

F. No. I/8/2024-CC-MOSPI
Government of India
Ministry of Statistics and Programme Implementation
Data Informatics and Innovation Division

Subject: Request for Proposal for Problem Statements under Data Innovation Lab Scheme of Ministry of Statistics and Programme Implementation

The Data Innovation Lab component of Capacity Development (CD) Scheme has been operationalised vide Notification No. I-1101/1/2019-CC-MoSPI dated 23rd July 2024 and the same is available under announcement section of the website of the ministry (https://www.mospi.gov.in/sites/default/files/announcements/DI_lab_guidelines.pdf).

2. Data Innovation Lab will create an ecosystem for experimentation, offering of new ideas and their Proof-of-Concept through wider participation of amongst others, entrepreneurs, start-ups, individual researchers and academic-research organisations. It will promote innovation, adoption of Information Technology in the field of official statistics, including official statistics related methodology and address the challenges of the National Statistical System (NSS) of MoSPI. Technologically Innovative solution through this scheme component will strengthen National Statistical System (NSS) which envisages strategic partnerships with institutes/organizations of repute for their participation in the ecosystem.

3. The request for proposals along with Terms of Reference has been advertised online including on the Central Public Procurement Portal [<https://eprocure.gov.in/eprocure/app/>] (TenderID:2024_MOSPI_831889_1) and website of MoSPI (https://www.mospi.gov.in/sites/default/files/RFP_DI_LAB.pdf). Solutions/ proposals shall also be received online through Central Public Procurement Portal in the prescribed format.

4. The Forwarding letter with proposal may be addressed to

Ms. Madhura Roy
Director, MoSPI
Email: madhura.roy@gov.in

Yours sincerely



Tapan Das
Head of Office

Data Informatics and Innovation Division

Date: 22/x/24

RFP - Reference No. I/8/2024-CC-MoSPI

REQUEST FOR PROPOSAL (RFP)

**AI-enabled intelligent search solutions for documents
and**

AI-based Legacy Data Extraction and Processing Tool

Under

Data Innovation Lab

Ministry of Statistics & Programme Implementation (MoSPI)

**Government of India
Ministry of Statistics & Programme Implementation
Khurshid Lal Bhawan, Janpath
New Delhi-110001**

Important Dates

Sr. No	Event	Deadline
1.	Publication of RFP document	22/10/2024
2.	Bid download start	22/10/2024
3.	Pre-Bid Meeting	04/11/2024; 1100 Hrs
4.	Last date and time for bid Submission	12/11/2024; 1800 Hrs
5.	Date and time of opening of bids	13/11/2024; 1100 Hrs
6.	Date of time of opening of Financial Bids	To be announced
7.	Nodal officer	Ms Madhura Roy Director, DIID madhura.roy@gov.in

Fact Sheet

Sl. No.	Key Information	Details
1.	Assignment Title	Request for Proposal for AI-enabled intelligent search solutions for documents and AI-based Legacy Data Extraction and Processing Tool
2.	Purchaser	DIID, Ministry of Statistics and Program Implementation (MOSPI), Government of India
3.	Location	New Delhi, India
Bid Submissions and Evaluation		
4.	Bid Submissions	E-procurement submission at https://eprocure.gov.in/eprocure/app
5.	Technical Bid	E-form for Technical bid through the https://eprocure.gov.in/eprocure/app
7.	Financial Bid	E-form for Financial bid through the https://eprocure.gov.in/eprocure/app
Bid Conditions		
8.	Method of Selection / Bid Parameter	<p>The evaluation of proposals shall be on the principle of Quality Cum Cost Based Selection (QCBS) based on the final weighted score.</p> <ul style="list-style-type: none"> ● Only the Bidders matching the pre-qualification criteria will be selected for further evaluation ● The technical proposal shall be evaluated based on technical evaluation as mentioned under “Technical Proposal Evaluation Criteria”. Each responsive Proposal will be given a technical score(st). Technical weightage (Tw) will be 70%

		<ul style="list-style-type: none"> The proposal with the lowest cost (Fm) shall be given financial score (Fs) of 100 points. The financial scores of other proposals shall be computed as follows:- $Fs = 100 \times Fm/F$ <p>Where F=amount of Financial Proposal</p> <p>Financial Weightage (Fw) 30%</p> <ul style="list-style-type: none"> Combined Quality and Cost Evaluation- The total score shall be obtained by weighting the combined quality/technical and cost scores and adding them, as follows:- $S = St \times Tw + Fs \times Fw$
9.	EMD/Bid Security	Rs. 50,000/- only (INR Fifty Thousand only) in the form of Demand Draft.
10.	Validity for EMD/Bid Security	90 days from bid submission date.
11.	Performance Security	5 % of the Cost discovered through tender process
12	Tender Processing fee	Nil
13	Tender Cost	Can be downloaded from https://eprocure.gov.in/eprocure/app free of cost.
14	Availability of RFP	Free Distribution of the RFP Document on www.mospi.gov.in website & https://eprocure.gov.in/eprocure/app

15	Pre-bid meeting	4th November 2024 1100 Hrs at Conference hall, DIID, East Block 10 RK Puram, New Delhi 110066
16.	Due Date for Submission of Bids	Upto 18:00 Hrs (IST) on 12 November, 2024
17.	Date and Time of Opening of Bids	13 November, 2024, 11:00 Hrs or any other date as intimated by DIID
18.	Technical Presentation of Bidders	To be intimated later.
19.	Opening of Price/Financial Bid	To be intimated later.
20.	Issue of LoA	To be intimated later.
21.	Language	English
22.	Assignment Duration	Duration of assignment will be for 8 months from the date of issuance of work order .

The contract with detailed terms and conditions will be signed while awarding the contract.

REQUEST FOR PROPOSAL

Data Innovation Lab:

Data Innovation Lab will create an ecosystem for experimentation, offering of new ideas and their Proof-of-Concept through wider participation of amongst others, entrepreneurs, start-ups, individual researchers and academic-research organisations. It will promote innovation, adoption of Information Technology in the field of official statistics, including official statistics related methodology and address the challenges of the National Statistical System (NSS) of MoSPI. Technologically Innovative solution through this scheme component will strengthen National Statistical System (NSS) which envisages strategic partnerships with institutes/organizations of repute for their participation in the ecosystem. Participation is called for submission of solutions to problem statements attached at Annexure-IA & IB.

Expected Outcome:

Selected entity will execute the project to offer a methodological cum technological solution to the problem statements as a prototype proven by a pilot. The details are provided in Annexure-IIA & IIB (Terms of Reference)

Eligibility Criteria:

- **Start-up companies:** Start-up qualification will be decided according to the latest notification by the Department of Industrial Policy and Promotion (DIPP).
- **Autonomous Bodies:** Autonomous bodies, including public sector organizations, non-profit organizations, research institutions / universities and other academic institutions, are eligible to participate.
- **Others:** Indian Faculty members/student from institutes/ universities of repute established in India under appropriate legal provisions, individually or collaboratively can participate. Other organizations and institutes from the relevant field/Individuals with proven credence in the field of Data science/technology working with Indian start up/ Companies can also participate individually or as team.

Note:

1. Only Indian Citizens/Companies/Startups/ other entities as indicated above are eligible to participate.
2. To encourage participation from across academia and industry, participating teams are not mandatorily required to be registered Indian Companies/startups to apply for the challenge. However, teams shortlisted at the Ideation Stage will be required to register themselves as Indian Startup/Company and submit relevant proof at the stage of submission of the Prototype to facilitate contract signing.
3. If individual innovators are associated with any company, they should provide a No Objection Certificate (NOC) from their employer, stating that the company will have no rights to the financial assistance and/or intellectual property rights (IPR). Additionally, individuals should inform their employer of their participation in this RFP.
4. There is no bar on collaboration amongst institutions, they will identify a project lead amongst them and contract will be signed with the organization of the Team Lead. It may be noted that an individual

researcher from an institution (e.g., a university) will be treated as institution for the purpose of this activity as s/he will be representing the University and shall sign the contract after taking NOC /approval from the University.

