



The Manager (Project Purchase)

Date: 02.09.2022

Open Tender Reference No: EE/UME/40/2022/DATACOLLECT
GEM NAR ID: GEM/GARPTS/02092022/Y4R3TOA8HH2B

Due Date/Time: 22.09.2022@ 3:00 PM

Dear Sir/Madam,

On behalf of the Indian Institute of Technology Madras, offers are invited for

Speech Data Collection in Indian Languages and Manually Transcription for building various Speech Technologies

conforming to the specifications.

Tender Documents may be downloaded from Central Public Procurement Portal <https://etenders.gov.in/e procure/app>. Aspiring Bidders who have not enrolled / registered in e-procurement should enroll / register before participating through the website <https://etenders.gov.in/e procure/app>. The portal enrolment is free of cost. Bidders are advised to go through instructions provided at **“Help for contractors”**. [Special Instructions to the Contractors/Bidders for the e-submission of the bids online through this eProcurement Portal”]

Bidders can access tender documents on the website (For searching in the NIC site, kindly go to Tender Search option and type ‘IIT’. Thereafter, click on “GO” button to view all IIT Madras tenders). Select the appropriate tender and fill them with all relevant information and submit the completed tender document online on the website <https://etenders.gov.in/e procure/app> as per the schedule attached.

1)	Pre-bid Meeting Details	:	NA
2)	ICSR Vendor Registration	:	<p>Vendor registration code. Vendor registration with IC&SR (IITM) is mandatory for bidders to participate in tenders.</p> <p>** For Vendor Registration & Guidelines, Please follow the website : https://icandsr.iitm.ac.in/vendorportal; Helpdesk: vendorhelpdesk@icsrpis.iitm.ac.in</p>

No manual bids will be accepted. All tender documents including Technical and Financial bids should be submitted in the E-procurement portal.

Last date for receipt of tender	:	22.09.2022 @ 3:00 PM
Date & time of opening of tender	:	23.09.2022 @ 3:00 PM

3. Instructions to the Bidder:

<u>A)</u>	Searching for tender documents	:	<ul style="list-style-type: none"> • There are various search options built in the CPP Portal, to facilitate bidders to search active tenders by several parameters. These parameters could include Tender ID, organization name, location, date, value, etc. There is also an option of advanced search for tenders, wherein the bidders may combine a number of search parameters such as organization name, form of contract, location, date, other keywords etc. to search for a tender published on the CPP Portal. • Once the bidders have selected the tenders they are interested in, they may download the required documents / tender schedules. These tenders can be moved to the respective “My Tender” folder. This would enable the CPP Portal to intimate the bidders through SMS / email in case there is any corrigendum issued to the tender document. • The bidder should make a note of the unique Tender ID assigned to each tender, in case they want to obtain any clarification / help from the Helpdesk.
<u>B)</u>	Assistance to bidders	:	<ul style="list-style-type: none"> • Any queries relating to the tender document and the terms and conditions contained therein should be addressed to the Tender Inviting Authority for a tender or the relevant contact person indicated in the tender. • Any queries relating to the process of online bid submission or queries relating to CPP Portal in general may be directed to the 24x7 CPP Portal Helpdesk. The contact number for the helpdesk is [0120-4200462, 0120-4001002, 0120-4001005]
<u>C)</u>	Enrollment Process to Bidders	:	<p><u>REGISTRATION</u></p> <ul style="list-style-type: none"> • Bidders are required to enroll on the e-Procurement module of the Central Public Procurement Portal URL:https://etenders.gov.in/eprocure/app by clicking on “Online Bidder Enrollment”. Enrollment on the CPP Portal is free of charge. • As part of the enrolment process, the bidders will be required to choose a unique username and assign a password for their accounts. • Bidders are advised to register their valid email address and mobile numbers as part of the registration process. These would be used for any communication from the CPP Portal. • Upon enrolment, the bidders will be required to register their valid Digital Signature Certificate (Class II or Class III Certificates with signing key usage) issued by any Certifying Authority recognized by CCA India (e.g. Sify / TCS / nCode / eMudhra etc.) • Only one valid DSC should be registered by a bidder. Please note that the bidders are responsible to ensure that they do not lend their DSCs to others which may lead to misuse. • Bidder then may log in to the site through the secured log-in by entering their user ID / password and the password of the DSC / eToken. • Possession of a Valid Class II/III Digital Signature Certificate (DSC) in the form of smart card/e-token in the company's name

