REQUEST FOR PROPOSAL (RFP)

Odia Annotation of Student Voice Recordings from Odisha

Closing Date for Submission
<February 27, 2026>

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, Wadhwani 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.

Wadhwani 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.

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. In line with its mission to support projects of national and social significance, LEHS also undertakes rigorous monitoring, evaluation, and learning (MEL) activities to assess the impact, usability, and scalability of different programmatic interventions.

II.   BACKGROUND

A significant number of students in Indian public schools (Grades 1–8) lack foundational reading skills, impacting their academic progression across all subjects. This challenge is compounded by the absence of scalable, data-driven tools for assessing reading in diverse languages and for identifying specific proficiency levels to enable targeted interventions.

In response, Wadhwani AI has developed an AI-powered tool that **automates the assessment of foundational reading skills**. The tool leverages highly contextualized and fine-tuned Automated Speech Recognition (ASR) models to analyze word and paragraph

Head Office: 70, Ring Road, 2nd Floor,
Lajpat Nagar - III, New Delhi - 110024, India
Registration No.: S/47696 of 2003

WADHWANI AI
A program of the AI Unit of LEHS

reading, generating key performance metrics and identifying specific reading miscues. Our solution also applies an intelligent layer to cohort students into reading levels, providing teachers with actionable insights for planning targeted classroom interventions.

To date, Wadhwani AI has assessed **over 7 million students in Gujarat and Rajasthan** for the Gujarati and Hindi languages**.** We are now adapting the tool for the Odia language for the state of Odisha. To build a highly contextualized ASR model for Odia across grades, we need human-generated transcripts of students' voice recordings reading paragraphs, which we have collected from select districts of Odisha in partnership with the Odisha Government.

III. PROJECT OBJECTIVES

To create timestamped human transcriptions of student voice recordings from Grades 3–8 reading sets of paragraphs in Odia, collected from select districts in Odisha. For Grades 3–8, each individual word heard in the recording must be timestamped and transcribed verbatim in Odia exactly as heard by the human annotator (no corrections for grammar, spelling, or pronunciation). WIAI will provide the partner with:

- Audio Files: Recordings of students reading Odia paragraphs.
- Reference Text: The exact Oodia paragraph the student was instructed to read; this text will be used to identify miscues.

IV. SCOPE OF WORK

Summary of Activities

Sr.no	Activity Description	Timeline
1	Co-Working & Onboarding Phase	Jan End - Feb 2026
2	Data Annotation for ASR: Create timestamped human transcriptions of the collected data following Wadhwani AI's finalized data annotation protocol.	

1. Co-Working & Onboarding Phase

The selected partner shall participate in a structured co-working and onboarding phase during the first week following contract finalisation. Participation in this phase is mandatory and is a prerequisite for commencing full-scale annotation.

The objective of this phase is to establish end-to-end operational alignment between Wadhwani AI and the partner prior to full-scale annotation, including a demonstration of the

Head Office: 70, Ring Road, 2nd Floor,
Lajpat Nagar - III, New Delhi - 110024, India
Registration No.: S/47696 of 2003

WADHWANI AI
A program of the AI Unit of LEHS

annotation tool by the partner, an overview of the Odia annotation protocol, and alignment on quality assurance and review processes.

Full-scale annotation may commence only after formal readiness sign-off by Wadhwani AI.

2. Data Annotation for ASR

Create timestamped human transcriptions of the collected student voice recordings in Odia, following **Wadhwani AI's finalized data annotation protocol**, including verbatim transcription, word-level and sub-lexical timestamping, and application of defined labels

Total Number of Hours to be annotated: **~150 Hours for Odia**

Data Annotation Volume:

Grades 3 to 8	
Paragraph	~150 Hours of Recording
	Collected from a subset of approximately 13,000 students across Grades 3–8

The selected partner is expected to provide:

Human Transcription

- Each individual student recording needs to be transcribed in **Odia** with timestamps for each individual word.

- **Crucial Note:**

  ○ The transcription must reflect the speech exactly as heard (verbatim). Annotators must NOT correct grammar, fix pronunciation, fill in intended words, or infer meaning.

- **Timestamping:**

  ○ For paragraphs, the transcription must include **precise timestamps** (e.g. HH:MM: SS:MS) for *each individual word*. Where words are broken down sub-lexically (e.g., sound-by-sound reading), timestamps must be provided for the **sub-lexical portions** rather than the word as a whole.

Labels

- In addition to the speech transcript, wherever applicable, labels are to be added to the audio to denote the presence of Unintelligible Speech, Fillers and Pauses, Background Noise, Adult voice, Prompting etc.

- The definitive labels for Odia annotations will be provided by WIAI in the detailed data annotation protocol.

Quality Review

- An internal system of quality checks should be in place for every transcription, both for the quality of transcription and for the quality of labels.

