

REQUEST FOR PROPOSAL (RFP)
For Multilingual Data Collection Services for AgriAI Collect

Closing Date for Submission

June 15, 2026, at 23:59 PM

I. INTRODUCTION

LEHS is a charitable organization in India whose purpose is to offer basic health and education for the poor. LEHS in furtherance of charitable objectives through its flagship programs Wadhvani AI which aims to build equitable and sustainable systems by making quality primary healthcare available and accessible to the underserved population and to bring the benefits of modern AI technology to underserved populations by building and deploying AI solutions for social impact across domains such as healthcare, agriculture, governance and education in India. LEHS aims to promote the integration of technologies, particularly in emerging domains like artificial intelligence and innovations into the Indian mainstream primary healthcare, education, and agriculture systems through a partnership with the State and National Government, apex institutions, international agencies, and private sector partners e.g. innovators, social enterprises and other ecosystem contributors in line with its stated objectives for the betterment of society particularly focusing on projects of national and social significance.

Wadhvani AI, a unit of LEHS, focuses on developing, deploying, and evaluating artificial intelligence solutions to address critical social challenges in India, particularly in domains such as healthcare, agriculture, and education. As part of its ongoing initiatives in the agriculture sector, Wadhvani AI is developing an **image-to-text extraction model** aimed at identifying and extracting key textual information from **agricultural input product images**—such as fertilizer bags, pesticide containers, and seed packets. These AI models are being trained to accurately recognize and extract product names, compositions, usage instructions, and other relevant information **in English**, enabling better access to product-related information for farmers and agri-extension workers. In alignment with its quality assurance framework, Wadhvani AI will conduct rigorous evaluations to assess the accuracy, consistency, and applicability of the extracted textual data to ensure the model's effectiveness and reliability in real-world agricultural use cases.

II. BACKGROUND

India's agriculture sector is driven by over **93 million smallholder farming households**, with nearly 90% owning less than 2 hectares of land ([NDTV Profit](#)). While schemes like PM-KISAN offer income support ([Wikipedia](#)), smallholders still face barriers to accessing premium markets. **Certifications** such as India Organic and [GLOBALG.A.P.](#) are essential economic tools—organic produce can fetch **20–50% price premiums**, and the organic market nearly doubled from ₹169 crore in 2017–18 to ₹330 crore in 2022–23 ([Mint](#)). However, **manual and inconsistent data practices** delay audits and exclude many farmers from timely certification. To tackle this problem among many others in the agriculture data ecosystem, India is rolling out AgriStack, a national digital agriculture infrastructure, with **over 3.7 million Farmer IDs** created and **Digital Crop Surveys** deployed in **436 districts** ([Indian Express](#), [AgriStack](#)).

The current ecosystem is fragmented, analog, and heavily reliant on manual processes. Farm-level data—critical for audits and certification—is often captured through farmer diaries maintained by field workers or input digitally by farmers themselves. These diaries include important questions such as:

- What is the name of the fertilizer or pesticide you used?
- What is the brand name or product label?

- What input did you apply and when?

However, many smallholder farmers are unable to answer these questions due to low literacy, language barriers, or lack of familiarity with product labeling. Recognizing this gap, **Wadhvani AI has proposed a new data collection approach**: allowing farmers to simply **upload images of agricultural input products** (e.g., fertilizers, pesticides, seed packets) in response to these questions.

To support this, **Wadhvani AI is developing an image-to-text extraction model** that can automatically extract relevant information from these product images. The model is trained to detect and extract key product details, such as:

- Brand Name
- Product Name
- Chemical Composition
- Manufacturing and Expiry Date
- Product Label Color Code

This innovation would enable farmers to respond more accurately to diary questions without needing to read or transcribe complex product information. The extracted data is then automatically populated into the farmer diary, improving data completeness, consistency, and accuracy. By minimizing manual entry, reducing errors, and standardizing input records, this AI-powered solution strengthens traceability, supports certification workflows, and helps smallholder farmers better access premium markets—while reinforcing India's credibility in global agri-trade.

This RFP invites qualified agencies to undertake **multilingual data collection services** aligned with this objective.