			<p>is a prerequisite for registration and participating in the bid submission activities through https://etenders.gov.in/eprocure/app</p> <ul style="list-style-type: none"> • Digital Signature Certificates can be obtained from the authorized certifying agencies, details of which are available in the web site https://etenders.gov.in/eprocure/app under the “Information about DSC”.
<u>D)</u>	Preparation of bids	:	<ul style="list-style-type: none"> • Bidder should take into account any corrigendum published on the tender document before submitting their bids. • Please go through the tender advertisement and the tender document carefully to understand the documents required to be submitted as part of the bid. Please note the number of covers in which the bid documents have to be submitted, the number of documents including the names and content of each of the document that need to be submitted. Any deviations from these may lead to rejection of the bid. • Bidder, in advance, should prepare the bid documents to be submitted as indicated in the tender document / schedule and generally shall be in PDF / XLS formats as the case may be. Bid documents may be scanned with 100 dpi with black and white option. • To avoid the time and effort required in uploading the same set of standard documents which are required to be submitted as a part of every bid, a provision of uploading such standard documents (e.g. PAN card copy, GSTIN Details, annual reports, auditor certificates etc.) has been provided to the bidders. Bidders can use “My Documents” area available to them to upload such documents. These documents may be directly submitted from the “My Documents” area while submitting a bid, and need not be uploaded again and again. This will lead to a reduction in the time required for bid submission process.
<u>E)</u>	Submission of bids	:	<ul style="list-style-type: none"> • Bidder should log into the site well in advance for bid submission so that he/she can upload the bid in time i.e. on or before the bid submission date and time. Bidder will be responsible for any delay due to other issues. • The bidder has to digitally sign and upload the required bid documents one by one as indicated in the tender document. • Bidder has to select the bid security declaration. Otherwise, the tender will be summarily rejected. • A standard BOQ format has been provided with the tender document to be filled by all the bidders. Bidders are requested to note that they should necessarily submit their financial bids in the format provided and no other format is acceptable. Bidders are required to download the BOQ file, open it and complete the detail with their respective financial quotes and other details (such as name of the bidder). If the BOQ file is found to be modified by the bidder, the bid will be rejected. • The server time (which is displayed on the bidders’ dashboard) will be considered as the standard time for referencing the deadlines for

		<p>submission of the bids by the bidders, opening of bids etc. The bidders should follow this time during bid submission.</p> <ul style="list-style-type: none"> • The Tender Inviting Authority (TIA) will not be held responsible for any sort of delay or the difficulties faced during the submission of bids online by the bidders due to local issues. • The uploaded tender documents become readable only after the tender opening by the authorized bid openers. • Upon the successful and timely submission of bids, the portal will give a successful bid submission message & a bid summary will be displayed with the bid no. and the date & time of submission of the bid with all other relevant details. • Kindly add scanned PDF of all relevant documents in a single PDF file of compliance sheet. • More information useful for submitting online bids on the CPP Portal may be obtained at: https://etenders.gov.in/eprocure/app. • All tender documents including pre-qualification bid, Technical Bid & Financial Bid should be submitted separately in online CPP portal as per the specified format only. Right is reserved to ignore any tender which fails to comply with the above instructions. No manual bid submission will be entertained.
F)	Marking on Pre-Qualification Bid	<ul style="list-style-type: none"> • The bidder eligibility criteria, technical specification and supply of item for this tender is given in Annexure A. • The Bidders shall go through the specification and submit the Pre-Qualification bid. • The Pre-qualification bid should be submitted in the proforma as per Annexure-B in pdf format only through online (e-tender). No manual submission of bid will be entertained. • The Pre-Qualification bid should have a page-wise heading as “Pre-qualification Bid” and page no. in all pages with seal and signature of authorized signatory. The total no. of pages should be mentioned at the last page of the documents
G)	Marking on Technical Bid	<ul style="list-style-type: none"> • The bidder eligibility criteria, technical specification and supply of item for this tender is given in Annexure A. • The Bidders shall go through the specification and submit the technical bid. • The Technical bid should be submitted in the proforma as per Annexure-C in pdf format only through online (e-tender). No manual submission of bid will be entertained. • The technical bid should have a page-wise heading as “Technical Bid” and page no. in all pages with seal and signature of authorized signatory. The total no. of pages should be mentioned at the last page of the documents. • The technical bid should consist of bidder eligibility criteria details and all technical details along with catalogue/ pamphlet which will give a detailed description of product with technical data sheet so that technical compliance can be verified.