Implementation Stages:

- **Idea Stage:** An idea for building a product/service, that addresses the problem areas listed under Annexure 1 will be submitted as per the submission form. Submissions will be evaluated by a multidisciplinary Selection Committee.
- **Solution development and testing Stage:** In this stage, shortlisted teams will develop prototypes from ideas, focusing on optimizing algorithms, improving interfaces, and adding features. This phase includes a mid-phase evaluation to assess progress and provide feedback on development and integration capabilities, leading to the submission of refined solutions as proof of concept. This phase will conclude with submission of Beta version as explained in annexure-IIA & IIB.
- **Finalization of Solution & Deployment(Optional):** On the request of MoSPI and based on mutually agreed terms and conditions, the selected entity who has developed the solution as PoC will prepare the detailed mainstreaming plan for the solution and Governing Committee will decide on continuity of the proposal and financial support, if any, required for the same. However, MoSPI, may also decide on continuity of the proposed solution at its own without having any formal engagement with the proposer of the solution

Payment Term:

- **10 percent or Rs. 50000/-** whichever is less on submission of an inception report detailing outlining the Proof of Concept of the solution including methodology and implementation mechanism
- **20 percent** on submission and approval of alpha release along with documentation. At this stage total 20 percent of the financial assistance is proposed to be released. However, if only Rs. 50000 has been released initially instead of 10 percent then accordingly adjustments will be made in the released amount so that a total of 20 percent released once alpha version is approved.
- **70 percent** on approval of Beta release along with code base and documentation
- **10 percent** on successful completion of post development support stage as per proposal submitted by the solution provider or decided by the Selection Committee.

Note: MoSPI will issue a Letter of Approval (LoA) with detailed terms and conditions and also sanction order for release of financial assistance. The authorized signatory should then issue a Bond (Annexure-VI) to MoSPI on Rs 100 Stamp paper within 15 days of the dispatch of the LoA

No utilization certificate would be required for releasing the installments. Only the Utilization of Funds & statement of expenditure (Annexure-V) would be submitted along with the project completion report at the end for releasing the final installment. If there are any unutilized funds, then it shall be refunded to MoSPI by bank drafts payable to Pay and Accounts Office, MoSPI. The details of the same would be communicated to the participating entity in the Notice Demand letter if issued. The quantum of funding is the sole discretion of the MoSPI. The approval criteria are indicated in ToR of the respective problem statements. Minimum cutoff for acceptance would be 80 percent for beta version and 60 percent for alpha version.

Parameters for Technical Evaluation:

The proposed solution from eligible participating entities will be evaluated on the following parameters:

Parameter		Description	Marks
1	Approach towards problem solving	Product idea, degree of innovation, simplicity of proposed solution, uniqueness & scalability of idea, novelty of approach,	30
2	Solution technical feasibility	Product features, scalability, interoperability, enhancement & expansion, underlying technology components & stack	25
3	Use case	Use case, USP and vision	10
5	Scalability roadmap and sustainability plan	Strategy for achieving population scale, longterm plan for revenue sustainability	10
6	Team ability	Experience of the previous work in the field of AI related project	10
7	Team Leader's profile	Relevant experience and Educational qualifications	5
8	Organization profile	Previous experience of organization in AI related projects	5
9	Adherence to responsible AI principles	Safety and reliability, equality, inclusivity and non-discrimination, privacy and security, transparency, accountability, protection and reinforcement of positive human values	5

Note: A minimum score of 50 will make the organization eligible for financial assistance. In case the entity does not achieve the cutoff criteria, the Committee may suggest changes to the original proposal and then reconsider the revised proposal subsequent to changes without having to apply again through the CPP portal. All the claims made regarding technical parameter needs to be supported with adequate documents duly signed.

Participation Process

The problem statements along with Terms of Reference shall be advertised online including on the Central Public Procurement Portal [<https://eprocure.gov.in/cppp/>] and website of MoSPI(www.mospi.gov.in) as part of Request for Proposals [RFP] and solutions/ proposals shall also be received online through the same portal in the prescribed format.

In such cases a two stage bidding process shall be followed where the weightages of Technical and Financial Bid would be 70:30 and procedure of QCBS as per Manual of procurement of Services for Consultancy Services will be adopted. In case of only one participating entity who achieve more than 50 points then the financial support would be decided by the Governing Council on recommendation of Selection Committee.

The eligible would submit the proposal within specified time (Three weeks from date of publications

of RFO on CPP portal) with documents as per the format given in Annexure-III. The eligible entities can apply for more than one problem statements.

Detail terms of Reference is at Annexure-IIA & IIB. The proposal will be governed by the guidelines of DI Lab issued vide notification no. 1-11011/1/2019-CC-MoSPI dated 23rd July, 2024. (https://www.mospi.gov.in/sites/default/files/announcements/DI_lab_guidelines.pdf).

Unless specifically mentioned/clarified in this document, the guidelines of DI Lab shall prevail in case of any doubt or contradiction regarding any of the clauses in this document with guidelines. The Technical and Financial submission forms may be filled separately and provided at Annexure-III & IV. All documents should be duly signed by the Team leader. The Team leader should be duly nominated through a letter issued under the signature of Authorized signatory of the entity. In case team of individual are participating a MoU for collaboration indicating the name of the team leader may be submitted.

Other Conditions:

- a. The outcome of this initiative can only be used by the participating team for the purpose of AI solution development.
- b. Selected entity will retain the rights to the solution/product developed. However, they should adhere to the terms and conditions defined in DI Lab guidelines related to AI solutions.
- c. The solution should not violate/breach/copy any copyrighted or patented concepts
- d. The solutions should not violate any data protection and governance regulations and policies.
- e. The solution should be in adherence to related cybersecurity standards and guidelines of the government of India.
- f. Solutions must adhere to ethical principles and guidelines for the development, deployment and use of AI technologies, including fairness, transparency, accountability, and non-discrimination

Termination:

This contract may be terminated by both parties upon 30 days' notice period. If the participating entity terminates the contract, then it shall refund the credited amount/fee with penal interest as applicable as per GFR. However, MoSPI may permit the Organisation/individual expert to retain part of the fee paid depending upon the services already rendered. MoSPI may also terminate this contract in whole or in part, in writing to the assigned selected entity. The initiation of arbitration proceedings in accordance with "settlement of disputes" stated below shall not be deemed a termination of this contract. In the event of any termination by the MoSPI, no payment shall be paid to the selected entity except for work and services satisfactorily performed in conformity with the terms and conditions of the contract. The selected entity shall take immediate steps to terminate the work and services in a prompt and orderly manner and to minimize losses and further expenditures. Should selected entity expert be adjudged

bankrupt, or be liquidated or become insolvent, or should the selected entity make an assignment for the benefit of its creditors, or should a receiver be appointed on account of the insolvency of the agency, the MoSPI may, without prejudice to any other right or remedy it may have, terminate this contract forthwith. Selected entity expert shall immediately inform the MoSPI of the occurrence of any of the above events.

Settlement of Disputes

- (i) Amicable settlement: The parties shall use their best efforts to settle amicably any dispute, controversy or claim arising out of, or relating to this contract or the breach, termination or invalidity thereof.
- (ii) Arbitration:
 - a. In the event of any dispute or difference between the parties hereto, such disputes or differences shall be resolved amicably by mutual consultation. If such resolution is not possible, then the unresolved dispute or difference shall be referred to arbitration of the sole arbitrator to be appointed in mutual consultation. In case it is felt necessary to do so, the Department of Legal Affairs would be consulted in an appropriate manner. The provisions of Arbitration and Conciliation Act, 1996 (No. 26 of 1996) shall be applicable to the arbitration.
 - b. All litigations will be subject to Delhi Jurisdiction. The venue of such arbitration shall be New Delhi only. The language of arbitration proceedings shall be English. The arbitrator shall make a reasoned award (the “Award”), which shall be final and binding on the parties.
 - c. The cost of the arbitration shall be shared equally by the parties to the agreement. However, expenses incurred by each party in connection with the preparation, presentation shall be borne by the party itself.
 - d. Pending the submission of and/or decision on a dispute, difference or claim or until the arbitral award is published; the Parties shall continue to perform all of their obligations under this scheme without prejudice to a final adjustment in accordance with such award.

Force Majeure

“Force Majeure” refers to any event or occurrence which results in either or both Parties are unable to perform their obligations under this MoU, without the fault, delay or negligence in performance of the Parties. Events that may be termed as force majeure events are including but not limited to fire, flood, strike, Act of God, acts of the public enemy or governmental authority in its sovereign or contractual capacity (including but not limited to declaration of lockdown), war, terrorism, epidemics, pandemics, civil unrest or riots, and/or power outage or grid failure or other similar causes beyond its control and without the fault or negligence of the delayed or non performing party.

Notification procedure for Force Majeure

- a) The affected Party shall notify the other Party of a Force Majeure event within seven (7) days of occurrence of such event. If the other Party disputes the claim for relief under Force Majeure, it shall give the claiming Party written notice of such dispute within thirty (30) days of such notice.

Such dispute shall be dealt with in accordance with the Dispute Resolution Mechanism in accordance to Clause 9.