III. Overview of Geography

1. The data collection is expected to be done in the mentioned states i.e. **Rajasthan, Madhya Pradesh, Gujarat and Maharashtra.**

IV. SCOPE OF WORK

The selected agency will be responsible for providing high-quality image collection services to support the development and evaluation of the Image to Text Extraction model. The scope of work includes, the following:

- Collect an image set of **20,000 unique agricultural input product images** (5,000 unique product sets per language: **English, Hindi, Marathi, and Gujarati**).
- The dataset must include, but not limited to, images from the following categories:
 - Seeds and Planting Material
 - Certified seeds
 - Hybrid seeds
 - Genetically modified seeds (GMOs)
 - Fertilizers
 - Chemical fertilizers
 - Nitrogenous (e.g., Urea, Ammonium Nitrate)
 - Phosphatic (e.g., DAP, SSP)
 - Potassic (e.g., Muriate of Potash)

- Micronutrients (e.g., Zinc Sulfate, Boron)
 - Macronutrients (e.g. Sulphur, Calcium, Magnesium)
- Organic fertilizers
 - Farmyard manure (FYM)
 - Vermicompost
 - Green manure
 - Oil cakes (neem, groundnut, etc.)
- Biofertilizers
 - Rhizobium
 - Azospirillum
 - Azotobacter
 - Phosphate-solubilizing bacteria (PSB)
 - Mycorrhiza
- Plant Protection Inputs
 - Chemical pesticides
 - Insecticides
 - Fungicides
 - Herbicides
 - Nematicides
 - Other option
- Biopesticides
 - Bacillus thuringiensis (Bt)
 - Trichoderma
 - Pseudomonas fluorescens
 - Neem-based products etc.,
- Soil and Water Management Inputs
 - Soil conditioners
 - pH regulators
- The agricultural input product images to be taken in settings, but not limited to, like:
 - a. Agricultural supply stores
 - b. Warehouses
 - c. Farms
 - d. Retail outlets
 - e. Outdoor environments
- The agency will use our internal data collection tool for Image Collection activity.
- Conduct training sessions for the data collectors on the usage of the tool prior to the commencement of the data collection activity.
- Image collection will be conducted in **batches** as given under "[Submission Batches and Review Timelines](#)".
- Each submitted batch will undergo a **25% quality review** by Wadhvani AI. If issues are found, feedback will be provided, and the agency will be required to **incorporate** the feedback in the upcoming batches.
- Each collection batch must be completed and submitted within the agreed timeline. **Under no circumstances should there be a delay of more than 5 days from the original submission**

due date, unless explicitly approved by Wadhvani AI. Timely delivery is critical to ensure smooth progress of the overall project.

Data Privacy and Compliance


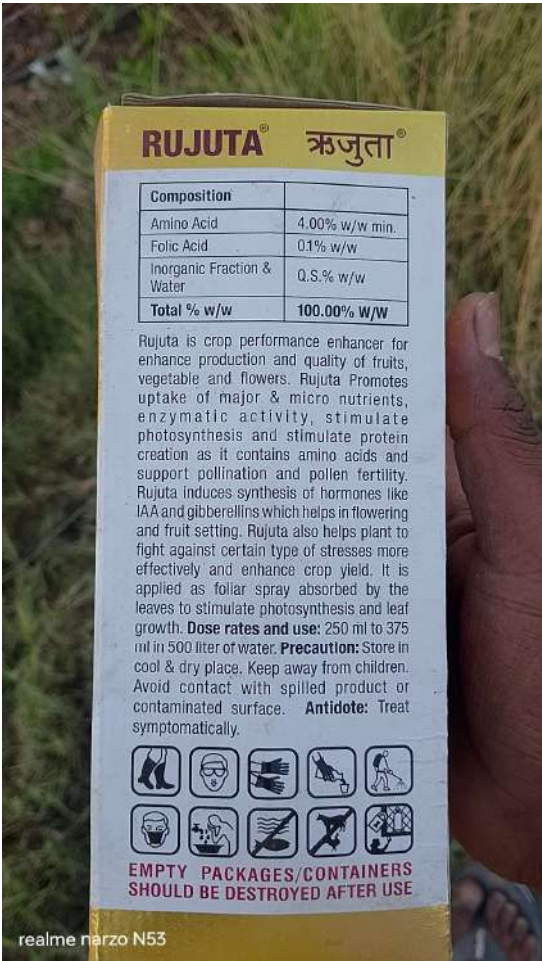
The agency must strictly adhere to the provisions of the **Digital Personal Data Protection (DPDP) Act, 2023**, as Wadhvani AI is also fully compliant with this law. The agency shall be fully responsible for ensuring the privacy, security, and protection of all personal data involved in this project. Any **data leakage, unauthorized access, or misuse**—whether by the **agency** or any **third-party platform** it engages—will be considered **a breach of the DPDP Act**. In such cases, the agency will be held accountable and subject to any legal or contractual consequences arising from such violations. Additionally, if the agency provides any **Personally Identifiable Information (PII)**—including **personal photos, selfies, or any out-of-domain and sensitive images**—it shall bear full responsibility in the event such data is inadvertently used in any of Wadhvani AI’s machine learning model training processes. It is therefore mandatory for the agency to refrain from providing any PII data during the entire data collection activity.