H)	Marking on Price Bid	<ul style="list-style-type: none"> Financial bid (BoQ) should be submitted in the prescribed proforma format as per Annexure-D in xls format through e-tender only. No manual or other form of submission of Financial Bid will not be entertained 																						
4)	<p>EMD: Rs.10,50,000 /- (Rupees Ten Lakhs and Fifty Thousand Only) has to be paid by means of bank transfer. Bank Account Details are given below. Proof of remittance has to be enclosed along with technical tender documents; else the bid will be rejected due to non-enclosure of EMD. This Earnest Money will be returned to the unsuccessful tenderers after the finalization of the tenders. EMD will not carry any interest. EMD will not be waived under any circumstances. EMD is exempted for Micro and Small Enterprises (MSE) as defined in MSE Procurement policy issued by the department of Micro, Small and Medium Enterprise (MSME). Relevant proof of registration of MSE with the competent authority should be submitted. For others, non-submission of EMD will lead to rejection of tender at the opening stage itself.</p> <table border="1" data-bbox="203 646 1419 1226"> <tr> <td>Institution Account Name</td> <td>The Registrar, Indian Institute of Technology - Madras</td> </tr> <tr> <td>Account No</td> <td>2722101003872</td> </tr> <tr> <td>Account Print Name</td> <td>The Registrar, IIT Madras</td> </tr> <tr> <td>IFSC CODE</td> <td>CNRB0002722</td> </tr> <tr> <td>Bank Name (in full)</td> <td>CANARA BANK</td> </tr> <tr> <td>Branch Name</td> <td>IIT -Madras Branch</td> </tr> <tr> <td rowspan="5">Complete Branch Address</td> <td>Canara Bank</td> </tr> <tr> <td>IIT -Madras Branch</td> </tr> <tr> <td>IIT -Madras Campus Post Office</td> </tr> <tr> <td>Sardar Patel Road</td> </tr> <tr> <td>Guindy, Chennai - 600 036</td> </tr> <tr> <td>MICR No.</td> <td>600015085</td> </tr> <tr> <td>Account Type</td> <td>Savings Account</td> </tr> </table>		Institution Account Name	The Registrar, Indian Institute of Technology - Madras	Account No	2722101003872	Account Print Name	The Registrar, IIT Madras	IFSC CODE	CNRB0002722	Bank Name (in full)	CANARA BANK	Branch Name	IIT -Madras Branch	Complete Branch Address	Canara Bank	IIT -Madras Branch	IIT -Madras Campus Post Office	Sardar Patel Road	Guindy, Chennai - 600 036	MICR No.	600015085	Account Type	Savings Account
Institution Account Name	The Registrar, Indian Institute of Technology - Madras																							
Account No	2722101003872																							
Account Print Name	The Registrar, IIT Madras																							
IFSC CODE	CNRB0002722																							
Bank Name (in full)	CANARA BANK																							
Branch Name	IIT -Madras Branch																							
Complete Branch Address	Canara Bank																							
	IIT -Madras Branch																							
	IIT -Madras Campus Post Office																							
	Sardar Patel Road																							
	Guindy, Chennai - 600 036																							
MICR No.	600015085																							
Account Type	Savings Account																							
5)	<p>Performance Security: -</p> <p>The successful bidder should submit Performance Security for an amount of 3% of the value of the contract/supply. The Performance Security may be furnished in the form of an Account Payee DD, FD Receipt in the name of “The Registrar, IIT Madras” from any scheduled commercial bank or Bank Guarantee from any scheduled commercial bank in India. The performance security should be furnished within 14 days from the date of the purchase order.</p> <p>Performance Security in the form of Bank Guarantee: - In case the successful bidder wishes to submit Performance Security in the form of Bank Guarantee, the Bank Guarantee should be routed directly to IIT Madras from the Bank.</p> <p>The Bank Guarantee should remain valid for a period of sixty days beyond the date of completion of all contractual obligations of the supplier including the warranty obligations.</p>																							
6)	<p>Price:</p> <p>The offer/bid should be exclusive of taxes and duties, which will be paid by the purchaser as applicable. However, the percentage of tax & duties should be clearly indicated separately.</p>																							

7)	<p>Tenderer shall submit along with this tender:</p> <ul style="list-style-type: none"> (i) Name and full address of the Banker and their swift code and PAN No. and GSTIN number. (ii) GST registration proof showing registration number, area of registration etc. a) (iii) All of your future correspondences including Invoices should bear the GST No. and Area Code.
8)	<p>Terms of Delivery:</p> <p>The tenderer should clearly indicate the time of Delivery. In case there is any deviation in the delivery schedule, liquidated damages clause will be enforced or penalty for the delayed supply period will be levied.</p>
9)	<p>Period for which the offer will remain open:</p> <p>The Tender shall remain open for acceptance/validity till: 120 days from the date of opening of the tender. However, the day up to which the offer is to remain open being declared closed holiday for the Indian Institute of Technology Madras, the offer shall remain open for acceptance till the next working day.</p>
10)	<p>Risk Purchase Clause</p> <p>In the event of failure of provision of required services within the stipulated delivery schedule, the purchaser has all the right to get services from other sources on the total risk of the supplier under risk purchase clause.</p>
11)	<p>Payment:</p> <p>No Advance payment will be made for Indigenous purchase. Payment will be made 100% after the acceptance of completeness and accuracy of the said report by the Project Coordinator at IPM Cell, IITM.</p>
12)	<p>Acceptance and Rejection:</p> <p>Failure to comply with any of the instructions stated in this document or offering unsatisfactory explanations for non-compliance will likely to lead to rejection of offers.</p> <ul style="list-style-type: none"> (i) I.I.T. Madras has the right to accept the whole or any part of the Tender or portion of the quantity offered or reject it in full without assigning any reason.
13)	<p>Debarment from Bidding:</p> <p>In case of breach of Terms & Conditions, Bidder may be suspended from being eligible for bidding in any contract with the IIT Madras up to 2 Years [as per Rule 151(iii) of GFR] from the date of Tender.</p>
14)	<p>Disputes and Jurisdiction:</p> <p>Settlement of Disputes: Any dispute, controversy or claim arising out of or in connection with this PO including any question regarding its existence, validity, breach or termination, shall in the first instance be attempted to be resolved amicably by both the Parties. If attempts for such amicable resolution fails or no decision is reached within 30 days whichever is earlier, then such disputes shall be settled by arbitration in accordance with the Arbitration and Conciliation Act, 1996. Unless the Parties agree on a sole arbitrator, within 30 days from the receipt of a written request by one Party from the other Party to so agree, the arbitral panel shall comprise of three arbitrators. In that event, the supplier will nominate one arbitrator and the Project Coordinator of IITM shall nominate an arbitrator. The Dean IC&SR will nominate the Presiding Arbitrator of the arbitral tribunal. The arbitration proceeding shall be carried out in English language. The cost of arbitration and fees of the arbitrator(s) shall be shared equally by the Parties. The seat of arbitration shall be at IC&SR IIT Madras, Chennai.</p> <ul style="list-style-type: none"> a. The Applicable Law: The Purchase Order shall be construed, interpreted and governed by the Laws of India. Court at Chennai shall have exclusive jurisdiction subject to the arbitration