- b) Upon cessation of the situation which led the Party claiming Force Majeure, the claiming Party shall, within seven (7) days, hereof notify the other Party in writing of the cessation and the Parties shall, as soon as practicable thereafter, continue performance of all obligations under this Agreement.

Contact Information: The Forwarding letter with proposal may be addressed to

Ms. Madhura Roy

Director, MoSPI

Email: madhura.roy@gov.in

**Manual Search within the Document impacts overall
efficiency of operations of MoSPI**

Details	
Particulars	Details
Name of Ministry/Department/Division	Ministry of Statistics and Programme Implementation/Computer Centre
Address	10, East Block, Rama Krishna Puram, New Delhi, 110066
Name of the Nodal Officer and Designation	Ms. Madhura Roy, Director
Phone Number (Nodal Officer)	8860570648
Email ID	madhura.roy@gov.in
Domain/Area: Problem Statement	Official Statistics in general
Category of problem statement <i>(Select all that are applicable)</i>	<input type="checkbox"/> Data Collection <input checked="" type="checkbox"/> Data Processing <input checked="" type="checkbox"/> Data Quality <input type="checkbox"/> Statistical Methodology <input checked="" type="checkbox"/> Data Accessibility <input type="checkbox"/> Data Integration and Interoperability <input checked="" type="checkbox"/> Data Timeliness <input type="checkbox"/> Data Standardization <input checked="" type="checkbox"/> Data Utilisation and Analysis <input type="checkbox"/> Data Visualization <input type="checkbox"/> Data Transparency <input type="checkbox"/> Technological Infrastructure <input checked="" type="checkbox"/> Data Management <input type="checkbox"/> Data Extraction and Pipeline Creation <input type="checkbox"/> Others (please specify)
What kind of support do you expect from the solution giver?	<input type="checkbox"/> Research and development of new methodology <input checked="" type="checkbox"/> Development or modification of tools/process
What kind of support/resources will you be able to share with MoSPI?	<input checked="" type="checkbox"/> Subject Matter Experts (SMEs) <input checked="" type="checkbox"/> Existing Datasets <input type="checkbox"/> Technical Resources <input type="checkbox"/> Collaborative Networks <input type="checkbox"/> IT Infrastructure <input checked="" type="checkbox"/> Physical Infrastructure (space for partners to work from) <input type="checkbox"/> Others (please specify)

*Mandatory

For non-mandatory questions, enter "N/A" in the given box if not applicable

Problem Statement

A. Problem statement Identified (Max 200 words) *

(Write a crisp and specific problem statement identified by the Ministry/Department/Division. Include aspects such as what the problem is, whom it impacts, and scale of impact. Try to give data figures wherever possible. Make sure the problem statement is related to official statistics areas.)

Ministry of Statistics and PI during its various activities needs to refer to manuals and other relevant documents mostly available in Portable Document Format (PDFs) or images. Though pdf is searchable but manual document retrieval requires specialised knowledge and consumes considerable time. Even if the manual is available in pdf or any other machine-readable format, search result may not yield the most relevant match and considerable effort is required to find the most pertinent information which remain elusive. For example, investigators often spend substantial time manually searching for specific information, interpretations, and guidelines within a manual during fieldwork, despite comprehensive training provided to them. This may result in productivity losses and delays in data collection and quality, hindering the effectiveness of the data-producing activities. It would also be helpful if there were an AI based tool available which facilitates quick search from the given document based on the input provided in writing or through a voice command. The tool should be able to search both the web or the given document(s).

Contains 171 words

B. Methodology used for identifying the problem statement. *

(Detail how the problem was identified, including any studies conducted, resources referred to, or methodologies applied.)

The problem was identified through a detailed review of existing processes and feedback received from various stakeholders, which highlighted the inefficiencies in manually searching documents. Feedback from officials, staff, and field officers was gathered through informal communication to assess the time taken for keyword searches in PDF documents. Insights from this exercise, combined with an analysis of common challenges faced in retrieving document content from the web, led to the identification of this problem.

*Mandatory

For non-mandatory questions, enter "N/A" in the given box if not applicable

C. Challenges imposed and need for solving them. *

(Explanation of the current situation, including relevant data and statistics that highlights the need for addressing this problem. List all key stakeholders affected by this problem, including internal teams, external partners, or end-users. Highlight the potential long-term impacts if the problem remains unsolved.)

The instructions manuals are typically stored as PDFs, making it difficult for users to efficiently search for specific content and interpretations, which should ideally come up in a handy way. The current process is highly time consuming, requiring individuals to perform keyword searches, and review the results—tasks that are time-consuming and prone to errors.

The lack of a streamlined, automated search process leads to significant delays in finding relevant information, reducing the overall productivity. The reliance on domain expertise means that newer staff or those unfamiliar with specific survey content face additional challenges, further slowing down operations. If left unaddressed, this problem could result in long-term inefficiencies, increased operational costs, and frustration among staff, ultimately impacting the quality of surveys and data driven insights.

Key Stakeholders Impacted:

Various Ministries, Departments, and Divisions including MoSPI, policy makers, partners awaiting crucial information, and end-users who rely on timely decisions, are all adversely affected by this issue.

D. Existing processes/systems in place to deal with the challenges (Max 150 words). *

(How is the Ministry/Department/Division currently addressing the problem statement? In case no way has been found to manage it, kindly mention that as well.)

The document is either available on the web or on the machine used for data collection in PDF/image. The user runs a manual search and uses its knowledge to reach the most relevant information. Overall, **no comprehensive, unified system is in place** to address the problem, leading to inconsistent search experiences and ongoing inefficiencies.

Contains 54 words

*Mandatory

For non-mandatory questions, enter "N/A" in the given box if not applicable

E. Expected outcome(s) for stakeholders' post resolution. *

(Clearly outline the benefits and improvements the impacted stakeholders will experience once the problem is resolved. Also mention the essential features of the solution.)

- **Improved Efficiency:** Quick retrieval of specific content and interpretations through advanced keyword and context-based search will significantly reduce document search time, facilitating faster understanding, decision-making, and easier adaptation to updated manual content.
- **Enhanced User Experience:** User-friendly interface making it easy for field officers to quickly access and become well-versed with manual content.
- **Better Resource Allocation:** Staff and field officers can focus on more critical tasks instead of manual searches.
- **Increased Accuracy:** Fewer errors in retrieving relevant information.

F. How urgent do you consider it to solve this problem? *

(What degree of impact does the problem have on operations?)

- High Priority:** The problem significantly impacts daily operations; needs immediate attention.
- Medium Priority:** The problem affects productivity or efficiency but does not halt operations. It should be addressed within a reasonable timeframe.
- Low Priority:** The problem has minimal impact on overall operations and can be resolved at a later time without major consequences.

G. Expected timeline for resolution?

(Mention the duration within which you expect a resolution.)

A resolution is expected within 6 to 9 months.

H. Share any global best practices you'd like to highlight?

(Mention any global best practices you know of that could address your issue or be implemented to solve it. Include links where possible.)

GPT and other foundation models provides for such functionality through their chat interfaces

*Mandatory

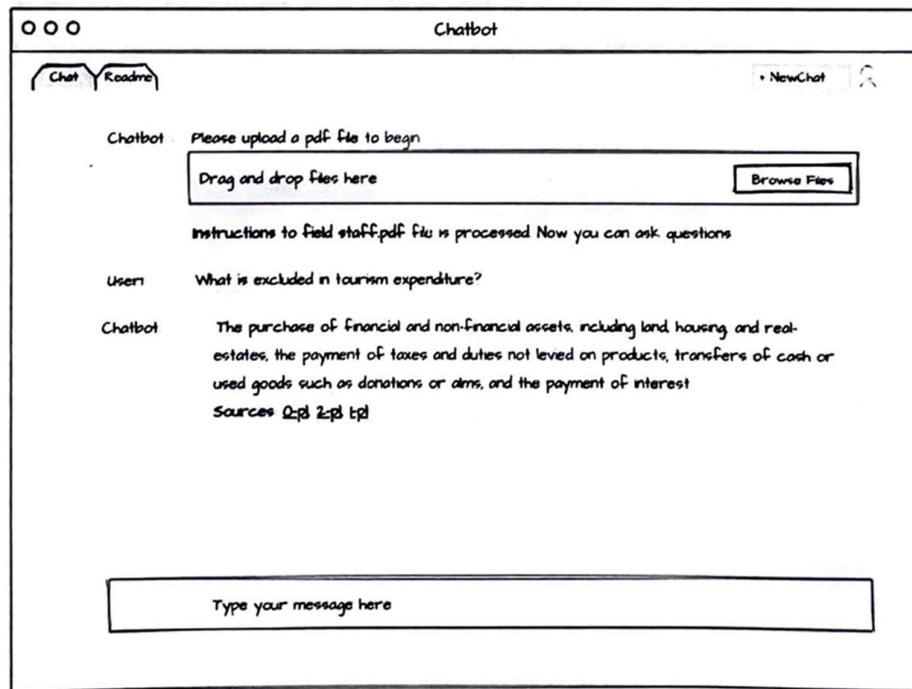
For non-mandatory questions, enter "N/A" in the given box if not applicable

(Complete this section only if the Ministry/Department/Division submitting the proposal has potential solutions in mind)

I. Proposed solutions (Max 200 Words)

(Provide an overview of the proposed solutions, including the key milestones and tentative timelines for each phase of implementation. Try to post your idea in points /diagrams /infographics /pictures.)

