

REQUEST FOR PROPOSAL (RFP) English Transcriptions of Student Voice Recordings of Reading of Rajasthan

Closing Date for Submission
<November 21, 2025, at 11:59 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, 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

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, Wadhvani AI has assessed **over 7 million students in Gujarat and Rajasthan** for Gujarati and Hindi languages. We are now adapting the tool for the English language for the state of Rajasthan, and to build a highly contextualised ASR model for English, we need human-generated transcripts of students' voice recordings, reading words and paragraphs that we have collected from all districts in Rajasthan in partnership with the Rajasthan Government.

III. Project Objectives:

To create timestamped human transcriptions of student voice recordings from Grade 3 to 8 reading sets of words and paragraphs in English collected from 41 districts in Rajasthan. Each individual word heard in the recording needs to be timestamped and transcribed as heard by the human annotator. WIAI will provide the partner with:

- Audio File: A recording of a student reading (either a list of words or a paragraph).
- Reference Text: The exact Word List or Spoken Paragraph the student was instructed to read. This text will be used to identify miscues.
- Standard voice recordings: 10-15 ideal readings of each data collection item (Word/Para). This will serve as a match reference for the annotators.

Language: English

Total Number of Hours to be annotated: **~150 hours**

IV. SCOPE OF WORK

The selected partner is expected to provide:

Human Transcription

- Each individual student recording needs to be transcribed in **English** 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** 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.

- For Words: A JSON with timestamps will be provided to the partner. Transcription for each time-stamped portion will have to be done by the partner.

- **Reference System for Correctness:**

- **Tier 1: Google's English Pronunciations:** If the student's pronunciation matches any of the **UK, US, or Indian English** pronunciations on Google, it **must** be treated as a correct reading. Each correctly pronounced English word needs to be transcribed in plain English (e.g., "house", "garden")
- **Tier 2: Contextual reference (Acceptable Variations):** Ideal readings provided by WIAI will serve as the acceptable variation match set for regional, non-error-inducing accent features. Annotators must familiarize themselves with these samples.
- The utterance/speech is considered a **miscue or mispronunciation** if the pronunciation deviates significantly, resulting in:
 - i. **Incorrect Word:** A different, existing English word (e.g., "house" for "horse"). In this case, type the incorrect word.
 - ii. **Invented Word:** A clearly spoken word that is *not* a valid English word (e.g., "blap" for "black"). In this case transcribe the invented word in Devanagari.
 - iii. **Unintelligible Segment:** The sound is too ambiguous or mumbled to identify. A special tag for mumbling needs to be added in this case.
- Annotators must strictly comply with the detailed protocol for English recording annotations, which shall be furnished by WIAI.

Labels

- In addition to the speech transcript, wherever appropriate, labels are to be added to the audio to denote the presence of Fillers, Noise, Long silences, Mumbling, Multiple Speakers, and accent-influenced pronunciation.
- Whenever a word is read in a broken manner with pauses between sounds, a special label indicating an intra-word pause needs to be marked at all instances of such pauses.
- The definitive labels for annotations will be provided by WIAI in the detailed transcription 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, WIAI reserves the right to reject the entire batch.

Data Policy

- Wadhwani AI will share anonymized data, stripped of PII, uploaded in an AWS S3 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 of providing or procuring a suitable annotation tool lies with the partner. This tool must meet all the aforementioned needed transcription functionalities.

Formatting & Delivery

Delivering translations in a JSON format as specified by Wadhwani AI

```
{
  "id": 13945,
  "file_name": "audio_student_04195.wav",
  "annotations": [
    {
      "start": "0:00:00.519350",
      "end": "0:00:01.209974",
```

```

    "Transcription": [
      "Boy"
    ],
  },
  {
    "start": "0:00:01.232074",
    "end": "0:00:01.651974",
    "Transcription": [
      "is"
    ],
  },
  . . . .
}
?

```

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/proficient speakers of English and be able to discern different accents of English in speech.

Total Number of Hours to be annotated: **~150 hours**

Sample Reading Tasks which will be recorded for transcription:

Words	fox, hen, lip, bat, cub, pet, run, pot, jam, milk, hand, nest, trip, plug, case
Paragraph	Waterfalls are natural wonders where rivers drop suddenly from a height, creating rushing water and mist. They form when rivers flow over layers of hard rock that resist erosion, while softer rock beneath wears away. Famous waterfalls like Niagara and Ange I Falls attract visitors worldwide. Animals and plants near waterfalls thrive in moist conditions, such as mosses, ferns, and fish that leap upstream. Waterfalls also provide hydroelectric power when engineers build dams nearby. Tourists enjoy their beauty , while scientists study erosion, flow speed, and surrounding ecosystems. Seasonal rainfall changes the size and strength of waterfalls, sometimes turning a trickle into a roaring cascade. Legends and folklore often surround waterfalls, showing their impor tance in culture and imagination. Standing near one, people can feel the force of nature, hear the thunder of falling water, and see rainbows forming in the spray.

Proposals should clearly outline methodology, data policy, team expertise, and pricing for the proposal.

The selected partner is expected to complete the transcription activity within 3 weeks and **submit the first batch of 20 hours within 1 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

S.No	Activity	Date
1	Release of the Request for Proposal (RFP)	13th November 2025
2	Q&A and Clarifications	18th November 2025
3	Last date for submission of the complete proposal	21st November 2025

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 amounting to ~150 hours**
- **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 Audios (**links will expire by 20th November**)

- **Sample 1 – [link](#)**
- **Sample 2 – [link](#)**
- **Sample 3 – [link](#)**
- **Sample 4 – [link](#)**

2. **Financial Proposal**
 - o Detailed cost breakdown
 - 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 English, demonstrable quality check process
- **Preferred:** Prior experience in transcribing student speech

Submission Details

- All proposals to this RFP must be received no later than November 21, **2025**. 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: **English Transcriptions of Student Voice Recordings of Reading - Rajasthan (Name of the partner)**.

VIII. Evaluation Criteria for Technical Proposal

Criteria	Weightage
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 English 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, reference (NGOs).
- NGOs need to provide 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 most recent renewal as a legal entity.
- Preferably 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).

Bidders selected by the LEHS may be asked for their presentations after reviewing technical proposals and negotiations.

If bidders do not hear from LEHS within two weeks of closing of the RFP, consider your proposal not selected. LEHS reserves the right to reject any or all proposals at any given time without giving any justification or response to the bidders.