	<p>clause.</p> <p>Any legal disputes arising out of any breach of contract pertaining to this tender shall be settled in the court of competent jurisdiction located within the city of Chennai in Tamil Nadu.</p>
15)	<p>Force Majeure: The Supplier shall not be liable for forfeiture of its performance security, liquidated damages or termination for default, if and to the extent that, it's delay in performance or other failure to perform its obligations under the Contract is the result of an event of Force Majeure.</p> <p>For purposes of this Clause, "Force Majeure" means an event beyond the control of the Supplier and not involving the Supplier's fault or negligence and not foreseeable. Such events may include, but are not limited to, acts of the Purchaser either in its sovereign or contractual capacity, wars or revolutions, fires, floods, epidemics, quarantine restrictions and freight embargoes.</p> <p>b. If a Force Majeure situation arises, the Supplier shall promptly notify the Purchaser in writing of such conditions and the cause thereof. Unless otherwise directed by the Purchaser in writing, the Supplier shall continue to perform its obligations under the Contract as far as is reasonably practical, and shall seek all reasonable alternative means for performance not prevented by the Force Majeure event.</p>
16)	<p>Preference to "class I Local Suppliers": preference will be given to "class 1 local suppliers" (subject to class -I local supplier's quoted price falling within the margin of purchase preference) as per public procurement (preference to make in India) order 2017 .O.M No P- 45021/2/2017 – pp(BE - 11) dt 04/06/2020 subject to the conditions that the "class 1 Local Supplier" should agree to supply goods / provide service at L1 rate and furnish a certificate with the technical bid document that the goods/service provided by them consists local content equal to or more than 50%.(certificate from Chartered Accountant in case value of contract exceeds Rs 10 crore).</p> <p>**Note: Local content percentage to be calculated in accordance with the definition provided at clause 2 of revised public procurement preference to Make in India Policy vide GoI Order no. P-45021/2/2017-PP (B.E.-II) dated 15.06.2017 (subsequently revised vide orders dated 28.05.2018, 29.05.2019and 04.06.2020) MOCI order No. 45021/2/2017-PP (BE II) Dt.16th September 2020 & P- 45021/102/2019-BE-II-Part(1) (E-50310) Dt.4th March 2021</p>

17	<p>Preparation of Tender: The bidders should submit the bids in three bid system as detailed below.</p> <p>Cover I _Pre-Qualification Bid</p> <p>The Pre- Qualification Bid for in the Bidder Eligibility criteria I & II along with all relevant proof needs to be submitted in the proforma given in Annexure – B.</p> <p>The proof attached should be indexed and page numbers of the attached proof should be clearly mentioned in the Pre-Qualification Bid format.</p> <p>Cover II _Technical Bid</p> <p>The technical bid should consist of technical specification compliance sheet as per Annexure-C.</p> <p>Cover III _Price Bid</p> <p>The price bid should be submitted in excel format (BoQ) as per the proforma (Annexure D) uploaded in the e-Tender web site. The Quoted price should be for the services offered and inclusive of all cost and statutory levies at IIT Madras. The vendor should Quote the cost on Per Hour Basis as per the BPQ in Annexure D.</p>
18	<p>Bidder Eligibility Criteria I:</p> <ul style="list-style-type: none"> a) As per the Government of India Order, only “Class - I Local Suppliers” and “Class - II Local Suppliers” can participate in this tender. Declaration to be submitted as per Annexure - E b) Bidder should confirm their acceptance that they comply with the provisions with report to “Guidelines for eligibility of a bidder from a country which shares a land border with India as detailed at Annexure-F. The bidder should submit Certificate for “Bidder from/ Not from Country sharing Land border with India & Registration of Bidder with Competent Authority” as per Order of DoE F.No.6/18/2019-PPD dated 23.07.2020 as mentioned. c) Proof of Submission of EMD. d) Consent Form as per Annexure -G e) Declaration for Blacklisting as per Annexure - H <p>Bidder Eligibility Criteria II:</p> <ul style="list-style-type: none"> a) The Bidder’s firm should have existed for a minimum of 4 years. (Enclose Company Registration Certificate) b) The bidder should have done at least 1000 hours of speech data Collection and 1500 hours of transcription to reputed organizations. PO copies or work completion certificates should be submitted as a proof of the same.

	<p>c) The bidder should have experience collecting transcribed Speech data for at least 3 of the following Indian languages (Indian English, Hindi, Malayalam, Kannada, Marathi, Gujarati, Punjabi, Assamese, Bengali, Odia, Manipuri and Tamil), including one From South India (Tamilnadu, Karnataka, Kerala, Andhra and Telangana) and one from other states in North India. PO copies or work completion certificates should be submitted as a proof of the same.</p> <p>d) Bidders should have experience of transcribing at least 100 hours of speech data per language in at least 3 Indian languages which may include Indian English. PO copies or work completion certificates should be submitted as a proof of the same.</p>
--	--

