénérateurs" is translated as "regenerative results" instead of "regenerative outcomes", which is less precise.

2. **Sentence Structure & Readability:**
   - DeepL produces direct, often literal translations, making the text less fluid than the professional translation.
   - Some sentences maintain French syntactic structures, making them sound unnatural in English.
   - Example:
     - DeepL: "Healthy soil results in stronger plants that don’t get sick as easily."
     - Professional Translation: "Healthy soil produces stronger plants that are more resistant to disease."
     - The professional translation is smoother and more precise.

3. **Accuracy Score for DeepL Translation:**
   - Approximate Accuracy: 87%
   - Main Issues:
     - Word choice & terminology inconsistencies
     - Literal translations leading to unnatural phrasing
     - Occasional formatting and structural errors

Comparison B: Google Translate vs. Professional Translation

Google Translate shows a clear drop in accuracy compared to DeepL. Here are the key issues:

1. **Terminology & Word Choice Issues:**
    - Google Translate frequently mistranslates technical agricultural terms.
    - Example:
        - French Original: "La rotation des cultures aide à améliorer le sol et à prévenir les ravageurs et les maladies."
        - Google Translate: "Crop rotation helps improve the soil and helps prevent pests and diseases."
    - While not entirely incorrect, "helps improve" is a weaker phrase than the professional translation: "Crop rotation enhances soil health and prevents pests and diseases."
    - "Culture intercalaire" is inconsistently translated as "intercrop", "crop alternation", or even "mixed crops", affecting clarity.

2. **Grammatical & Structural Errors:**
    - Google Translate often fails to adapt French sentence structures to English, leading to awkward phrasing.
    - Example:
        - Google Translate: "The presence of trees creates a more comfortable working environment (more shade, cooler microclimates)."
        - Professional Translation: "Trees provide shade and create a cooler microclimate, making the work environment more comfortable."
    - The professional version is more natural, while Google Translate tends to overuse direct, robotic structures.

3. **Literal Translations & Awkward Phrasing:**
    - Google Translate struggles with idioms and contextual meaning.
    - Example:
        - French Original: "Nous sommes sur le point de perdre suffisamment d'eau pour soutenir les sociétés humaines."
        - Google Translate: "We are on the verge of losing enough water to support human societies."
        - Professional: "We are at risk of losing the water needed to sustain human societies."
        - The professional translation better captures the intended meaning.

4. **Omissions & Misinterpretations:**
    - Google Translate omits or misinterprets some technical phrases.
    - Example:
        - French Original: "Les agriculteurs recevront une prime lorsqu'ils démontreront qu'ils se conforment aux pratiques requises."
        - Google Translate: "Farmers will receive a bonus if they show they comply with the required practices."
        - Professional: "Farmers will receive a premium price when they demonstrate compliance with required practices."

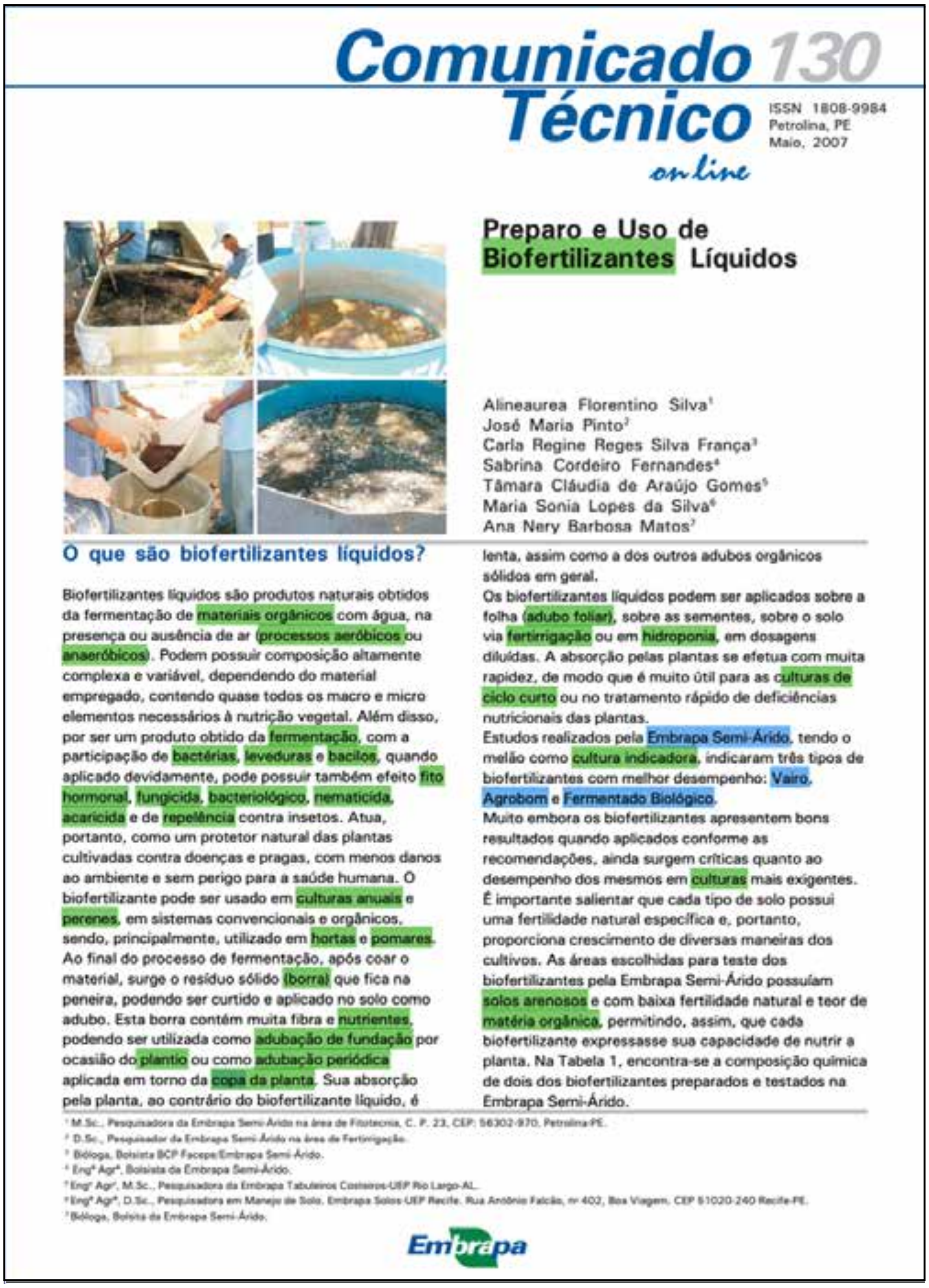
- The difference between "bonus" and "premium price" affects the meaning, since "premium price" refers to higher earnings, while "bonus" implies a separate reward.