The solution is a web-based chatbot that uses AI-powered search based on NLP, users to upload documents or search the web and receive answers to their questions in natural language, whether asked through written or verbal commands. The chatbot should quickly return accurate results, displaying the relevant page from the document to the user. The interface will allow easy document uploading and searching, similar to a provided design, though minor differences are acceptable as long as it functions as required. The bot should be able to use website of the ministry or web in general if required.



J. Analysis of the feasibility of the solution

(Evaluate the viability of the proposed solution, considering its technical, financial, and operational aspects, along with identifying potential challenges and risks.)

The proposed solution is technically feasible, leveraging existing NLP and AI technologies that have proven effective in similar contexts, thereby offering enhanced search capabilities. Operationally, the solution can potentially be integrated with current document management systems/websites, minimizing disruption during implementation.

*Mandatory

For non-mandatory questions, enter "N/A" in the given box if not applicable

Less than Optimal use of legacy Data due to its availability in non-machine readable format	
Particulars	Details
Name of Ministry/Department/Division	Ministry of Statistics and Programme Implementation/Computer Centre
Address	10, East Block, Rama Krishna Puram, New Delhi, 110066
Name of the Nodal Officer and Designation	Ms. Madhura Roy, Director
Phone Number (Nodal Officer)	8860570648
Email ID	madhura.roy@gov.in
Domain/Area: Problem Statement	Official Statistics in general
Category of problem statement (Select all that are applicable)	<input type="checkbox"/> Data Collection <input type="checkbox"/> Data Processing <input type="checkbox"/> Data Quality <input type="checkbox"/> Statistical Methodology <input type="checkbox"/> Data Accessibility <input checked="" type="checkbox"/> Data Integration and Interoperability <input type="checkbox"/> Data Timeliness <input checked="" type="checkbox"/> Data Standardization <input checked="" type="checkbox"/> Data Utilisation and Analysis <input type="checkbox"/> Data Visualization <input type="checkbox"/> Data Transparency <input type="checkbox"/> Technological Infrastructure <input type="checkbox"/> Data Management <input checked="" type="checkbox"/> Data Extraction and Pipeline Creation <input type="checkbox"/> Others (please specify)
What kind of support do you expect from the solution giver?	<input type="checkbox"/> Research and development of new methodology <input checked="" type="checkbox"/> Development or modification of tools/process
What kind of support/resources will you be able to share with MoSPI?	<input checked="" type="checkbox"/> Subject Matter Experts (SMEs) <input checked="" type="checkbox"/> Existing Datasets <input type="checkbox"/> Technical Resources <input type="checkbox"/> Collaborative Networks <input type="checkbox"/> IT Infrastructure <input checked="" type="checkbox"/> Physical Infrastructure (space for partners to work from) <input type="checkbox"/> Others (please specify)

Problem Statement

A. Problem statement Identified (Max 200 words) *

(Write a crisp and specific problem statement identified by the Ministry/Department/Division. Include aspects such as what the problem is, whom it impacts, and scale of impact. Try to give data figures wherever possible. Make sure the problem statement is related to official statistics areas.)

Legacy data of this ministry, including surveys and macroeconomic indicators, are stored in formats such as PDFs, CSVs, Excel files (with merged cells, Hindi text, etc.), and images. Extracting this data requires detailed coding. Without necessary coding knowledge or resources, the data's usability is significantly hindered, affecting efficient analysis and decision-making. There is a need for an AI-based tool that can extract this data from various documents and store it in a database for further processing. The tool should have the functionality to read tables from documents and calculate descriptive statistics such as mean, median, and mode. Additionally, it should include the provision to plot graphs based on the read data for better appreciation. In essence there is a need to rejuvenate legacy data by use of AI based tools for ease of access and mangement

The following process should be executed by the bot on text or verbal command:

Load document → Extract text → Store in database → Perform SQL query

Contains 142 words

B. Methodology used for identifying the problem statement. *

(Detail how the problem was identified, including any studies conducted, resources referred to, or methodologies applied.)

The issue was identified through an assessment of existing data management practices within MoSPI at the time of development of the eSankhyiki portal. The review involved analysing the efficiency of storing data in relational databases using tables compared to retaining it in document or image file formats. During the review, using modern large language models (LLMs) for data extraction and analytics was also identified as a potential area for further exploration, as they significantly reduce the need for coding skills and make the process more user friendly.

C. Challenges imposed and need for solving them. *

(Explanation of the current situation, including relevant data and statistics that highlights the need for addressing this problem. List all key stakeholders affected by this problem, including internal teams, external partners, or end-users. Highlight the potential long-term impacts if the problem remains unsolved.)

The format in which data is often stored gives rise to several challenges:

1. Incompatibility with Modern Tools:

The legacy data is generally stored in non-machine readable formats, hindering advanced analysis.

2. Operational Inefficiencies:

Extracting and processing data from these reports, CSVs, Excel formats etc. is time-consuming and requires code-intensive skills, which slows down data retrieval and decision-making.

3. Data Fragmentation:

Data in various non-usable formats is often in silos, making it challenging to integrate datasets for comprehensive analysis.

4. Limited Scalability:

As data volume increases, extraction and processing methods using intensive coding struggle to keep up, limiting the ability to perform large-scale analytics.

5. Lack of Standardization:

The diverse formats and languages used in legacy data complicate its integration with modern systems, leading to inconsistencies and potential errors.

Key stakeholders affected include data analysts, policymakers, and internal teams responsible for statistical analysis and dissemination. If left unresolved, this issue could reduce productivity, and slow down innovation.

D. Existing processes/systems in place to deal with the challenges (Max 150 words). *

(How is the Ministry/Department/Division currently addressing the problem statement? In case no way has been found to manage it, kindly mention that as well.)

Currently, data extraction from these formats is done through scripting, which is very time consuming and code intensive.

Contains 15 words

E. Expected outcome(s) for stakeholders' post resolution. *

(Clearly outline the benefits and improvements the impacted stakeholders will experience once the problem is resolved. Also mention the essential features of the solution.)

Once resolved, stakeholders will be able to leverage advanced analytics. This will result in:

- 1. Improved Data Accessibility:** AI-based low-code/no-code tools for extracting and converting data from unusable formats to usable ones will enable quicker and easier data retrieval, reducing dependencies on specialized coding knowledge.
- 2. Improved Data Utilization:** More users will be able to use the data effectively without needing specialized technical skills.
- 3. Enhanced Productivity:** Reduced time and effort in processing data, allowing staff to focus on higher-value tasks.

F. How urgent do you consider it to solve this problem? *

(What degree of impact does the problem have on operations?)

- High Priority:** The problem significantly impacts daily operations; needs immediate attention.
- Medium Priority:** The problem affects productivity or efficiency but does not halt operations. It should be addressed within a reasonable timeframe.
- Low Priority:** The problem has minimal impact on overall operations and can be resolved at a later time without major consequences.

G. Expected timeline for resolution?

(Mention the duration within which you expect a resolution.)

Short-term: 3-6 months

H. Share any global best practices you'd like to highlight?

(Mention any global best practices you know of that could address your issue or be implemented to solve it. Include links where possible.)

Automated Data Extraction Tools:

Many organizations globally are leveraging AI-powered tools like **ChatGPT** and **Microsoft Copilot** to automatically extract data from PDFs, images, and other non-usable formats. These tools quickly convert data quickly into usable formats, enhancing accessibility and reducing reliance on manual

processing.

(Complete this section only if the Ministry/Department/Division submitting the proposal has potential solutions in mind)

I. Proposed solutions (Max 200 Words)

(Provide an overview of the proposed solutions, including the key milestones and tentative timelines for each phase of implementation. Try to post your idea in points /diagrams /infographics /pictures.)

N/A

J. Analysis of the feasibility of the solution

(Evaluate the viability of the proposed solution, considering its technical, financial, and operational aspects, along with identifying potential challenges and risks.)

N/A

Terms of Reference for AI-enabled intelligent search solutions for documents.