19)	<p>Evaluation of Bids</p> <p>Bid evaluation will take place in three stages.</p> <p>Stage I Pre-Qualification Bid evaluation</p> <p>The Pre- Qualification Bid of the bidders will be evaluated and the bidders who have qualified all the criteria stipulated in bidder eligibility criteria I & II will alone be selected and their technical bids will be opened.</p> <p>Stage II Technical Bid evaluation</p> <ol style="list-style-type: none"> All bidders who qualify in the Stage I should transcribe 1 hour of data (covering all or subset of 10 languages), which will be evaluated for transcription correctness. Upon publishing the Pre-Qualification result IITM will forward the Data for transcription to bidders through email with time duration of 3 days for submission of transcript data. Stage 1 Qualified bidders need to make a power point presentation before a EPC Committee. Date for presentation will be informed to bidders through email. The technical submission, Data transcription and technical presentation together will be considered for technical bid evaluation. <p>The total marks for technical bid will be for 100 marks, out of which the bidders should score minimum 70 marks to qualify for opening of their financial bid. The Financial bids of those bidders who have scored less than 70 marks will not be opened.</p> <p>The detailed scoring system for arriving at the total technical score of the bidders who have qualified Stage I are:</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">S.No.</th> <th style="width: 70%;">Criteria</th> <th style="width: 20%;">Maximum marks</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">1</td> <td>Has the Bidder's firm collected data for any publicly listed companies /Academics or research institutions (If yes, enclose proof for the same)</td> <td style="text-align: center;">10</td> </tr> </tbody> </table>	S.No.	Criteria	Maximum marks	1	Has the Bidder's firm collected data for any publicly listed companies /Academics or research institutions (If yes, enclose proof for the same)	10
S.No.	Criteria	Maximum marks					
1	Has the Bidder's firm collected data for any publicly listed companies /Academics or research institutions (If yes, enclose proof for the same)	10					

	2	Does the Bidder's firm have collected more than 1000 hours of data? (Marks will be given based on the amount of data collected by the bidder in the past, enclose proof for the same)	10
	3	Does the Bidder's firm have experience collecting data for speaker-based and language-based labeling? (If yes, enclose proof for the same)	10
	4	Does the Bidder's firm have experience Collecting and Transcribing data in multiple Indian languages? (Marks will be given based on the Bidder's experience in collecting data in various languages, enclose proof for the same)	10
	5	The Bidder should provide a 2-page write-up on their plan of action to collect and transcribe data, with details on the quality check and manpower/ organization structure for the scope of work given in this tender.	15
	6	Bidder should share 15 minutes of conversational speech data per language collected by them in at least 4 Indian languages (at least One from South India(Tamilnadu, Karnataka, Kerala, Andhra and Telangana) and one from Other states in North India. To be submitted in USB drive.	10
	7	The Bidder should transcribe 1 hour of data (covering all or subset of 10 languages), which will be evaluated for transcription correctness. IITM will forward the Data to the Stage 1 Qualified Bidders only.	10
	8	The Bidder should prepare a presentation detailing their plan to Collect data, Workflow, Existing Manpower, Diversity, Presence in various regions. etc. Detailed PPT document to be submitted along with Technical bid.	25
	<p>Stage III: Price Bid Evaluation</p> <p>The price bid evaluation will be based on the rate quoted by the bidder for the Collection of Speech Data in Indian Languages and Manually Transcription for building various Speech Technologies. to arrival of Lowest Bid (L1) value.</p>		
20)	<p>Selection of successful bidder and Award of Order</p> <p>The successful bidder will be selected based on assessment of skills, experience, and understanding/analysis of the project scope as per Technical Bid.</p>		
21)	<p>Apportion of work</p> <p>1. The next 3 Bidders following the L1 will be offered a chance to match the L1 price and will be selected as successful contractors.</p>		

	<p>2. If either or all L2, L3 and L4 have not agreed to match the price, then the offer will be extended to the next bidders with Least cost and this process can be continued until we select 4 Successful contractors for award of work.</p> <p>3. Estimate requirement of Data Collection - 5000 hrs.</p> <p>4. Minimum of 750 hrs. will be apportioned to each vendor and balance quantity will be apportioned later based on quality and timeline of completion of work by selected bidders.</p>																				
22)	<p>Period of Service: The Collection of Speech Data in Indian Languages and Manually Transcription for building various Speech Technologies will be initially awarded for an estimated time 5000 Hrs. Additional requirement may arise above 5000 hours and the successful bidders shall extend the service on the same agreed rate.</p>																				
23)	<p>Delivery and Payment Schedule:</p> <p>The timeline for delivery and payment is as below.</p> <ol style="list-style-type: none"> 1. At the end of every month the data will be evaluated for the quality check by IITM. 2. The maximum acceptable data Errors is 6-8 %. 3. No Payments will be made for the data with more than 15% Errors. 4. The total hours of work allocated, and completion will be inclusive of Data submitted with Errors. <table border="1"> <thead> <tr> <th>Sl. No</th> <th>Duration</th> <th>Amount of Work</th> <th>Payment</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1st Month from Date of PO.</td> <td>Minimum 15% of overall data collection covering 4 languages (2 North & 2 South Indian)</td> <td>15% of the total cost</td> </tr> <tr> <td>2</td> <td>2nd Month from Date of PO.</td> <td>Minimum 15% of overall data collection (inclusive of Month 1) covering all languages</td> <td>15% of the total cost</td> </tr> <tr> <td>3</td> <td>3rd Month from Date of PO.</td> <td>Minimum 35% of overall data collection covering all languages</td> <td>35% of the total cost</td> </tr> <tr> <td>4</td> <td>4th Month from Date of PO.</td> <td>Minimum 35% of overall data collection covering all languages</td> <td>35% of the total cost</td> </tr> </tbody> </table>	Sl. No	Duration	Amount of Work	Payment	1	1 st Month from Date of PO.	Minimum 15% of overall data collection covering 4 languages (2 North & 2 South Indian)	15% of the total cost	2	2 nd Month from Date of PO.	Minimum 15% of overall data collection (inclusive of Month 1) covering all languages	15% of the total cost	3	3 rd Month from Date of PO.	Minimum 35% of overall data collection covering all languages	35% of the total cost	4	4 th Month from Date of PO.	Minimum 35% of overall data collection covering all languages	35% of the total cost
Sl. No	Duration	Amount of Work	Payment																		
1	1 st Month from Date of PO.	Minimum 15% of overall data collection covering 4 languages (2 North & 2 South Indian)	15% of the total cost																		
2	2 nd Month from Date of PO.	Minimum 15% of overall data collection (inclusive of Month 1) covering all languages	15% of the total cost																		
3	3 rd Month from Date of PO.	Minimum 35% of overall data collection covering all languages	35% of the total cost																		
4	4 th Month from Date of PO.	Minimum 35% of overall data collection covering all languages	35% of the total cost																		
24)	<p>All information including selection and rejection of technical or financial bids of the prospective bidders will be communicated through e-Tender portal. In terms of Rule 173(iv) of General Financial Rule 2017, the bidder shall be at liberty to question the bidding conditions, bidding process and/or rejection of bids.</p>																				