Accuracy Score for Google Translate:

- Approximate Accuracy: 78%
- Main Issues:
  - Frequent incorrect terminology choices
  - Literal translations leading to unnatural English
  - More structural and grammatical errors than DeepL
  - Occasional loss of meaning due to misinterpretation

9. Translation Example

Following are images of one page of the actual translations for Document #4: *Preparo de biofertilizantes líquidos, Embrapa Semi-Árido.* (section 5 above). Prior to translation we added highlights to the text: green indicates a technical agricultural term or expression, while blue indicates country-specific words or expression. Pink indicates an error in the translation.



Comunicado 130
Técnico
on line

ISSN 1808-9984
Petrolina, PE
Maio, 2007

Preparo e Uso de Biofertilizantes Líquidos

Alineaurea Florentino Silva[1]
José Maria Pinto[2]
Carla Regine Reges Silva França[3]
Sabrina Cordeiro Fernandes[4]
Tâmara Cláudia de Araújo Gomes[5]
Maria Sonia Lopes da Silva[6]
Ana Nery Barbosa Matos[7]

O que são biofertilizantes líquidos?

Biofertilizantes líquidos são produtos naturais obtidos da fermentação de materiais orgânicos com água, na presença ou ausência de ar (processos aeróbicos ou anaeróbicos). Podem possuir composição altamente complexa e variável, dependendo do material empregado, contendo quase todos os macro e micro elementos necessários à nutrição vegetal. Além disso, por ser um produto obtido da fermentação, com a participação de bactérias, leveduras e bacilos, quando aplicado devidamente, pode possuir também efeito fito hormonal, fungicida, bacteriológico, nematicida, acaricida e de repelência contra insetos. Atua, portanto, como um protetor natural das plantas cultivadas contra doenças e pragas, com menos danos ao ambiente e sem perigo para a saúde humana. O biofertilizante pode ser usado em culturas anuais e perenes, em sistemas convencionais e orgânicos, sendo, principalmente, utilizado em hortas e pomares. Ao final do processo de fermentação, após coar o material, surge o resíduo sólido (borra) que fica na peneira, podendo ser curtido e aplicado no solo como adubo. Esta borra contém muita fibra e nutrientes, podendo ser utilizada como adubação de fundação por ocasião do plantio ou como adubação periódica aplicada em torno da copa da planta. Sua absorção pela planta, ao contrário do biofertilizante líquido, é lenta, assim como a dos outros adubos orgânicos sólidos em geral.

Os biofertilizantes líquidos podem ser aplicados sobre a folha (adubo foliar), sobre as sementes, sobre o solo via fertirrigação ou em hidroponia, em dosagens diluídas. A absorção pelas plantas se efetua com muita rapidez, de modo que é muito útil para as culturas de ciclo curto ou no tratamento rápido de deficiências nutricionais das plantas.

Estudos realizados pela Embrapa Semi-Árido, tendo o melão como cultura indicadora, indicaram três tipos de biofertilizantes com melhor desempenho: Vairo, Agrobom e Fermentado Biológico.

Muito embora os biofertilizantes apresentem bons resultados quando aplicados conforme as recomendações, ainda surgem críticas quanto ao desempenho dos mesmos em culturas mais exigentes. É importante salientar que cada tipo de solo possui uma fertilidade natural específica e, portanto, proporciona crescimento de diversas maneiras dos cultivos. As áreas escolhidas para teste dos biofertilizantes pela Embrapa Semi-Árido possuíam solos arenosos e com baixa fertilidade natural e teor de matéria orgânica, permitindo, assim, que cada biofertilizante expressasse sua capacidade de nutrir a planta. Na Tabela 1, encontra-se a composição química de dois dos biofertilizantes preparados e testados na Embrapa Semi-Árido.