- The transcripts should be delivered to Wadhwani AI in batches of **~20 hours each**, and two rounds of quality check feedback might be given by Wadhwani AI on each batch. WIAI's review will include a review of transcriptions as well as a review of labels.

- WIAI will utilize an acceptance sampling plan to assess the quality of the delivered batches. A statistically significant sample (sample size determined using the standard formula of estimating proportions) from each batch will be inspected. Should 10-20% or more of the reviewed sample fail to meet the required quality standards (Timestamp accuracy within a 25ms margin of Error and Labeling Accuracy), WIAI reserves the right to reject the entire batch.

Data Policy

- Wadhwani AI will share anonymized data, stripped of PII, uploaded in a GCP/AWS bucket with the selected partner, from the data collected.

- The partner needs to have processes in place to prevent the download of data by individual annotators.

- The Partner is required to sign an undertaking guaranteeing that this voice data will **not be used for any proprietary purposes** outside of this contractual agreement.

Tool for Annotation

- The responsibility for providing or procuring a suitable annotation tool rests with the partner. The selected tool must fully support all required transcription and labeling functionalities outlined above, including accurate timestamping, word-level segmentation, and tag application in accordance with the defined annotation protocol.

Formatting & Delivery

Head Office: 70, Ring Road, 2nd Floor,
Lajpat Nagar - III, New Delhi - 110024, India
Registration No.: S/47696 of 2003

WADHWANI AI
A program of the AI Unit of LEHS

Delivering transcriptions in a JSON format as specified below:

```
{
  "id": 13945,
  "file_name": "audio_student_04195.wav",
  "annotations": [
    {
      "start": "0:00:00.519350",
      "end": "0:00:01.209974",
      "Transcription": [
        "ଉପରେ"
      ],
    },
    {
      "start": "0:00:01.232074",
      "end": "0:00:01.651974",
      "Transcription": [
        "ଲେଖାରେ"
      ],
    }
    ....
}
```

Note: Transcriptions should **strictly** adhere to the terminology provided in this JSON format. Entries will be rejected if not compliant.

Regional Representation

- Annotators should be **native or proficient Odia speakers** with the ability to accurately discern and interpret different Odia accents.

Sample Odia paragraphs reference text, which will be shared:

| Paragraph | ଝିଅଟିଏ ଥିଲା । କବିତା ଲେଖାରେ ତା'ର ଭାରି ମନ । କି ଦିନ କି ରାତି, ସେ ଏକୁଟିଆ ବସେ । ମନେ ମନେ କ'ଣ ଗୁଣୁଗୁଣୁ ହୋଇ ଖାତାରେ ଲେଖି ପକାଏ । ବେଳ ପାଇଲେ ବାହାରେ ବୁଲି ସୁନ୍ଦର ସୁନ୍ଦର ଫୁଲ, ଚଢେଇ, ଗଛଲତା ସବୁକୁ ଅନେଇ କ'ଣ ଭାବେ । ସତେ ଯେପରି ଏମାନେ ତା'ର ଅତି ଆପଣାର । ଏକଥା ମାଆଙ୍କ ଛଡା ଅନ୍ୟ କେହି ଜାଣିପାରନ୍ତି ନାହିଁ । ମାଆ ଝିଅକୁ ଉସାହ ଦିଅନ୍ତି । ବାପା କିନ୍ତୁ ଚାହାନ୍ତି ଓଲଟା । ବାପାଙ୍କ ଇଚ୍ଛା, ଝିଅ ଅଙ୍କ କଷ୍ଟ, ହିସାବ କିତାବ ଭଲ ଜାଣୁ । ଠାରେ ବାପା ଝିଅକୁ କେତେଖଣ୍ଡ ଅଙ୍କ କଷିବାକୁ ଦେଲେ । ମାତ୍ର ଝିଅର ଅଙ୍କ କଷିବାରେ ମନ ନାହିଁ । ଅଙ୍କ କଷିବାପାଇଁ ବାଧକରି ବାପା ତାକୁ ଗୋଟିଏ କୋଠାରୀ ଭିତରେ ରଖି ବାହାରୁ କବାଟ କିଲିଦେଲେ । |

	କେତେ ସମୟ ପରେ ଆସି କବାଟ ଖୋଲିଲେ । ମାତ୍ର ଏ କ'ଣ ? ଝିଅଟି ଅଙ୍କ ନ କଷି ଲେଖି ପକାଇଛି । ଗୋଟିଏ ସୁନ୍ଦର କବିତା । କବିତାଟି ଲେଖାଯାଇଛି ଇଂରାଜୀ ଭାଷାରେ ।

Proposals should clearly outline the methodology, an overview of the data annotation tool, data policy, team expertise, and pricing.