25)	The tenderer shall certify that the tender document submitted by him / her are of the same replica of the tender document as published by IIT Madras and no corrections, additions and alterations made to the same. If any deviation found in the same at any stage and date, the bid / contract will be rejected / terminated and actions will be initiated as per the terms and conditions of the contract.
26)	The bidders will not be entertained to participate in opening of Bids, since the tender is e-tender, the opening of the bids may be checked using the respective logins of the bidders.

ACKNOWLEDGEMENT

It is hereby acknowledged that I/We have gone through all the points listed under “Specification, Guidelines, Terms and Conditions” of tender document. I/We totally understand the terms and conditions and agree to abide by the same.

**SIGNATURE OF TENDERER ALONG WITH SEAL OF
THE COMPANY WITH DATE**

Scope of the Work.

Speech Data Collection in Indian Languages and Manually Transcription for building various Speech Technologies

Tender No. EE/UME/40/2022/DATACOLLECT

Deliverables

1. **Bidders need to Mandatorily quote and work on all 10 languages(Hindi, Malayalam, Kannada, Marathi, Gujarati, Punjabi, Assamese, Bengali, Odia and Tamil).**
2. **Bidders need to collect Minimum 50 hrs of Data per language and for each 50 Hrs of Data**
 - a) **The minimum word vocabulary should be 20,000 words,**
 - b) **Minimum No of speakers should be 300.**
 - c) **And a Minimum time of 10 mins per speaker.**
3. **Mode Of Delivery: In electronic form.**

Data Collection Guidelines:

The data collected **should be predominantly conversational in nature:**

1) Extempore Speech (monolingual)

- a) Collected data should contain the speech of about 1500 speakers [5min/speaker]
- b) Speaker's age should be 18 to 60.
- c) Number of males and females should be balanced in the Data.
- d) Collected data should contain different domains (weather, different types of news, entertainment, health, agriculture, education, jobs, finance).
- e) BPO type of data – this can be a mockup or it could be actual data (if we can get this from any agencies) – also mostly monolingual (at best a few words in English) – Aadhar, passport, ration card, services – banking, repairs.
- f) Data should be collected from different major dialects (**at least about 4 major dialects** for every language).

2) Conversational Speech

- a) 2-6 people chatting about a topic of interest (for example sports, news, weather, entertainment, politics, business, everyday problems like public transport, e-governance, government schemes like aadhar, dhan yojana, ayush bharat etc).

- b) Speakers should prepare a little before the conversation. Many times, in the whole conversation they keep saying: “you tell”, “what’s up”, “okay” “what shall i say’ and no real conversation. So please ask them to have some points in their mind before the start of the conversation.
- c) At least **10 different domains** must be chosen.
- d) Each conversation could last about 3-5mins between different speakers.
- e) The number of males and females should be more or less equal in multiple sessions.
- f) About a total of **2500-3000 conversations per language** must be collected with about **1000-1500** speakers.
- g) The conversations can be made across various smartphones using voice/data calls.
- h) In addition, the data should be collected from various places like a railway station, an airport, or driving in a car/vehicle
- i) The conversations should also vary over networks 2G-4G, (5G if possible). Apparently, the coding schemes are different for all the networks. You can collect this from different handsets. Sometimes different varieties of Gs are available in the bowels of India. We need to account for this.
- j) **About 500 conversations** from f) above should be multilingual both code mixed and code switched. Prefer if it is **MIX OF English, and another local language**. (e.g. Tamil+English (Tanglish) or Hindi+English (Hinglish) etc)

3) Some general rules for both 1) and 2)

- a) The data must be transcribed verbatim, and time stamps must be accurate to **about 300-500ms**.
- b) Declassification of data – the identity of the speaker must be changed but should be uniquely named.
- c) Speaker information such as age and gender to be included while naming a speaker(speaker_ID) in an utterance/segment of speech data.

Data Labeling Guidelines:

1. Transcribe “verbatim.” Do not correct for grammatical errors.
2. Use a dictionary for spelling when in doubt.
3. Avoid abbreviations e.g Ft George – transcribe as Fort George, % as percent (प्रतीशत, சதவிதம்). Similarly for rupees, dollars, paise etc.
4. Terminate each utterance by a . | or whatever symbol is used in a particular language.
5. If there are multiple speakers in a conversations, then:

<Speaker1>: verbatim of transcript of Speaker 1 in UTF8 of the corresponding language

<Speaker2>: verbatim of transcript of Speaker 2 in UTF8 of the corresponding language

You may also additional tags like <Speaker x angry> <Speaker x sad> etc to

Indicate the emotion if you can identify in the speech.