[1] M.Sc., Pesquisadora da Embrapa Semi-Árido na área de Fitotecnia, C. P. 23, CEP: 56302-970, Petrolina-PE.
[2] D.Sc., Pesquisador da Embrapa Semi-Árido na área de Fertirrigação.
[3] Bióloga, Bolsista BCP Facepe/Embrapa Semi-Árido.
[4] Engª Agrª, Bolsista da Embrapa Semi-Árido.
[5] Engª Agrª, M.Sc., Pesquisadora da Embrapa Tabuleiros Costeiros-UEP Rio Largo-AL.
[6] Engª Agrª, D.Sc., Pesquisadora em Manejo de Solo, Embrapa Solos-UEP Recife, Rua Antônio Falcão, nº 402, Boa Viagem, CEP 51020-240 Recife-PE.
[7] Bióloga, Bolsista da Embrapa Semi-Árido.

Embrapa

Document #4: Original text in Portuguese.

Technical Communiqué 130

ISSN 1808-9984
Petrolina, PE
Maio, 2007

on line

Preparation and Use of Liquid Biofertilizers

Alineaurea Florentino Silva[1]
José Maria Pinto[2]
Carla Regine Reges Silva França[3]
Sabrina Cordeiro Fernandes[4]
Tâmara Cláudia de Araújo Gomes[5]
Maria Sonia Lopes da Silva[6]
Ana Nery Barbosa Matos[7]

What are liquid biofertilizers?

Liquid biofertilizers are natural products obtained from the fermentation of organic materials with water, in the presence or absence of air (aerobic or anaerobic processes). They can have a highly complex and variable composition, depending on the material used, containing almost all the macro and micro elements necessary for plant nutrition. Furthermore, as it is a product obtained from fermentation, with the participation of bacteria, yeasts, and bacilli, when properly applied, it can also have phytohormonal, fungicidal, bacteriological, nematicidal, acaricidal and insect repellent effects. It therefore acts as a natural protector of cultivated plants against diseases and pests, with less damage to the environment and no danger to human health. The biofertilizer can be used on annual and perennial crops, in conventional and organic systems, and is mainly used in vegetable gardens and orchards. At the end of the fermentation process, after sieving the material, there is a solid residue (sludge) which remains in the sieve and can be cured and applied to the soil as fertilizer. This sludge contains a lot of fiber and nutrients and can be used as base dressing when planting or for recurring fertilization around the top of the plant. Unlike liquid biofertilizer, it is absorbed slowly by the plant, just like other solid organic fertilizers in general.

Liquid biofertilizers can be applied to the leaf (foliar fertilizer), to seeds, to the soil via fertirrigation or in hydroponics, in diluted dosages. It is absorbed by plants very quickly, so it is very useful for short-cycle crops or for the rapid treatment of plant nutritional deficiencies. Studies carried out by Embrapa Semi-Arido, with melon as the indicator crop, indicated three types of biofertilizers with the best performance: *Vairo*, *Agrobom* and *Fermentado Biológico*.

Although biofertilizers show good results when applied according to the recommendations, there are still criticisms about their performance on more demanding crops. It is important to point out that each type of soil has a specific natural fertility and therefore provides growth in different ways for crops. The areas chosen for testing the biofertilizers by Embrapa Semi-Arido had sandy soils with low natural fertility and organic matter content, thus allowing each biofertilizer to express its ability to nourish the plant. Table 1 shows the chemical composition of two of the biofertilizers prepared and tested at Embrapa Semi-Arido.

[1] M.Sc., Embrapa Researcher Semi-Árido in the field of Fitotecnia, C. P. 23, CEP: 56302-970, Petrolina-PE.
[2] D.Sc., Embrapa Researcher Semi-Árido in the field of Fertirrigation.
[3] Biologist, Fellow, BCP Facepe/Embrapa Semi-Árido.
[4] Agr. Eng., Embrapa Semi-Árido Fellow.
[5] Agr. Eng., M.Sc., Embrapa Researcher Tabuleiros Costeiros-UEP Rio Largo-AL.
[6] Agr. Eng., D.Sc., Soil Management Researcher. Embrapa Solos-UEP Recife. Rua Antônio Falcão, n° 402, Boa Viagem. CEP 51020-240 Recife-PE.
[7] Bióloga, Embrapa Semi-Árido Fellow.

Document #4: Professional translation into English.

2007 - EMBRAPA - Biofertilizantes PT

Preparation and Use of Liquid Biofertilizers

What are liquid biofertilizers?

Liquid biofertilizers are natural products obtained from the fermentation of organic materials with water, in the presence or absence of air (aerobic or anaerobic processes). They can have a highly complex and variable composition, depending on the materials used, containing almost all the macro and microelements necessary for plant nutrition. Additionally, as a product obtained from fermentation with the participation of bacteria, yeasts, and bacilli, when applied properly, they may also have phytohormonal, fungicidal, bactericidal, nematicidal, acaricidal effects, and act as a repellent against insects. Therefore, they act as a natural protector of cultivated plants against diseases and pests, with less damage to the environment and no danger to human health. Biofertilizer can be used in annual and perennial crops, in conventional and organic systems, and is mainly used in gardens and orchards. At the end of the fermentation process, after straining the material, the solid residue (sludge) remains on the sieve. This sludge can be cured and applied to the soil as fertilizer. This sludge contains a lot of fiber and nutrients and can be used as a base fertilizer at planting or as a periodic fertilizer applied around the plant canopy. Its absorption by the plant, unlike liquid biofertilizers, is slow, as with other solid organic fertilizers in general.

Liquid biofertilizers can be applied to leaves (foliar fertilizer), seeds, soil via fertigation, or in hydroponics, in diluted dosages. The absorption by plants occurs very quickly, making it very useful for short-cycle crops or in the quick treatment of plant nutritional deficiencies.