1. Background:

The Ministry of Statistics and Programme Implementation (MoSPI) handles large volumes of manuals, guidelines, and reports in formats such as PDFs and images. Efficient retrieval of specific information from these documents is critical for MoSPI's operations. However, current manual search processes are time-consuming and often return suboptimal results, hindering productivity. To address these inefficiencies, MoSPI is looking for an AI-powered intelligent search solution that can leverage advanced natural language processing algorithms to enable semantic based document search and text generation for results. The solution will enable staff to retrieve accurate and contextually relevant results through both text and voice-based queries, thereby improving operational efficiency and accuracy.

This proposal is governed by the **Data Innovation Lab's** guidelines regarding eligibility, financial assistance, and governance mechanisms, ensuring transparency and alignment with the lab's objectives.

2. Objectives:

The primary objective is to develop an AI-powered tool that enhances MoSPI's ability to quickly search, interpret, and retrieve information from large volumes of PDF and image-based documents. The tool should:

- **Improve Search Accuracy:** Ensure that the most relevant results are returned based on both text and voice queries.
- **Reduce Search Time:** Minimize the time required to find specific information.
- **Enhance User Experience:** Provide a user-friendly interface for both technical and non-technical users.
- **Increase Productivity:** Allow staff to focus on critical tasks by automating manual document searches.

3. Scope of Work:

The selected solution provider will be responsible for development of tool with following features and other tasks:

1. **AI-enabled intelligent semantic Search:** Capable of retrieving relevant information and generate the response of the queries in natural language from document files like PDFs, images, doc etc. stored in MoSPI's repository. The search should not only capture the syntax of question but as well as semantics of the query.
2. **Language and mode of user inputs:** Capable of taking user inputs in both voice and text form in natural language and able to support India's official languages part from English.
3. **Customizable as per MoSPI's requirements:** The tool must allow re-training or fine-tuning of ai/ml models as per MoSPI-specific documents and updates over time.
4. **User Interface Design:** An intuitive, simple and user-friendly web interface accessible via desktop, tablet and mobile platforms.

5. **Integration with Existing Systems:** Ensure the solution integrates seamlessly with MoSPI's current document management and database systems, allowing efficient retrieval across multiple document formats.
6. **Search Relevance and Speed:** Ensure the tool provides accurate results quickly, enhancing user productivity.
7. **Code and Documentation Submission:** Provide the source code via Git with detailed documentation on installation, architecture, and usage.
8. **Training and Support:** Conduct live training sessions for staff, and provide support documentation, including FAQs and troubleshooting guides.

4. Deliverables:

The solution provider must deliver the following:

1. **AI-powered Document searching tool:**
 - A functional tool capable of handling PDF, image-based, and other document types, providing accurate, contextually relevant results.
 - Support for text and voice-based search queries.
 - Efficient use of Natural Language Processing (NLP) to understand and process natural language queries.
2. **Code Submission via Git:**
 - A Git repository (e.g., GitHub, GitLab) containing the full source code, structured for easy collaboration and version control.
 - Regular updates with meaningful commit messages.
 - A detailed README.md file explaining the setup, repository structure, and usage instructions.
3. **Comprehensive Documentation:**
 - **Technical documentation** detailing the architecture, design, and operational flow of the system.
 - **Installation and deployment guides** for setting up the system with clear steps and dependencies.
 - **User guides** for technical and non-technical users, including instructions for text and voice-based searches.
4. **Explanatory Video:**
 - A video demonstrating installation, usage, and key functionalities, tailored for both technical and non-technical users.
5. **Training and Support:**
 - Live training sessions for MoSPI staff.
 - Support materials, including FAQs, troubleshooting guides, and technical assistance during the post-deployment phase (3-6 months).
6. **Integration with Existing Systems:**
 - Ensure the solution integrates with MoSPI's document management and repository systems, offering efficient and seamless access.

5. Delivery Timeline

A prototype of the fully developed, tested, and deployable solution may be provided within a period of 4 months of start of the contract. The contract start date will be considered after one month of awarding the contract.

The prototypes should be released in phases:

- Alpha Release: Initial version of the idea implementation along with code displaying capabilities of the solution should be released for technical staff available in MoSPI. They will access the quality, functionality; user experience of the solution on ministry's data. The solution will be substantiated with detailed presentation including but not limited to approach, methodology, timeline to all stakeholders.
- Beta Release: The feedbacks from alpha release should be analysed for feasibility. All the feedbacks which require improvement/changes and are feasible will be incorporated in the solution. The final release with all features will be released for User Acceptance Testing (UAT). The solution will be substantiated with detailed presentation including but not limited to approach, methodology, timeline to all stakeholders.

After completion of the project, the solution provider will provide 3 months of technical support required for consumption and adoption of prototype in the ministry.

6. Evaluation Mechanism:

Tool will be evaluated based on the parameters given below by a Committee. The alpha release will be evaluated based on the criteria's mentioned in the Sr. No. 1 of the table given below and the minimum score required for solution to be acceptable is 30 out of 45. The beta release will be evaluated based on all the criteria's mentioned in the table below and should get at least score 60 out of 100 to be acceptable by a Committee.

A. The proposal will be evaluated based on the following criteria:

Sr. No.	Criterion	Description	Marks
1.	Technological Approach	<ul style="list-style-type: none"> • Methodology and Architecture: The proposed solution shall be evaluated based on methodology and architecture design of the solution. The detailed methodology, solution architecture, flow charts, relevant technical documents and examples will be evaluated (20 points). • AI/ML Model Cost Effectiveness: The models required for solution development are: (5 points) <ul style="list-style-type: none"> ○ Self-hosted open source or community contributed models/APIs which are freely available for usage for AI/ML capabilities. (5 points) ○ Commercial paid version of AI/ML model as APIs (3 points) ○ Using commercially paid version of AI/ML applications with a few customization (0 points) • AI/ML Capability: The proposed solution is leveraging the state-of-the-art AI/ML capabilities (Max. 10 points). 	45

		<ul style="list-style-type: none"> ○ Able to process tabular data available as images, text and scanned documents, and take text and voice instructions for querying (10 points) ○ Able to process tabular data available as text, scanned documents and take only text instructions for querying (7 points) ○ Able to process tabular data available as text documents and take only text instructions for querying (5 points) ● Software Requirement: Open-source software tools required for development of the solution are accounting for (5 points) <ul style="list-style-type: none"> ○ More than 90% of all software tools required (5 point) ○ Between 60-90% of all software tools required (3 point) ○ Below 60% of all software tools required (1 point) ● Compute Infrastructure Requirement: The solution will be able to run on cloud/on-premise hardware with capacity of (5 points) <ul style="list-style-type: none"> ○ Max of 64 GB CPU RAM for 100 users using it simultaneously (5 points) ○ Max of 16GB vRAM and 64 GB CPU RAM for 100 users using it simultaneously (3 points) ○ More that 16vRAM and 64GB compute resource for 100 users using it simultaneously (2 points). 	
2	Performance	<ul style="list-style-type: none"> ● Accuracy of Data Extraction: Measures the extraction performance of the solution for extraction of tabular data from document files on test data-set shared by MoSPI (10 points) <ul style="list-style-type: none"> ○ Able to do with more than 80% accuracy (10 points) ○ Between 60-80% times (5 points) ○ Below 60% (0 points) ● Accuracy and Relevance of results: Measures the accuracy of information processed in summaries/insights/graphs and their relevance to the context on test data-set shared by MoSPI (10 points) <ul style="list-style-type: none"> ○ Able to generate accurate results more than 70% times (10 points) ○ Able to generate accurate results between 50-70% times (5 points) ○ Able to generate accurate results less than 50 % times (0 points) 	45

		<ul style="list-style-type: none"> • Query output latency: The solution will be able to display more than 99% of results generated from test data-set shared by MoSPI to user within (10 points) <ul style="list-style-type: none"> ○ 5 sec (10 points) ○ Between 5-10 sec (5 points) ○ Above 10 sec (0 points) • Document processing latency: The solution will be able to process more than 99% of documents in the test data-set shared by MoSPI within (5 points) <ul style="list-style-type: none"> ○ 30 sec (5 points) ○ Between 30-60 sec (3 points) ○ Above 60 sec (0 points) • Scalability of solution on large volume of dataset: The solution will be able to scale for ministry's data file more than 1000 documents (5 points) <ul style="list-style-type: none"> ○ Yes (5 points) ○ No (0 points) • Indian Official language support: The solution will be able to support Indian official languages in addition to English language (5 points) <ul style="list-style-type: none"> ○ 3 or more languages (5 points) ○ 1-2 language (3 points) ○ 0 languages (0 points) 	
3	Project Implementation and Support	<ul style="list-style-type: none"> • Project Timeline: After starting the contract and disbursing first instalment of financial assistance, the beta version of the proof of concept will be delivered within (5 points): <ul style="list-style-type: none"> ○ 4 months (5 points) ○ 4-6 months (3 points) ○ More than 6 months (0 points) • Technical Support: After expiry of mandatory support time of 3 months, the technical support in terms of training, minor customizations etc. required for adoption of solution in the ministry will be provided for (5 points): <ul style="list-style-type: none"> ○ 0 months (0 points) ○ 3 months (2 points) ○ 6 months (4 points) ○ 12 months (5 points) 	10

7. Code Submission and Quality Assurance Process:

- **Git Repository:** The full codebase must be submitted via Git for version control and collaboration.
 - **Commit History:** Frequent commits with meaningful descriptions are mandatory.