6. If there are multiple non-speech sounds, and you can identify them mark them in square brackets (as indicated above) [laughter][coughing]

7. If multiple languages are used, and you do not know a language that is being spoken, then mark this.

For example, if the speaker spoke in Tamil and then in Bangla:

Then transcribe this, if you know the script of the language,

நான் ஊருக்கு செல்கிறேன். அமி ஋ஃபரி நா

If you do not know the script but are definite about the language then transcribe as:

நான் ஊருக்கு செல்கிறேன். <start time> <Bangla> <end time>

If you do not know the language or the script then transcribe as:

நான் ஊருக்கு செல்கிறேன். <start time> <unknown_language> <end time>

Please do mark the time where a different language was spoken, and ended especially if you can not transcribe it.

If the language is English, transcribe into English. Timing information is not required if you can transcribe it in the given language.

8. Wherever possible indicate when two people spoke simultaneously

Bye, bye # if you are able to make out the words spoken. Or alternatively just mark it as <start time> <multi_speaker> <end time> – please mark the start and end time if you are not able to transcribe.

9. **Audio file format** : WAV format (‘.wav’) Sampling Frequency=16Khz, 16-bit Linear PCM and channel = 1

10. Any whole audio file can have multiple speakers, but please ensure that an audio segment does not.

11. Audio segments should not be more than 15 seconds.

12. Long audio files need not be physically split into multiple smaller segments, instead make use of the kaldia format segments file to indicate the start_time and end_time of the corresponding audio segment.

PRE-QUALIFICATION BID PROFORMA
Tender No. EE/UME/40/2022/DATACOLLECT

Item Name: Speech Data Collection in Indian Languages and Manually Transcription for building various Speech Technologies

Bidder Eligibility Criteria:

I	Bidder Eligibility Criteria-I (Public Procurement – Preference to Make in India)	Class I / Class II	Local Content value	Reference, Page No.
a	Only 'Class-I local suppliers' and 'Class-II local suppliers', as defined under DIPP, MoCI Order No. P-45021/2/2017-PP (BE II) dated 16 th September 2020 and other subsequent orders issued therein. Declaration to be submitted as per Annexure - E			
b	Bidder should confirm their acceptance that they comply with the provisions with report to “Guidelines for eligibility of a bidder from a country which shares a land border with India as detailed at Annexure-F. The bidder should submit Certificate for “Bidder from/ Not from Country sharing Land border with India & Registration of Bidder with Competent Authority” as per Order of DoE F.No.6/18/2019-PPD dated 23.07.2020 as mentioned.			
c	Proof of Submission of EMD.			
d	Consent Form as per Annexure -G			
e	Declaration for Blacklisting as per Annexure - H			
II	Bidder Eligibility Criteria-II	Compliance (Yes/No)	Reference Page No.	Remarks, If any
a	The Bidder’s firm should have existed for a minimum of 4 years. (Enclose Company Registration Certificate)			
b	The bidder should have done at least 1000 hours of speech data Collection and 1500 hours of transcription to reputed organizations. PO copies or work completion certificates should be submitted as a proof of the same.			
c	The bidder should have experience collecting transcribed Speech data for at least 3 of the following Indian languages (Indian English, Hindi, Malayalam, Kannada, Marathi, Gujarati, Punjabi, Assamese, Bengali, Odia, Manipuri and Tamil), including one From South India(Tamilnadu, Karnataka, Kerala, Andhra and Telangana) and one from Other states in North India. PO copies or work completion certificates should be submitted as a proof of the same.			
d	Bidders should have experience of transcribing at least 100 hours of speech data per language in at least 3 Indian languages which may include Indian English. PO copies or work completion certificates should be submitted as a proof of the same.			

TECHNICAL BID PROFORMA
Tender No. EE/UME/40/2022/DATACOLLECT

Item Name: Speech Data Collection in Indian Languages and Manually Transcription for building various Speech Technologies

The Bidders should submit a detailed proposal for the scope of works of this tender covering the proposed,

S.NO	TECHNICAL SPECIFICATION	COMPLIED/NOT COMPLIED	REMARKS
1	Has the Bidder's firm collected data for any publicly listed companies /Academics or research institutions (If yes, enclose proof for the same)		
2	Does the Bidder's firm have collected more than 1000 hours of data? (Marks will be given based on the amount of data collected by the bidder in the past, enclose proof for the same)		
3	Does the Bidder's firm have experience collecting data for speaker-based and language-based labeling? (If yes, enclose proof for the same)		
4	Does the Bidder's firm have experience Collecting and Transcribing data in multiple Indian languages? (Marks will be given based on the Bidder's experience in collecting data in various languages, enclose proof for the same)		
5	The Bidder should provide a 2-page write-up on their plan of action to collect and transcribe data, with details on the quality check and manpower/ organization structure for the scope of work given in this tender.		
6	Bidder should share 15 minutes of conversational speech data per language collected by them in at least 4 Indian languages (at least One from South India(Tamilnadu, Karnataka, Kerala, Andhra and Telangana) and one from Other states in North India. To be submitted in USB drive.		
7	The Bidder should transcribe 1 hour of data (covering all or subset of 10 languages), which will be evaluated for transcription correctness. IITM will forward the Data to the Stage 1 Qualified Bidders only.		
8	The Bidder should prepare a presentation detailing their plan to Collect data, Workflow, Existing Manpower, Diversity, Presence in various regions. etc. Detailed PPT document to be submitted along with Technical bid.		