Studies conducted by Embrapa Semi-Arid, using melon as an indicator crop, identified three types of biofertilizers with the best performance: Vairo, Agrobom, and Biological Ferment.

Although biofertilizers show good results when applied according to recommendations, there are still criticisms regarding their performance in more demanding crops. It is important to note that each type of soil has a specific natural fertility and, therefore, promotes different growths of crops. The areas chosen for biofertilizer testing by Embrapa Semi-Arid had sandy soils with low natural fertility and organic matter content, allowing each biofertilizer to express its capacity to nourish the plant. Table 1 shows the chemical composition of two biofertilizers prepared and tested at Embrapa Semi-Arid.

For high yields and productivity, soil fertility must be improved through the use of composts and plant residues, which cannot be achieved in a short time. Some studies have shown that this result is only achieved with the combination of several agricultural practices and not just with the application of isolated products. In addition,

Document #4: GPT-4 translation into English.

ENGLISH TRANSLATION via CLAUDE > 2007 - EMBRAPA - Biofertilizantes PT

Preparation and Use of Liquid Biofertilizers

Alineaurea Florentino Silva1 José Maria Pinto2 Carla Regine Reges Silva França3 Sabrina Cordeiro Fernandes4 Tâmara Cláudia de Araújo Gomes5 Maria Sonia Lopes da Silva6 Ana Nery Barbosa Matos7

1 M.Sc., Researcher at Embrapa Semi-Arid in the area of Plant Science, P.O. Box 23, ZIP: 56302-970, Petrolina-PE. 2 D.Sc., Researcher at Embrapa Semi-Arid in the area of Fertigation. 3 Biologist, BCP Facepe/Embrapa Semi-Arid Fellow. 4 Agricultural Engineer, Embrapa Semi-Arid Fellow. 5 Agricultural Engineer, M.Sc., Researcher at Embrapa Coastal Tablelands-UEP Rio Largo-AL. 6 Agricultural Engineer, D.Sc., Researcher in Soil Management. Embrapa Soils-UEP Recife. Antônio Falcão Street, no. 402, Boa Viagem. ZIP 51020-240 Recife-PE. 7 Biologist, Embrapa Semi-Arid Fellow.

130 ISSN 1808-9984 Petrolina, PE May, 2007

What are liquid biofertilizers?

Liquid biofertilizers are natural products obtained from the fermentation of organic materials with water, in the presence or absence of air (aerobic or anaerobic processes). They can have a highly complex and variable composition, depending on the material used, containing almost all the macro and microelements necessary for plant nutrition. Furthermore, because it is a product obtained from fermentation, with the participation of bacteria, yeasts and bacilli, when properly applied, it can also have phytohormonal, fungicidal, bacteriological, nematicidal, acaricidal effects and repel insects. It therefore acts as a natural protector of cultivated plants against diseases and pests, with less damage to the environment and without danger to human health. Biofertilizer can be used on annual and perennial crops, in conventional and organic systems, being mainly used in vegetable gardens and orchards.

At the end of the fermentation process, after straining the material, a solid residue (sludge) remains in the sieve, which can be composted and applied to the soil as fertilizer. This sludge contains a lot of fiber and nutrients and can be used as foundation fertilizer at planting time or as periodic fertilizer applied around the plant canopy. Its absorption by the plant, unlike liquid biofertilizer, is slow, as are other solid organic fertilizers in general.

Liquid biofertilizers can be applied to the leaf (foliar fertilizer), to seeds, to the soil via fertigation or in hydroponics, in diluted doses. Absorption by plants takes place very quickly, so it is very useful for short-cycle crops or for the rapid treatment of plant nutritional deficiencies.

Studies carried out by Embrapa Semi-Arid, using melon as an indicator crop, indicated three types of biofertilizers with better performance: Vairo, Agrobom and Biological Fermented.

Document #4: Claude 3.5 translation into English.

Comunicado Técnico online 130

ISSN 1808-9984
Petrolina, PE
May, 2007

Preparation and Use of Liquid Biofertilizers

Alineaurea Florentino Silva[1] José Maria Pinto[2]
Carla Regine Reges Silva França[3]
Sabrina Cordeiro Fernandes[4] Tâmara Cláudia de Araújo Gomes[5] Maria Sonia Lopes da Silva[6]
Ana Nery Barbosa Matos[7]

What are liquid biofertilizers?

Liquid biofertilizers are natural products obtained from the fermentation of organic materials with water, in the presence or absence of air (aerobic or anaerobic processes). They can have a highly complex and variable composition, depending on the material used, containing almost all the macro and micro elements necessary for plant nutrition. Furthermore, as it is a product obtained from fermentation, with the participation of bacteria, yeasts and bacilli, when properly applied, it can also have phytohormonal, fungicidal, bacteriological, nematicidal, acaricidal and insect repellent effects. It therefore acts as a natural protector of cultivated plants against diseases and pests, with less damage to the environment and no danger to human health. Biofertilizer can be used on annual and perennial crops, in conventional and organic systems, and is mainly used in vegetable gardens and orchards. At the end of the fermentation process, after straining the material, a solid residue (sludge) is left in the sieve, which can be tanned and applied to the soil as fertilizer. This sludge contains a lot of fiber and nutrients and can be used as a foundation fertilizer when planting or as a periodic fertilizer applied around the top of the plant. Unlike liquid biofertilizer, it is absorbed by the plant slow, just like other solid organic fertilizers in general. Liquid biofertilizers can be applied to the leaf (foliar fertilizer), to seeds, to the soil via fertigation or in hydroponics, in diluted dosages. It is absorbed by plants very quickly, so it is very useful for short-cycle crops or for the rapid treatment of plant nutritional deficiencies.