- **Branches:** Branches for different development stages (e.g., development, testing, production) should be maintained.
- **Code Quality:** The code will be reviewed for quality, maintainability, and adherence to best practices.
 - Proper commenting and documentation should accompany the code.
- **Documentation in Git:**
 - **README.md** file: Must include an overview, installation steps, usage instructions, and repository structure.
 - **Docs folder:** Additional technical documentation should be included in the repository's `docs/` folder.
- **Explanatory Video:** An instructional video demonstrating installation, usage, and troubleshooting should be included.

Terms of Reference for AI-based Legacy Data Extraction and Processing Tool

1. Background:

The Ministry of Statistics and Programme Implementation (MoSPI) holds a vast collection of legacy data, including but not limited to survey and macroeconomic indicator data, in tabular form available as text documents, scanned documents, images and stored in non-machine-readable/interoperable formats such as PDFs, Excel files, CSVs, and jpeg etc. This data is vital for analysis and decision-making. Extracting, processing and presentation of data tables from documents requires mix of manual and coding efforts which in return limits it's usability. To address this, MoSPI is looking for an AI/ML based multimodal conversational agent capable of extraction, ingestion, presentation of data and insight generation for end users. The solution should be capable of automatically ingesting structured as well as unstructured data from documents like pdf, excel files to relational or vector database and able to generate plots, insights like mean, median, trend in time series data using instruction given in natural languages like English.

This proposal is governed by the **Data Innovation Lab's** guidelines for all parameters including but not limited to eligibility, financial assistance, and governance mechanisms etc.

2. Objectives:

To develop an AI/ML software solution that can rejuvenate legacy data tables (samples attached for reference) by automating data extraction, processing, and visualization from document formats like pdf, csv, excel etc. The tool should:

- **Automate Data Ingestion:** Extract data from PDFs, CSVs, Excel files, and images and store it in relational / vector database.
- **Improve Data Accessibility:** Automating the data extraction and ingestion will reduce dependency on specialized coding skills addressing the skill barriers.
- **Enhance Data Utilization:** Enable users to perform basic data analysis like mean, median, spread, trend analysis, visualizations like bar charts, line charts, density and frequency plots etc. and summary insights about the data using instructions given in natural language like English or Hindi.

3. Scope of Work:

The selected solution provider will be responsible for development of tool with following features and other tasks:

- a) AI-based Data Extraction from documents: Capable of extracting tabular data from multiple formats, including PDFs, images, Excel files (with complex structures like merged cells and Hindi text).
- b) Automated Data Storage and Querying: Capable of automatically storing extracted data in a relational database and perform queries based on user input.
- c) Data Analytics and Visualization: Capable of generating descriptive statistics (mean, median, mode) and automate plotting graphs on the data tables extracted from documents.

- d) Language and mode of User Inputs: Support both text and voice commands to get the user input required for processing on the extracted data. Indian languages should be supported apart from English should be desirable.
- e) User Interface: A simple and intuitive web-based interface that enables users to upload documents, get user inputs, run queries, and generate visualizations.
- f) Code and Documentation Submission: Provide the source code via Git with detailed documentation on installation, architecture, and usage. The solution should be packaged as a docker image running on docker container and open-source scripting/programming languages, tools should be are open-source tools like Python, R, PostgreSQL.
- g) Training and Support: Conduct live sessions with MoSPI for demonstration of the features and provide technical support for deployment of the solution in pilot.
- h) Scalability: The tool should be scalable and should quickly be able to ingest documents available on the website of the ministry

4. Deliverables:

The solution provider must deliver the following:

- a) AI-based Data Extraction and Processing Tool:
 - a. A functional tool that automates data extraction from non-machine-readable formats and stores the data in a relational database and supports generation of descriptive statistics and visualizations.
 - b. The entire solution is available as a docker image.
- b) Code Submission via Git:
 - a. A Git repository (e.g., GitHub) containing the full source code, structured for easy collaboration and version control.
 - b. Regular updates with meaningful commit messages.
 - c. A detailed `README.md` file explaining the setup, repository structure, and usage instructions.
- c) Comprehensive Documentation:
 - a. Technical documentation: Detailed architecture, design, and operational flow of the system, unit tests, detailed wireframes of any UI developed as part of the solution.
 - b. Installation and deployment guides for setting up the system, including system requirements and configuration.
 - c. User guides for technical and non-technical users, including instructions for document upload, data extraction, querying, and visualization.
- d) Explanatory Video:
 - a. A video demonstrating installation, usage, and key functionalities, tailored for both technical and non-technical users.
- e) Training and Support:
 - a. Live demonstration and technical sessions for MoSPI staff.
 - b. Provide support materials, including FAQs, troubleshooting guides, and technical assistance
- f) Integration with Existing Systems:

- a. The solution should integrate with MoSPI's existing data infrastructure and be compatible with the ministry's databases and management systems as proof of concept along with a pilot.

5. Delivery Timelines:

A prototype of the fully developed, tested, and deployable solution may be provided within a period of 4 months of start of the contract. The contract start date will be considered after one month of awarding the contract.

The prototypes should be released in phases:

- **Alpha Release:** Initial version of the idea implementation along with code displaying capabilities of the solution should be released for technical staff available in MoSPI. They will access the quality, functionality; user experience of the solution on ministry's data. The solution will be substantiated with detailed presentation including but not limited to approach, methodology, timeline to all stakeholders.
- **Beta Release:** The feedbacks from alpha release should be analysed for feasibility. All the feedbacks which require improvement/changes and are feasible will be incorporated in the solution. The final release with all features will be released for User Acceptance Testing (UAT). The solution will be substantiated with detailed presentation including but not limited to approach, methodology, timeline to all stakeholders.

After completion of the project, the solution provider will provide 3 months of technical support required for consumption and adoption of prototype in the ministry.

6. Evaluation Mechanism:

Tool will be evaluated based on the parameters given below by a Committee. The alpha release will be evaluated based on the criteria's mentioned in the Sr. No. 1 of the table given below and the minimum score required for solution to be acceptable is 30 out of 45 to be eligible for financial assistance. The beta release will be evaluated based on all the criteria's mentioned in the table below and should get at least score 60 out of 100 to be acceptable for financial assistance.

A. The proposal will be evaluated based on the following criteria:

Sr. No.	Criterion	Description	Marks
1.	Technological Approach	<ul style="list-style-type: none"> • Methodology and Architecture: The proposed solution shall be evaluated based on methodology and architecture design of the solution. The detailed methodology, solution architecture, flow charts, relevant technical documents and examples will be evaluated (20 points). • AI/ML Model Cost Effectiveness: The models required for solution development are: (5 points) <ul style="list-style-type: none"> ○ Self-hosted open source or community contributed models/APIs which are 	45

		<p>freely available for usage for AI/ML capabilities. (5 points)</p> <ul style="list-style-type: none"> ○ Commercial paid version of AI/ML model as APIs (3 points) ○ Using commercially paid version of AI/ML applications with a few customization (0 points) <ul style="list-style-type: none"> ● AI/ML Capability: The proposed solution is leveraging the state-of-the-art AI/ML capabilities (Max. 10 points). <ul style="list-style-type: none"> ○ Able to process textual data available as images, text and scanned documents, and take text and voice instructions for querying (10 points) ○ Able to process textual data available as text, scanned documents and take only text instructions for querying (7 points) ○ Able to process textual data available as text documents and take only text instructions for querying (5 points) ● Software Requirement: Open-source software tools required for development of the solution are accounting for (5 points) <ul style="list-style-type: none"> ○ More than 90% of all software tools required (5 point) ○ Between 60-90% of all software tools required (3 point) ○ Below 60% of all software tools required (1 point) ● Compute Infrastructure Requirement: The solution will be able to run on cloud/on-premise hardware with capacity of (5 points) <ul style="list-style-type: none"> ○ Max of 64 GB CPU RAM for 100 users using it simultaneously (5 points) ○ Max of 16GB vRAM and 64 GB CPU RAM for 100 users using it simultaneously (3 points) ○ More that 16vRAM and 64GB compute resource for 100 users using it simultaneously (2 points). 	
2	Performance	<ul style="list-style-type: none"> ● Accuracy of Data Extraction: Measures the extraction performance of the solution for extraction of textual data from document files on test data-set shared by MoSPI (10 points) <ul style="list-style-type: none"> ○ Able to do with more than 80% accuracy (10 points) ○ Between 60-80% times (5 points) ○ Below 60% (0 points) 	45