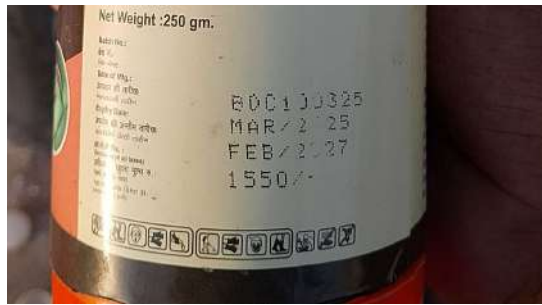


V. Required Deliverables

An **Image Collection Requirement Document** will be shared along with this RFP. Please find the following deliverables:

- Collect an image set of **20,000 unique agricultural input product images** (5,000 unique product sets per language: **English, Hindi, Marathi, and Gujarati**).
- The form will contain questions as given below for each question, you have to upload an image containing that information and enter the metadata for each image in the text field, as given below:

Key Product Details (questions in the form)	Image to be uploaded (upload image field)	Metadata (text field)
Brand Name		IFFCO

<p>Product Name</p>		<p>NANO DAP</p>										
<p>Chemical Compositions</p>	 <table border="1" data-bbox="598 974 869 1120"> <thead> <tr> <th>Composition</th> <th></th> </tr> </thead> <tbody> <tr> <td>Amino Acid</td> <td>4.00% w/w min.</td> </tr> <tr> <td>Folic Acid</td> <td>0.1% w/w</td> </tr> <tr> <td>Inorganic Fraction & Water</td> <td>Q.S.% w/w</td> </tr> <tr> <td>Total % w/w</td> <td>100.00% W/W</td> </tr> </tbody> </table> <p>Rujuta is crop performance enhancer for enhance production and quality of fruits, vegetable and flowers. Rujuta Promotes uptake of major & micro nutrients, enzymatic activity, stimulate photosynthesis and stimulate protein creation as it contains amino acids and support pollination and pollen fertility. Rujuta induces synthesis of hormones like IAA and gibberellins which helps in flowering and fruit setting. Rujuta also helps plant to fight against certain type of stresses more effectively and enhance crop yield. It is applied as foliar spray absorbed by the leaves to stimulate photosynthesis and leaf growth. Dose rates and use: 250 ml to 375 ml in 500 liter of water. Precaution: Store in cool & dry place. Keep away from children. Avoid contact with spilled product or contaminated surface. Antidote: Treat symptomatically.</p> <p>EMPTY PACKAGES/CONTAINERS SHOULD BE DESTROYED AFTER USE</p> <p>realme narzo N53</p>	Composition		Amino Acid	4.00% w/w min.	Folic Acid	0.1% w/w	Inorganic Fraction & Water	Q.S.% w/w	Total % w/w	100.00% W/W	<p>Amino Acid - 4.00% w/w min. Folic Acid - 0.1% w/w Inorganic Fraction & Water - Q.S% w/w Total% w/w - 100.00% w/w</p>
Composition												
Amino Acid	4.00% w/w min.											
Folic Acid	0.1% w/w											
Inorganic Fraction & Water	Q.S.% w/w											
Total % w/w	100.00% W/W											

<p>Expiry Date (Exp Date)</p>		<p>FEB/2025</p>
<p>Manufacturing Date (Mfg Date)</p>		<p>27/11/24</p>
<p>Product Label Color Code</p>		<p>Blue: Danger Yellow: Danger</p>

- **Note:** The required information may need to be captured through multiple images if it appears on different sides of the product (such as front, back, sides, top, or bottom). Additionally, the same

product name may appear in different languages on the packaging, but it should not be repeated in the same language.