Studies carried out by Embrapa Semi-Arid, with melon as the indicator crop, indicated three types of biofertilizers with the best performance: Vairo, Agrobom and Biological Ferment.

Although biofertilizers show good results when applied according to recommendations, there is still criticism of their performance on more demanding crops. It is important to point out that each type of soil has a specific natural fertility and therefore provides growth in different ways for crops. The areas chosen for testing the biofertilizers by Embrapa Semi-Arido had sandy soils with low natural fertility and organic matter , thus allowing each biofertilizer to express its ability to nourish the plant. Table 1 shows the chemical composition of two of the biofertilizers prepared and tested at Embrapa Semi-Arid.

[1] M.Sc., Researcher at Embrapa Semi-Arido in the area of Phytotechnology, C. P. 23, CEP: 56302-970, Petrolina-PE.
[2] D.Sc., Researcher at Embrapa Semi-Arido in the area of Fertirrigation.
[3] Biologist, BCP Facepe/Embrapa Semi-Arido Fellow.
[4] Agronomist, Embrapa Semi-Arid Fellow.
[5] Agricultural Engineer, M.Sc., Researcher at Embrapa Tabuleiros Costeiros-UEP Rio Largo-AL.
[6] Agrª Engª, D.Sc., Researcher in Soil Management. Embrapa Solos-UEP Recife. Rua Antônio Falcão, nº 402, Boa Viagem. CEP 51020-240 Recife-PE.
[7] Biologist, Embrapa Semi-Arid Fellow.

Embrapa

Document #4: DeepL translation into English.

Machine Translated by Google

Comunicado 130
Técnico

ISSN 1808-9984
Petrolina, PE
May, 2007

online online

Preparation and Use of Liquid Biofertilizers

Alineaurea Florentino Silva1
Jose Maria Pinto2
Carla Regine Reges Silva França3
Sabrina Lamb Fernandes4
Tamara Claudia de Araujo Gomes5
Maria Sonia Lopes da Silva6
Ana Nery Barbosa Matos7

What are liquid biofertilizers?

Liquid biofertilizers are natural products obtained from the fermentation of organic materials with water, in the presence or absence of air (aerobic or anaerobic processes). They can have a highly complex and variable composition, depending on the material used, containing almost all the macro and micro elements necessary for plant nutrition. Furthermore, as it is a product obtained from fermentation, with the participation of bacteria, yeasts and bacilli, when applied properly, it can also have phytohormonal, fungicidal, bacteriological, nematicidal, acaricidal and insect repellent effects. Therefore, it acts as a natural protector of cultivated plants against diseases and pests, with less damage to the environment and without risk to human health. Biofertilizers can be used in annual and perennial crops, in conventional and organic systems, and are mainly used in vegetable gardens and orchards.

At the end of the fermentation process, after straining the material, the solid residue (sludge) remains in the sieve, which can be cured and applied to the soil as fertilizer. This sludge contains a lot of fiber and nutrients and can be used as foundation fertilizer at the time of planting or as periodic fertilizer applied around the crown of the plant. Its absorption by the plant, unlike liquid biofertilizer, is

slow, as well as other solid organic fertilizers in general.

Liquid biofertilizers can be applied to the leaves (foliar fertilizer), to the seeds, to the soil via fertigation or in hydroponics, in diluted doses. Absorption by plants occurs very quickly, making them very useful for short-cycle crops or for the rapid treatment of nutritional deficiencies in plants.

Studies carried out by Embrapa Semi-Arido, using melon as an indicator crop, indicated three types of biofertilizers with the best performance: Vairo, Agrobom and Fermented Biological.

Although biofertilizers present good results when applied as recommended, criticism still arises regarding their performance in more demanding crops. It is important to emphasize that each type of soil has a specific natural fertility and, therefore, provides growth in different ways for crops. The areas chosen for testing biofertilizers by Embrapa Semi-Arido had sandy soils with low natural fertility and phosphorus content.

organic matter, thus allowing each biofertilizer to express its ability to nourish the plant. Table 1 shows the chemical composition of two of the biofertilizers prepared and tested at Embrapa Semi-Arido.

[1] M.Sc., Researcher at Embrapa Semi-Arido in the area of Phytotechnics, CP 23, CEP: 56302-970, Petrolina-PE.
[2] D.Sc., Researcher at Embrapa Semi-Arido in the area of Fertigation.
[3] Biologist, BCP Facepe/Embrapa Semi-Arid Scholarship Holder.
[4] Agricultural Engineer, Embrapa Semi-Arid Scholarship Holder.
[5] Enga Agra., M.Sc., Researcher at Embrapa Coastal Tablelands UEP Rio Largo-AL.
[6] Agricultural Engineer, D.Sc., Researcher in Soil Management, Embrapa Solos UEP Recife, Postal Code 51020-240 Recife-PE.
[7] Biologist, Embrapa Semi-Arid Scholarship.

Embrapa

Document #4: Google Translate translation into English.