		<ul style="list-style-type: none"> • Accuracy and Relevance of results: Measures the accuracy of information returned for queries on the document and their relevance to the context on test data-set shared by MoSPI (10 points) <ul style="list-style-type: none"> ○ Able to generate accurate results more than 90% times (10 points) ○ Able to generate accurate results between 70-90% times (5 points) ○ Able to generate accurate results less than 70 % times (0 points) • Query output latency: The solution will be able to display more than 99% of results generated from test data-set shared by MoSPI to user within (10 points) <ul style="list-style-type: none"> ○ 5 sec (10 points) ○ Between 5-10 sec (5 points) ○ Above 10 sec (0 points) • Document processing latency: The solution will be able to process more than 99% of documents in the test data-set shared by MoSPI within (5 points) <ul style="list-style-type: none"> ○ 30 sec (5 points) ○ Between 30-60 sec (3 points) ○ Above 60 sec (0 points) • Scalability of solution on large volume of dataset: The solution will be able to scale for ministry's all documents available on the website (5 points) <ul style="list-style-type: none"> ○ Yes (5 points) ○ No (0 points) • Indian Official language support: The solution will be able to support Indian official languages in addition to English language (5 points) <ul style="list-style-type: none"> ○ 3 or more languages (5 points) ○ 1-2 language (3 points) ○ 0 languages (0 points) 	
3	Project Implementation and Support	<ul style="list-style-type: none"> • Project Timeline: After starting the contract and disbursing first instalment of financial assistance, the beta version of the proof of concept will be delivered within (5 points): <ul style="list-style-type: none"> ○ 4 months (5 points) ○ 4-6 months (3 points) ○ More than 6 months (0 points) • Technical Support: After expiry of mandatory support time of 3 months, the technical support in terms of training, minor customizations etc. 	10%

		<p>required for adoption of solution in the ministry will be provided for (5 points):</p> <ul style="list-style-type: none"> ○ 0 months (0 points) ○ 3 months (2 points) ○ 6 months (4 points) ○ 12 months (5 points) 	
--	--	---	--

7. Code Submission and Quality Assurance Process:

- a) **Git Repository:** The full codebase must be submitted via Git for version control and collaboration.
 - a. **Commit History:** Frequent commits with meaningful descriptions are mandatory.
 - b. **Branches:** Branches for different development stages (e.g., development, testing, production) should be maintained.
- b) **Code Quality:** The code will be reviewed for quality, maintainability, and adherence to best practices.
 - a. Proper commenting and documentation should accompany the code.
- c) **Documentation in Git:**
 - a. **README.md** file: Must include an overview, installation steps, usage instructions, and repository structure.
 - b. **Docs folder:** Additional technical documentation should be included in the repository's `docs/` folder.
- d) **Explanatory Video:** An instructional video demonstrating installation, usage, and troubleshooting should be included.
- e) **Docs folder:** Additional technical documentation should be included in the repository's `\docs\` folder.

TECHNICAL PROPOSAL SUBMISSION FORM

Annexure-III

The draft proposal form is highlighted below. However, the details to be sought would be customized as per the problem statement in consideration.

1. Personal details and contact information of lead researcher	
Date of Application:	
Entity Submitting Proposal (Organization/ Individual):	
Name of Organisation:	
Organisation type: <i>Put 'X' next to the relevant organisation type.</i>	(a) Academic institution: (b) Company: Whether start-up: Yes/No (c) Research agency: (d) Individual researcher: (e) Others (Please specify): <i>(Note: Each of these would be referred to as agency/ entity in this document.)</i>
Registered Legal Status (in case of an organization, e.g., Company, Society, etc.)	
Address:	
Phone Number:	
Email:	
Credentials of Core Team Members	
<Name (Designation)>: <i>(Please create an extra row for each core team member followed by his/her credentials including qualifications and experience, as appropriate and applicable)</i>	<div style="background-color: #cccccc; padding: 5px; display: inline-block;"><credentials here></div>

<p>References:</p> <p><i>Indicate/ mention the attachment name of at least 2 persons who have notable works in the domain who are able to vet this proposal.</i></p>	
<p>2. Summary of proposed research</p>	
<p>Project title:</p>	
<p>Proposed start date:</p>	
<p>Estimated duration of required (months):</p> <p><i>Whether agency has own infrastructure/equipment</i></p>	

<p>Brief description of the project:</p> <p><i>Insight of what has been understood from the problem statement</i></p> <p><i>(Word limit: 1,000 words)</i></p>	
<p>Aims of the project:</p> <p><i>The results that will definitely be generated by agency.</i></p> <p><i>Also mention if any additional results may be provided.</i></p>	
<p>How does your proposal contribute to the requirements of MoSPI?</p> <p><i>Direct and indirect benefits may be provided</i></p>	
<p>3. Approach for Project Execution (Technology/Research design and methods)</p>	
<p>Technology/Research design/methods:</p> <p><i>Please provide a description of the approach for implementation of the project</i></p> <p><i>Details of the solution provided to the problem statement is being proposed</i></p> <p><i>Technology used, research design and methods you are planning to use and any other related information</i></p>	

4. Outputs, Deliverables & timelines

Intended outputs:

Please provide details on the outputs you intend to produce

Deliverables and Timelines

Clear deliverables and milestone along with the complete project time frame. e.g.: 1st week: Task x to be completed.

Deliverable Description	Proposed Timeline for deliverable

Publication and IPR:

Is agency ready to comply to agreement that IPR and external publication of any results produced from execution of this project would be joint effort of oSPI and agency.

Mentorship:

- (a) Whether mentorship is required?*
- (b) If so, the details the type of mentorship required.*

Deliverables, Milestones and Timelines

Please chalk out a detailed work dissemination plan. Associating this block with the funding requirements is critical.

Details of Project	Deliverable Description	Proposed Timeline for deliverable

5. Datasets			
MoSPI Datasets/Variables:			
<i>Please provide details on the datasets and variables you require for your projects</i>			
Data access need:			
<i>Please state any other datasets (not in public domain) which agency would need access to, if not how would they access</i>			
External Data sources:			
<i>Mention if datasets from other ministries are also required, whether they are in public domain etc., If not how would they access</i>			

Note:

- a. *If IP of any other organization or institution is involved, NOC from the said organization or institution should be obtained. In any of the above cases, if there are anyone/ team who has worked on the IP and if any of them is not part of the team, NOC all be obtained from such members for usage of the IP. Any dispute arising due to use/ misuse of IPR the participants/entity will be fully liable for such dispute and Ministry will have no responsibility to any such dispute.*
- b. *Attach the details of organisation related documents and credentials of the core team.*
- c. *An appropriate scoring method based on the project objective and its various component as approved by GC will serve as a guiding principle in selection of the project along with the project statement/project details.*
- d. *The agency should not be receiving any other government grant for the same project/ problem statement*
- e. *The agency/ entity shall provide appropriate documentation indicating a breakdown of Remuneration and Reimbursable in the contract form, commensurate with the indicative estimate presented in its EOI/ RFP.*

INDICATIVE FORMAT FOR PROVIDING INFORMATION ON STATEMENT OF SOLUTION

BACKGROUND/INTRODUCTION

ABBREVIATIONS USED

EXPECTED OUTCOMES/SOLUTIONS

--

DEFINITIONS USED

Note: The above format is indicative in nature. Additional headings/ sub-headings may be added as per the requirement.

The below information may be mandatorily provided in the format prescribed above as part of the technical proposal. In case few of the fields are not available then additional row/column may be added at the appropriate place.

Team Information

1. Team Leader Information:

- Full Name:
- Email Address:
- Phone Number:
- LinkedIn Profile:

2. Team Members:

- List each member with their full name, role, email, and LinkedIn profile.

3. Prior Experience in project implementation and research work

4. Prior experience in collaboration with government and/or private entities

5. Organization Details (if applicable):

- Organization Name:
- Registration Number:
- Type (Startup/Company/Other):
- Date of Incorporation:
- Address:
- Website:

Section 2: Project Proposal

1. Title of the Solution:

2. Problem Statement Addressed:

- Choose from Annexure I.