- In addition to digital collection through the tool, the agency is expected to submit a **weekly summary report** detailing:
 - Number of products and images captured per week
 - Distribution by category (e.g., fertilizers vs pesticides)
 - Any challenges or anomalies faced in our tool or on the field during data collection.

VI. Submission Batches and Review Timelines

To streamline the review and quality assurance process, submissions must be divided into 4 batches across all 4 languages. Wadhvani AI (WIAI) will conduct a **25% quality review** for each batch before the next submission.

Batch	No. of Images	Languages (Hindi, Marathi, Gujarati, English)	Submission Deadline	WIAI Review Timeline (25% sample)
Batch 1	2000	500 unique product images for each language	Week 1	2 days
Batch 2	2000	500 unique product images for each language	Week 2	2 days
Batch 3	2000	500 unique product images for each language	Week 3	2 days
Batch 4	2000	500 unique product images for each language	Week 4	2 days
Batch 5	2000	500 unique product images for each language	Week 5	2 days
Batch 6	2000	500 unique product images for each language	Week 6	2 days
Batch 7	2000	500 unique product images for each language	Week 7	2 days
Batch 8	2000	500 unique product images for each language	Week 8	2 days

Batch 9	2000	500 unique product images for each language	Week 9	2 days
Batch 10	2000	500 unique product images for each language	Week 10	2 days

- Each batch must be submitted upon completion, followed by a **review and approval cycle**.
- Wadhvani AI will conduct a **25% quality review** for each batch (across all languages).
- Corrections and re-submissions (if required) must be completed alongside ongoing collection activities if more than **15% images** do not follow protocol.
- No batch may be delayed by more than **5 days** without explicit written approval from WIAI.

VII. Evaluation Criteria for Technical Proposal

Financial proposals will only be evaluated for participants who qualify in the technical evaluation. The financial assessment will consider several key criteria: the total cost, a detailed cost breakdown, cost reasonableness, cost-effectiveness, and the proposed payment terms. Furthermore, the financial stability of the participating organization will be a factor.

Crucially, all participants must demonstrate compliance with the Digital Personal Data Protection (DPDP) Act, 2023, specifically in the responsible collection, storage, and usage of students' personal data.

Domain	Component	Evaluation Criteria
Technical Proposal	Understanding of Scope and Targets	<ol style="list-style-type: none"> 1. Demonstrated clarity on the requirement of collecting 20,000 unique agricultural input product images across 4 languages (English, Hindi, Marathi, Gujarati). 2. Understanding of 5,000 unique product sets per language without duplication. 3. Awareness of target geographies (Rajasthan, Madhya Pradesh, Gujarat, Maharashtra) and diverse collection environments (retail shops, farms, warehouses, etc.). 4. Clarity on capturing complete product information through multiple images where required.
	Collection Methodology and Execution Plan	<ol style="list-style-type: none"> 1. Detailed approach for multilingual data collection across 4 states. 2. Deployment plan for field staff (language-wise and state-wise allocation). 3. Strategy for device readiness and compatibility