10. **Translation of Visuals and Layout**

The two SLMs (DeepL and Google Translate) maintained all photos, graphs, illustrations, layout and format in their translations with very few errors (see appendix A3). However, the two LLM translations (GPT-4 and Claude 3.5) did not maintain these elements, and this is important because these elements are often important in reinforcing or giving context to portions of the text. The issues we discovered with the two LLM translations (which you can also see in appendix A3) included:

- most photos and graphic images in the translated versions are not included in translations, and the relevant captions are often missing as well;
- graphs, charts and tables that are digital images are mostly dropped in the same way as photos, the exception being tables that are comprised of segments of digital text, which are generally translated well; and
- layouts and formatting are not maintained, often making it harder to easily follow the written text.

A good example of the challenge with graphics is taken from Document #4 (yellow highlights used to track the image reference number below the image and in the text). The subject is how to make biofertilizer, and the images are key in showing the reader the steps involved.

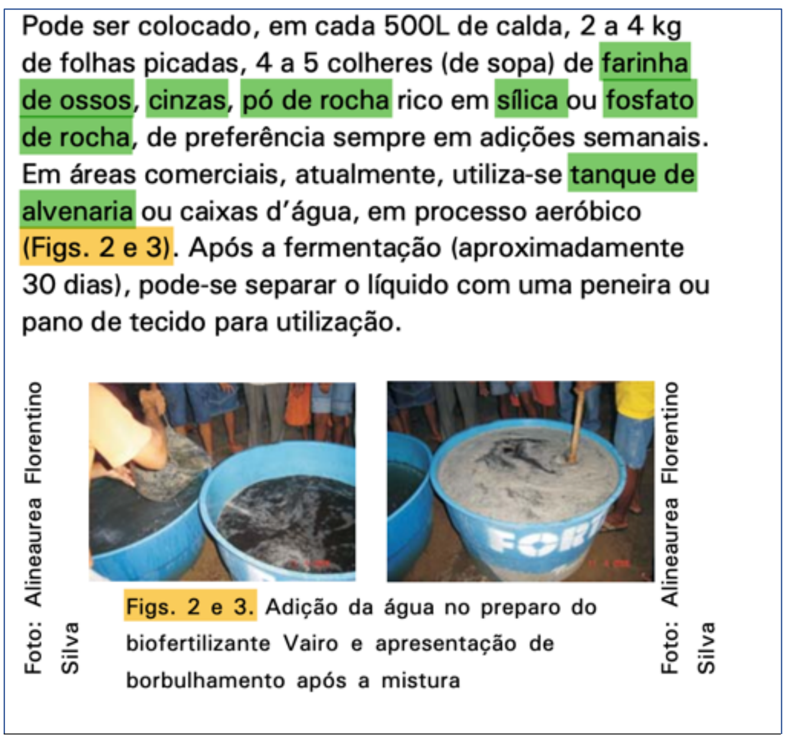

Pode ser colocado, em cada 500L de calda, 2 a 4 kg de folhas picadas, 4 a 5 colheres (de sopa) de farinha de ossos, cinzas, pó de rocha rico em sílica ou fosfato de rocha, de preferência sempre em adições semanais. Em áreas comerciais, atualmente, utiliza-se tanque de alvenaria ou caixas d'água, em processo aeróbico (Figs. 2 e 3). Após a fermentação (aproximadamente 30 dias), pode-se separar o líquido com uma peneira ou pano de tecido para utilização.

Foto: Alineaurea Florentino Silva

Foto: Alineaurea Florentino Silva

Figs. 2 e 3. Adição da água no preparo do biofertilizante Vairo e apresentação de borbulhamento após a mistura

Image from Document #4: Above is the original.

For every 500L of solution, 2 to 4 kg of chopped leaves, 4 to 5 tablespoons of bone meal, ashes, silica-rich rock dust, or rock phosphate can be added, preferably always in weekly additions. In commercial areas, masonry tanks or water tanks are currently used in an aerobic process. After fermentation (approximately 30 days), the liquid can be separated with a sieve or cloth for use.

Image from Document #4: This is the GPT-4 translation, and both the image and the reference to the image in the text "(Figs. 2 e 3)" is missing.

2 to 4 kg of chopped leaves, 4 to 5 tablespoons of bone meal, ashes, silica-rich rock powder or rock phosphate can be added to every 500L of mixture, preferably always in weekly additions.

In commercial areas, masonry tanks or water tanks are currently used, in an aerobic process (Figs. 2 and 3). After fermentation (approximately 30 days), the liquid can be separated with a sieve or cloth fabric for use.

[Image captions: Fig. 2 and 3. Addition of water in the preparation of Vairo biofertilizer and presentation of bubbling after mixing]

Image from Document #4: The Claude 3.5 translation above includes the reference to the image in the text, and also includes a translation of the caption. But the image itself is missing.

A second example of the challenge with graphics is taken from Document #1:

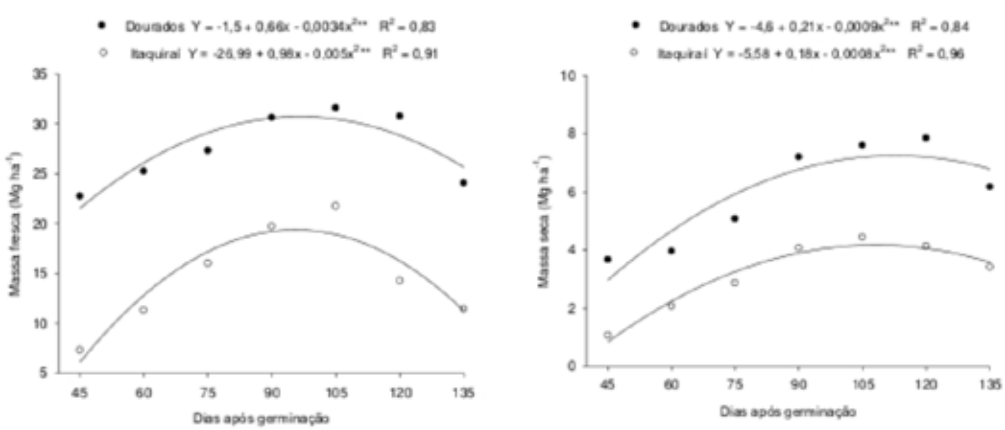
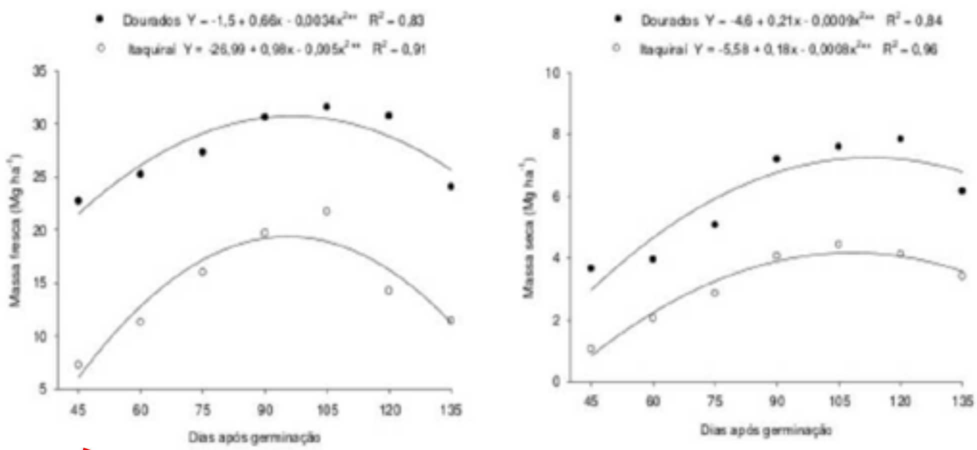


Figura 1. Produção de massa seca e fresca do feijão-de-porco ao longo do ciclo de cultivo em Dourados, MS (2007/2008) e Itaquiraí, MS (2008/2009).

**Significativo a 1% de probabilidade.

Image from Document #1: Original



Massa fresca	Fresh mass
Dias após germinação	Days after germination

Figure 1. Production of dry and fresh matter of jack beans throughout the cultivation cycle in Dourados, MS (2007/2008) and Itaquiraí, MS (2008/2009).
**Significant at 1% probability.

Image from Document #1: In this case, the professional translator did a separate translation next to the image, as indicated by the red arrow:

For this example from Document #1, GPT-4 or Claude 3.5 maintained the image but did not translate anything from it, although the text refers to it as “Figura 1.”

USING TRANSLATION GLOSSARIES TO IMPROVE THE ACCURACY OF LLMs + SLMs

APPENDIX

A5

Contents

1. Summary

The RARE Feasibility Study undertook a comprehensive analysis regarding the use of translation glossaries to enhance the accuracy of large language models (LLMs) and small language models (SLMs) in translating research documents on the subject of smallholder sustainable/regenerative agriculture. Based on comparative testing of GPT-4, Claude 3.5, DeepL Pro, and Google Translate, while LLMs achieved approximately 90% accuracy for text translation but struggled with visual and layout preservation, SLMs demonstrated around 80% text accuracy while better maintaining document visuals and formatting. To address the 10-20% accuracy gap, which becomes significant when scaling to thousands of documents across multiple languages, the implementation of translation glossaries is recommended. The analysis presents two primary approaches: traditional Excel/CSV-based glossaries, which offer static, structured terminology lists, and retrieval augmented generation (RAG) systems, which provide dynamic, context-aware translation support. The study recommends adopting a RAG-based glossary system, which can be continuously updated and refined through the proposed RARE Regional Network of academics, agronomists, and researchers. The proposed glossary structure includes essential fields such as source terms and translations, along with optional elements like contextual information, usage examples, and domain-specific details to ensure accurate and culturally appropriate translations.

2. Understanding How Glossaries Work

Appendix A4 presents a comparative analysis of AI-powered and professional translations of agricultural research papers and a training manual, evaluating four AI models (GPT-4, Claude 3.5, DeepL Pro, and Google Translate) on accuracy and terminology consistency for the text component as well as preservation of visuals and layout.

In this comparative analysis, AI translations achieved accuracy ratings of between 80 to 90% for the text component. When you then factor in eventually translating tens of thousands of research documents into multiple languages, this 10 to 20% gap in accuracy is considerably exacerbated.

While LLMs (GPT-4 and Claude 3.5) handled general text with around 90%, accuracy, they did not maintain visuals and layout. SLMs (DeepL Pro and Google Translate) performed with around 80% accuracy for text, but were significantly better at preserving visuals and layout. This led to making a recommendation to use SLMs as the primary AI translation tool for the RARE platform, but with the caveat that a translation glossary would be absolutely essential to increase the accuracy of the text translations.

All AI translation models recommend using a glossary to improve their accuracy in translating technical and cultural terms and phrases. There are two basic approaches when it comes to creating these glossaries.

The first involves manually creating an Excel or CSV format table with 1) the original word or phrase, 2) the translation of that word or phrase, and 3) various linguistic and contextual details such as part of speech, example sentences, domain or category, synonyms, definitions, priority levels, and source references to ensure accuracy and consistency in translation.