3. What is the research & development work that has already been done in this area? (Globally and in India)

4. Description of the Solution:

- Provide a detailed description of the solution, including its uniqueness and innovation.

5. Stage of Development in India and the World:

- Conceptual Idea
- Prototype Developed
- Existing Solution in Use

6. Brief on proposed Solution Capabilities

7. Technology Stack:

- List the technologies and tools used/to be used (Features & Specifications -AI algorithms, programming languages, platforms etc.)

8. Target Audience and Impact:

- Who will benefit from this solution? Describe the potential impact.

9. Use Case Scenarios

- Provide specific scenarios where the solution will be applied and what is the USP of the solution/innovation in addition to what has been given as part of the problem statement

10. Scalability and Sustainability:

- How can the solution be scaled? What are the long-term sustainability plans?

Section 3: Technical Details

- **Architecture and Design:**

Architectural diagrams, process flow, flowcharts, and other relevant design documents.

- 2. Data Management and Privacy:**

- Explain how data will be collected, stored, and used. Include measures taken to ensure compliance with relevant regulations and standards like data privacy and security measures.

- 3. Compliance with Responsible AI Principles:**

- Describe how the solution adheres to principles of fairness, transparency, accountability, and non-discrimination.

- 4. Briefly describe key technical hurdles that need to be overcome to implement your innovation and the resources required to do so (if any)**

- 5. Does your innovation have relevance to an existing or emerging technical standard? (A technical standard is a set of requirements for ensuring interoperability among devices or promoting reliability, productivity, efficiency, or safety of devices)**

- 6. Intellectual Property (IP) Status:**

- Detail any patents or IP rights associated with the solution.

Section 6: Declaration

- 1. Declaration by Team Leader:**

- I hereby declare that the information provided is accurate and complete to the best of my knowledge. I agree to abide by the rules and guidelines of the Data Innovation Lab Scheme component and of Request for Proposal.
- Date: _____

FINANCIAL PROPOSAL SUBMISSION FORM

Details of Funding Requirements (If any)

Please mention in detail your indicative funding requirements, i.e., an overall budget, comprehensive detailing on the breakup etc., in the table below

<Funding requirement should clearly commensurate with efforts/investment towards development of the solution by the propose.

Requirement specified should be, to large extent, be measurable and verifiable through timelines and deliverables. >

Cost Category	Requested Funding (INR)	Remarks if any
Manpower* (Consultants, free-lancers, researchers, etc. with estimated person- months to be mandatorily provided)		
Services (Subscription charges)		
Travel (Training, workshops,		
Goods/Equipment/data tools (non-perishable)		
Misc if any		

Note:

- Not more than four manpower will be provided financial assistance
- Travel Assistance will be provided for domestic tours as per level 12 officers of Government of India
- Subscription should be relevant to the project
- Not more than one-day workshop will be supported
- The equipment etc. procured need to be returned for the government of India on completion of the project
- Tax if any may be indicated separately
- A certificate of expenditure at the end of the project as per Annexure-V is required to be submitted along with other relevant documents.

Certificate of Expenditure

It is hereby certified that the funds have spent for the purpose for which it has been sanctioned as per the details below:

Cost Category	Amount Received (INR)	Amount Spent(INR)
Manpower* (Consultants, free-lancers, researchers, etc. with estimated person-months to be mandatorily provided)		
Services (Subscription charges)		
Travel (Training, workshops,		
Goods/Equipment/data tools (non-perishable)		
Misc if any		
Total		

Total Rs..... (.....in words) has been spent/utilized for the “Name of the project”

Name and Signature of the Team Leader

Date:

Place:

Name and Signature of the Authorized signatory of the Institution

Date:

Place:

Bond Form

- 1 Know all persons by these present that we the

 (Name of Samiti / Sansthan / Institution / An Association/ researcher) registered under the (Societies / Trust Registration Act (Specify the respective Act)/Companies Act/Affiliated to University) and having its office at _____ in the State of _____ through Shri / Smt./Ms. _____, duly authorized vide power of attorney / resolution/Order dated of the Samiti / Sansthan/ Institution / An Association/ Trust/University, (hereinafter called the "Obligator") and (I) Sh./Smt./Ms. Son/Daughter of Sh. Resident of, Aadhaar/Voter/ PAN/ ID card No (Attested copy must be enclosed) (II) Sh. Son/daughter of Sh.
 Resident ofAadhaar/ Voter/ ID card No(Attested copy must be enclosed) (Hereinafter sureties) are held and firmly bound to the President of India here in after called the Government, in the sum of Rs..... (Rupees (in words only)) together with interest thereon at the rate of 11.50% compounded annually (which would be adjusted as per the Government of India notification issued from time to time) from the date of receipt of the said amount by the Obligator up to the date of refund thereon to the Government. The term, obligator or sureties, unless repugnant to the context, shall mean and include the respective agents, assigns, heirs, successors etc., as the case may be.
- 2 Signed this day ofin the year
- 3 Whereas on the Obligator's request the Government has as per letter No.dated of Ministry of Statistics and Programme Implementation and here-in-after referred to as the Letter of Approval (which forms an integral part of these presents and copy of this is annexed as Annexure 'A'), agreed to make in favour of the Obligator for the purpose ofa fee of Rs. (Rupees.....) out of which Rs..... (Rupees... [amount of first installment]) have already been paid to the Obligator (the receipt) where the obligator do hereby admit and acknowledge two sureties in the terms and conditions manners contained herein after which the obligator and at its request the sureties have agreed to do.
- 4 Now in consideration of the aforesaid letter of sanction, the obligator herein binds itself and undertakes to comply with the conditions of the Letter of Approval referred to herein and if the obligator shall duly fulfill and comply with all its conditions mentioned in the Letter of Approval mentioning the fee then this bond or obligator's obligation therein shall be void and of no effect, but otherwise it shall remain in full force, effect and virtue, and the Government shall be at liberty to

enforce this bond against the obligator or the surety jointly and/or severally, as it may deem fit and on its option. These presents further witness that:

- a. That decision of the Secretary, MoSPI Govt. of India on the question whether there has been breach or violation on the part of the Obligator or any of the terms and conditions mentioned in the letter of approval, shall be final and binding on the Obligator.
- b. The liability of the sureties here under shall not be impaired or discharged by reason of time being granted by the Government or any forbearance, act or omission by or on the part of the Government whether with or without the knowledge or consent of the sureties in respect or in relation to the obligation or conditions to be performed or discharged in the obligation or by reason of the matter or thing whatsoever, which under the law relating to the sureties shall but for this provision have the effect of so releasing the sureties from such liability, nor shall it be necessary for the Government, to pursue the obligator before using the sureties or either of them the amount due here under.
- c. The **Obligator purein** agrees and undertake to surrender / pay to Government the monetary value of all such pecuniary or other benefits which it may receive or derive / have received or derived through / upon unauthorized use of fee for purpose other than that of which the fee was intended or the assets, purchased largely from out of "Government Funds". The decision of the Secretary, MoSPI as regards the monetary value of aforementioned to be surrendered /paid to the Government will be final and binding on the Society/Trust/Institution/ Organisation/ Researcher.
- d. The Obligator or the sureties shall, in the event of breach or violation of the terms and conditions mentioned in the Letter of Approval, refund to the Government on demand and without demur the entire amount of Rs. (Rupees) or such part thereof as may be mentioned in the **Notice Demand** issued by the Government along with the interest thereon at the rate of 11.50% compounded annually (which would be adjusted as per the Government of India notification issued from time to time) from the date of receipt of the said amount by the Obligator up to the date of refund thereof to the Government.
- e. The obligator and the surety assure and confirm that they have understood the scheme of assistance of sanction mentioned therein and they have executed this bond voluntarily and out of their free will.
- f. The Government of India has agreed to bear the stamp duty, if any, chargeable on these presents.

5. In witness thereof these presents have been executed on behalf of the Obligator and the Sureties the day and year here in above written and accepted for (Name and Designation) on the day and year appearing against his signature.

Signed for and on behalf of the Institution with date (Stamp/Seal)

(Name and Designation)

1. Signature of the Surety (Name and Designation of Surety with Identity Card such as Aadhaar/ Voter/ PAN/ ID card No (attach copy).....)

2. In the presence of:

(i).....

..... (Name & Address of witness)

(Signature)

Aadhaar/ Voter/ PAN/ ID card No (attach copy).....

(ii).....

..... (Name & Address of witness)

(Signature)

Aadhaar/ Voter/ PAN/ ID card No (attach copy).....

3. Accepted for and on behalf of the President of India by.