		with the provided data collection tool. 4. Plan for batch-wise submission (10 batches) within defined timelines.
	Quality Assurance Mechanisms	<ol style="list-style-type: none"> 1. Robust processes to ensure adherence to image capture protocols and metadata accuracy. 2. Approach to maintain high-quality, non-duplicate, and complete datasets. 3. Mechanism to incorporate feedback from Wadhvani AI's 25% quality review per batch. 4. Plan for correction and resubmission if quality thresholds (>15% errors) are not met.
	Feedback and Issue Resolution Mechanisms	<ol style="list-style-type: none"> 1. System to capture real-time feedback from field teams. 2. Define escalation and resolution processes for operational and technical challenges.
	Operational Monitoring and Reporting	<ol style="list-style-type: none"> 1. Framework for monitoring field progress against timelines. 2. Plan to ensure no delays beyond the permissible 5-day limit per batch. 3. Approach for regular reporting, dashboards, and progress updates.
	Training Plan for field staff	<ol style="list-style-type: none"> 1. Structured training plan for data collectors on tool usage and image capture protocols. 2. Approach for continuous learning, refresher sessions, and quality reinforcement.
	Coverage Strategy	<ol style="list-style-type: none"> 1. Strategy to ensure comprehensive coverage of all specified product categories (seeds, fertilizers, pesticides, etc.). 2. Plan to ensure diversity in brands, products, and real-world conditions across geographies.
Organizational Profile and Past Experience	Relevant Experience	Prior experience with NGOs, government programs, agriculture, or development sector projects.
	Technology Deployment and Experience	<ol style="list-style-type: none"> 1. Experience in large-scale field data collection using digital tools. 2. Familiarity with image-based or AI/ML data

		collection workflows is preferred.
	Work Samples/ Testimonials	Submission of case studies, reports, or client testimonials demonstrating similar project execution.
Team Composition and Experience	Agriculture Domain Expertise	Demonstrated experience in agriculture extension services or field-level engagement.
	Team Qualifications and Experience	Qualifications and relevant experience of field coordinators, supervisors, and project managers.
	Field Operations Capacity	1. Ability to recruit, train, deploy, and manage large distributed field teams. 2. Proven capability to execute time-bound, high-volume data collection projects.
Compliance and Safeguards	Data Privacy and Security	1. Strict adherence to the Digital Personal Data Protection (DPDP) Act, 2023. 2. Clear measures to ensure: <ul style="list-style-type: none"> • No collection of Personally Identifiable Information (PII) (e.g., personal photos, selfies). • Secure data storage, transmission, and access controls. • Prevention of data leakage, misuse, or unauthorized access.
	Ethical Safeguards	1. Practices ensuring responsible and ethical data collection in the field. 2. Safeguards for informed participation and non-intrusive data capture.

Only participants in RFP qualifying on technical evaluation will be eligible for financial evaluation. Financial proposals will be evaluated on criteria including **total cost, cost breakdown, cost reasonableness, cost-effectiveness, payment terms, and financial stability** of the participating organization.

VIII. Proposal Submission Guidelines

Submission Details

- All responses to this RFP must be received no later than **June 15, 2026**. The proposal should be submitted only through e-mail in PDF format addressed to the Procurement Team in the below-given e-mail id: rfp.lehs@wadhwaniai.org.

- The email's subject line must contain the reference number and title of the RFP: **Proposal For Multilingual Data Collection Services for AgriAI Collect - WIAI (Name of the Agency)**.
- Any proposals exceeding the upper cap of **INR 7,08,000** will be liable to be rejected.
- Any proposals received by Wadhvani-AI after the deadline for submission of proposals prescribed in the timeline of this document are liable to be rejected.

IX. Mandatory Enclosure

Please submit the following as enclosures or attachment with your quotation. Bidders must provide all the information requested below. Quotations that do not provide the required information and certificates, or do not follow the submission requirements may not be reviewed.

- Company Profile, testimonials, work completion reports, purchase orders, reference (NGOs).
- GST registration certificate.
- PAN registration.
- TAN registration.
- Audited Financials for last three years.
- Income Tax Return for last three years.
- Cancelled cheque.
- Copy of registration documents/certificate and most recent renewal as a legal entity.
- Preferably Private Ltd company and has sound business records.

All responses to this RFP must be received no later than **June 15, 2026**. The proposal should be submitted only through e-mail in PDF format addressed Procurement Team in the below-given e-mail id: rfp.lehs@wadhwaniai.org

For clarifications and queries, please direct all queries to rfp.lehs@wadhwaniai.org by June 10, 2026.

Note: Only shortlisted vendors will be contacted for presentations or negotiations. If you do not hear from us within two weeks of submission, consider your proposal not selected. LEHS reserves the right to reject any or all proposals and to negotiate terms and conditions with the selected vendor.