The selected partner is expected to complete the transcription activity within ~4 weeks and **submit the first batch of 20 hours within the first week of finalisation of the contract.**

Payment eligibility will be linked to the successful submission of the entire scope of work. The partner may raise a single consolidated invoice after delivery. Additionally, the partner will be required to participate in weekly progress reviews to ensure timely completion and adherence to quality standards.

V.   TIMELINE AND ROUNDS

Sr.no	Activity	Date
1	Release of the Request for Proposal (RFP)	February 16, 2026
2	Q&A and Clarifications	February 24, 2026
3	Last date for submission of the complete proposal	February 27, 2026

VI.   REQUIRED DELIVERABLES

The selected partner will be responsible for delivering the following within the agreed project timelines and quality standards:

- **JSON files of transcriptions of recordings in Odia**

- **Each recording to have a separate JSON file, format as specified by Wadhwani AI**

VII. PROPOSAL SUBMISSION GUIDELINES

Your proposal must include:

1. **Technical Proposal**
   o   Understanding of scope and deliverables
   o   Process and tools proposed to be used for the task
   o   Team composition and qualifications
   o   Quality assurance mechanisms
   o   Data Policy
   o   Annotation of the Sample Odia Audios
      Sample 1
      Sample 2
      Sample 3
      Sample 4

2. **Financial Proposal**
   o   Detailed cost breakdown
   o   Schedule of Payments Linked to Milestones Achieved
   o   Applicable taxes and contingencies

3. **Organizational Profile**
   o   Relevant project experience
   o   Sample of similar work

Criteria for Application

Applicants must meet the following criteria.

- **Mandatory:** Demonstrated team capacity to provide transcriptions in Odia, demonstrable quality check process

- **Preferred:** Prior experience in transcribing Odia student speech.

Submission Details

- All proposals to this RFP must be received no later than **<February 27th>.** The proposal should be submitted only through e-mail in PDF format addressed to The Procurement Team at 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: **Odia Transcriptions of Student Voice Recordings of Reading - Odisha (Name of the partner)**.

- Any proposals received by Wadhwani-AI after the deadline for submission of proposals prescribed in the timeline of this document are liable to be rejected.

VIII. EVALUATION CRITERIA FOR TECHNICAL PROPOSAL

Criteria	Weightage

Head Office: 70, Ring Road, 2nd Floor,
Lajpat Nagar - III, New Delhi - 110024, India
Registration No.: S/47696 of 2003

![WADHWANI AI - A program of the AI Unit of LEHS]

Understanding the scope	5%
Technical soundness of approach and samples received (tools used to create timestamped transcripts)	35%
Data Quality Assurance	10%
Work plan and staffing	15%
Credentials of the organization in transcribing in the Odia language, including phonetic miscues.	25%
Credentials of the proposed team	10%
Technical Total	**100%**

Only participants in the RFP qualifying on the technical evaluation will be eligible for the 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. Also, the participants should show a demonstrated compliance with the provisions of the Digital Personal Data Protection (DPDP) Act, 2023, ensuring responsible collection, storage, and use of the students' personal data.

IX. MANDATORY ENCLOSURE

Please submit the following as enclosures or attachments with your quotation. Bidders must provide all the information requested above in section VIII. 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, and reference (NGOs).

- NGOs need to provide a valid CSR-1 issued in favor of the NGO by the MCA.

- GST registration certificate.

- PAN & TAN registration.

- Audited Financials for the last three years.

- Income Tax Return for the last three years.

- Cancelled cheque.

- Copy of registration documents/certificate and the most recent renewal as a legal entity.

- Preferably a Private Ltd company and has sound business records.

- Authorized signatory letter.

- Conflict of interest declaration.

- Contact person details.
- Memorandum of Association & Articles of Association (or Trust Deed / Bye-laws as applicable).

- List of Governing Body Members/Board Members of the organization.

- 12A and 80G Registration Certificates and Revalidation Document.

- PF, ESI, and Professional Tax Registration Certificates (if applicable).

- Finance Policy, HR Policy, Anti-Sexual Harassment Policy, Procurement Policy, and any other relevant internal policies.

- List of Pending Litigation/Cases against the Entity (if any).

- Latest TDS Return filed (Form 26Q and 24Q).

- GSTR filings for the last 3 returns (if applicable).

- Sample Audited Utilization Certificate (UC) submitted to any donor for CSR Funding.

- Audit Report in Form 10B/BB for the last 3 years.

- Foreign Contribution (FC) Audited Financials for the last 3 years (if applicable).

- FCRA Returns (FC-4) for the last 3 years (if applicable).

Note: Only shortlisted partners 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 chosen partner.

Head Office: 70, Ring Road, 2nd Floor,
Lajpat Nagar - III, New Delhi - 110024, India
Registration No.: S/47696 of 2003

WADHWANI AI
A program of the AI Unit of LEHS