**SIGNATURE OF BIDDER ALONG WITH
SEAL OF THE COMPANY WITH DATE**

FINANCIAL BID (PROFORMA) - BILL OF QUANTITIES (BOQ)

**Item Name: Speech Data Collection in Indian Languages and Manually Transcription for building various
Speech Technologies
Tender No. EE/UME/40/2022/DATACOLLECT**

It. No	Description of work	Quantity	Units	Basic Rate/Hour in INR	GST %	Total Amount with taxes in INR
1	Speech Data Collection in Indian Languages and Manually Transcription for building various Speech Technologies	1	Hrs.			
	Grand Total					

Total Amount Rupees in words _____

FORMAT FOR AFFIDAVIT OF SELF-CERTIFICATION UNDER PREFERENCE TO MAKE IN INDIA – PER ITEM

Tender Reference Number:

Name of the item / Service:

Date: _____

I/We _____ S/o, D/o, W/o, _____
Resident of _____

Hereby solemnly affirm and declare as under:

That I will agree to abide by the terms and conditions of the Public Procurement (Preference to Make in India) Policy vide GoI Order no. P-45021/2/2017-PP (B.E.-II) dated 15.06.2017 (subsequently revised vide orders dated 28.05.2018, 29.05.2019 and 04.06.2020) MOCI order No. 45021/2/2017-PP (BE II) Dt.16th September 2020 & P- 45021/102/2019-BE-II-Part (1) (E-50310) Dt.4th March 2021 and any subsequent modifications/Amendments, if any and

That the local content for all inputs which constitute the said item/service/work has been verified by me and I am responsible for the correctness of the claims made therein.

Tick (✓) and Fill the Appropriate Category	
<input type="checkbox"/>	I/We _____ [name of the supplier] hereby confirm in respect of quoted items that Local Content is equal to or more than 50% and come under “Class-I Local Supplier” category.
<input type="checkbox"/>	I/We _____ [name of the supplier] hereby confirm in respect of quoted items that Local Content is equal to 20% but less than 50% and come under “Class-II Local Supplier” category.

- The details of the location (s) at which the local value addition is made and the proportionate value of local content in percentage

Address _____ Percentage of Local content: _____ %

For and on behalf of (Name of firm/entity)

Authorized signatory (To be duly authorized by the Board of Directors)

<Insert Name, Designation and Contact No.>

[Note: In case of procurement for a value in excess of Rs. 10 Crores, the bidders shall provide this certificate from statutory auditor or cost auditor of the company (in the case of companies) or from a practicing cost accountant or practicing chartered accountant (in respect of suppliers other than companies) giving the percentage of local content.]

This letter should be on the letterhead of the quoting firm and should be signed by a competent authority. Non-submission of this will lead to Disqualification of bids.

(To be given on the letter head of the bidder)

Tender Ref. No. _____

Dated: _____

CERTIFICATE

(Bidders from India)

I have read the clause regarding restrictions on procurement from a bidder of a country which shares a land border with India and hereby certify that I am not from such a country.

OR (*whichever is applicable*)

(Bidders from Country which shares a land border with India)

I have read the clause regarding restrictions on procurement from a bidder of a country which shares a land border with India and hereby certify that I from _____ (Name of Country) and has been registered with the Competent Authority. I also certify that I fulfil all the requirements in this regard and is eligible to be considered. *(Copy/ evidence of valid registration by the Competent Authority is to be attached)*

Place:

Date:

Signature of the Tenderer
Name & Address of the
Tenderer with Office Stamp

(To be given on the letter head of the bidder)

Tender Ref. No. _____

Dated: _____

TO WHOMSOEVER IT MAY CONCERN

In response of IIT Madras’s Bid for the “Call For Collaborative Participation in Collection of Speech Data in Indian Languages and Manual Transcription for building various Speech Technologies”, I _____ (Representative of the Company/Organization) on behalf of _____ (Name of the company/organization), bearing CIN. _____ (preferred), having registered office at _____ do hereby give my unconditional irrevocable consent to IIT Madras, to collect, retain and use in various ways, the transcribed speech data and/or audio recorded on _____ (date of recording).

Further, in exchange of sufficient consideration, _____(name of the organization/company) do unconditionally irrevocably assigns all its rights and/or ownership in the transcribed speech data, recorded video and audio and provide a lifetime access to IIT Madras.

_____ (name of the company/organization) shall be personally responsible for the validity of the rights and ownership of the data collected and shall hold IITM harmless from any complaint, infringement proceedings, and/or any claim made by a third-party.

_____ (name of the company/organization) do understand that IIT Madras can lawfully use the transcribed speech data, recorded video and audio.

Signature:

Name:

Place :

Date:

Self-Declaration that the Service Provider has not been blacklisted

(To be given on the letter head of the bidder)

I

.....

..... S/o R/o

..... police station District

..... Director/ partner/ sole proprietor (Strike out whichever is not applicable) of

..... (Firm or Company) do hereby declare and solemnly affirm:

- i. That the Firm has not been Blacklisted or declared insolvent by any of the Union or State Government / Organization.
- ii. That none of the individual / firm / Company Blacklisted or any partners or shareholder thereof has any connection directly or indirectly with or has any subsistence interest in the deponent business / firm company.
- iii. That neither the Firm nor any of its partner has been involved / convicted in any criminal case / economic offence nor any criminal case / economic offence is pending against firm or any partner of the Firm before any Court of Law / Police.

Place:

Signature of the Tenderer

Date:

Name & Address of the Tenderer with Office Stamp