The second approach is to use a retrieval augmented generation (RAG) system, which is a tool for creating and continually updating a more comprehensive multilingual translation glossary. A key feature of RAG systems is to act as an intermediary, retrieving definitions and contextual information from up-to-date technical glossaries, incorporating input from subject matter experts, and using the resulting glossary to improve the accuracy of the LLM or SLM during the translation process. This external knowledge injection allows the language model to generate more accurate, contextually appropriate, and professional translations of technical documents.

Exploring the glossary options in depth led to a recommendation that RARE use an external RAG-based multilingual translation glossary that can be updated using, among other sources, the extensive reach of LLMs.

3. Excel/CSV vs RAG

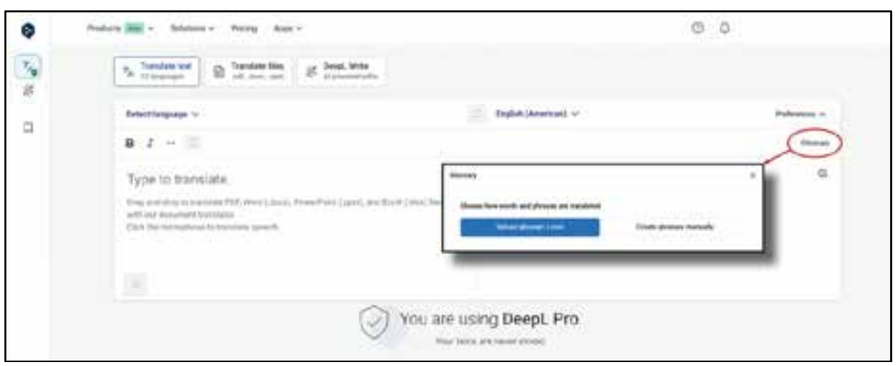
The use and integration of glossaries with AI-powered translation tools differ in key ways:

- **Excel/CSV Glossaries:** These are static, structured lists of terms and their preferred translations. They are typically uploaded or referenced in translation tools to ensure consistency, especially for technical and industry-specific terminology. All AI translation models, whether LLMs or SLMs, can use them when explicitly provided, but they do not "search" for glossary terms unless they are integrated into the translation prompt.

- **RAG-Based Glossaries:** These dynamically retrieve relevant glossary terms from a larger database in response to specific queries. Unlike static lists, RAG-powered glossaries allow LLMs and SLMs to pull definitions, translations, and contextual explanations on demand, making them more flexible for complex and evolving terminology needs.

While both formats help improve translation accuracy, Excel/CSV glossaries are best for enforcing strict terminology adherence, whereas RAG-based glossaries provide more adaptability, especially for nuanced or less frequently used terms.

Our recommendation is to go with a RAG-based glossary.



DeepL Pro page showing the Glossary button.

4. Role of the RARE Regional Network

In the course of reaching out to universities, government departments of agriculture, and smallholder-focused NGOs in connection with this Feasibility Study, it became clear that those we spoke with represent an invaluable and largely untapped resource when it comes to scaling smallholder regenerative agriculture.

This led to our recommendation to set up a RARE Regional Network of academics, agronomists and researchers to both find the existing documents as well as helping to coordinate and allocate new funding for research by members of the network.

Another important role will be to have this network contribute to the RAG-based glossary by collaboratively curating, refining, and expanding its content, ensuring that it remains a dynamic and authoritative resource for smallholder agriculture in the Global South. Drawing inspiration from Wikipedia's model, the glossary can be maintained through continuous input from academics, agronomists, and researchers who validate terms, add contextual insights, and integrate emerging regenerative agriculture practices.

Members can also engage by submitting new culturally specific terms, providing translations, and sharing real world case studies to enrich the glossary's practical relevance. A peer review mechanism can help uphold accuracy, while regular discussions and workshops foster knowledge exchange and adaptation to evolving agricultural challenges. By actively shaping and sustaining this resource, the network will ensure that smallholder farmers, policymakers, and extension workers have access to a continually evolving, context sensitive, and linguistically inclusive agricultural knowledge base.

One group that caught our attention was the Translation Glossary Project. They produced a glossary of development terminology with an indigenous community in Malawi which included dozens of Chichewa language terms that are not in published dictionaries.

5. Glossary Structure

A multilingual translation glossary, whether the manual Excel/CSV version or RAG-powered, has the same basic structure:

- **Essential Columns**

Column Name	Description
Source Term	The term in the original language (e.g., Portuguese).
Target Term	The translated term in the target language (e.g., English).
Part of Speech	Indicates whether the term is a noun, verb, adjective, etc.
Context/Example Sentence	A sentence showing how the term is used in context.
Domain/Category	Specifies the subject area (e.g., agriculture, legal, technical).
Notes	Additional explanations or instructions for usage.

- **Optional Columns**

Column Name	Description
Synonyms	Alternative terms with similar meanings.
Acronyms/Abbreviations	Abbreviated forms of the term, if applicable.
Definition	A brief definition of the term.
Priority Level	Indicates whether the term is preferred, deprecated, or neutral.
Source Reference	Notes on where the term was sourced from.

- **Example Table**

Source Term	Target Term	Part of Speech	Context/Example Sentence	Domain	Notes
Solo fértil	Fertile soil	Noun	A terra precisa de nutrientes para se tornar solo fértil.	Agriculture	Often used in soil management discussions.
Colheita	Harvest	Noun	A colheita deste ano foi abundante.	Farming	Avoid using "crop yield" in this context.