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

East Block No.10, R.K. Puram,

New Delhi, Date: 08.11.2024

Corrigendum

Subject: Clarification in respect of Central Public Procurement Portal Tender ID: 2024_MOSPI_831889_1.

Tender ID: 2024_MOSPI_831889_1 titled "AI Solutions and Data Extraction" was published on Central Public Procurement Portal on 22.10.2024.

2. The pre bid meeting was held on 04.11.2024 and certain clarifications were sought during the meeting the bidders /agencies attending the meeting.

3. Based on the inputs from the bidders and discussion during the pre-bid meeting, certain amendments in the tender document in respect of Central Public Procurement Portal Tender ID: 2024_MOSPI_831889_1 dated 22.10.2024 have been made and are placed in Annexure.

4. Last date and time of submission for bid has been extended to 19.11.2024 at 18:00 hrs.

5. Bidders shall bid as per updation defined in this corrigendum. Wherever there is a conflict, the clarifications herein shall prevail over the different aspects already mentioned in the bid document.

Yours sincerely

08/11/24

Tapan Das
Head of Office
Data Informatics and Innovation Division

Annexure

Subsequent to Pre-Bid Meeting held on 4th November, 2024, the following amendments are proposed in the RFP against paragraphs indicated below: -

Para No.	Original	May be read as
In RFP, under Heading "Payment Term", subsection Note-last line of 2nd paragraph	Minimum cut-off for acceptance would be 80 percent for beta version and 60 percent for alpha version.	Minimum cut-off for acceptance would be 30 out of 45 marks for alpha version and 60 out of 100 for beta version.
In RFP, under Heading "Parameters for Technical Evaluation", subsection Note	A minimum score of 50 will make the organization eligible for financial assistance	A minimum score of 60 will make the organization eligible for financial assistance
Annexure IIA Para 6: Evaluation Mechanism Sr No 1 Criterion: Technological Approach	• **AI/ML Capability:** The proposed solution is leveraging the state-of-the-art AI/ML capabilities (Max. 10 points). o Able to process tabular data available as images, text and scanned documents, and take text and voice instructions for querying (10 points) o Able to process tabular data available as text, scanned documents and take only text instructions for querying (7 points) o Able to process tabular data available as text documents and take only text instructions for querying (5 points)	• **AI/ML Capability:** The proposed solution is leveraging the state-of-the-art AI/ML capabilities (Max. 10 points). o Able to process textual data available as images, text and scanned documents, and take text and voice instructions for querying (10 points) o Able to process textual data available as text, scanned documents and take only text instructions for querying (7 points) o Able to process textual data available as text documents and take only text instructions for querying (5 points)
Annexure IIA Para 6: Evaluation Mechanism Sr No 1	• **Software Requirement:** Open-source software tools required for development of the solution are accounting for (5 points) o More than 90% of all software tools required (5 point)	• **Software Requirement:** Open-source software tools required for development of the solution are accounting for (5 points) o More than or equal to 90% of all software tools required (5 points)

Para No.	Original	May be read as
Criterion: Technological Approach	○ Between 60-90% of all software tools required (3 point) ○ Below 60% of all software tools required (1 point)	○ More than or equal to 70% but less than 90% of all software tools required (3 points) ○ More than or equal to 60% but less than 70% of all software tools required (2 points)
Annexure IIA Para 6: Evaluation Mechanism Sr No 1 Criterion: Technological Approach	**Compute Infrastructure Requirement:** The solution will be able to run on cloud/on-premise hardware with capacity of (5 points) ○ 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).	**Compute Infrastructure and Internet Connectivity Requirement:** The solution will be able to run on cloud/on-premise hardware with capacity of (5 points) ○ Max of 32 GB CPU RAM for 100 users using it simultaneously (5 points) ○ Max of 16GB vRAM and 32 GB CPU RAM or more than 32 GB CPU RAM for 100 users using it simultaneously (3 points) ○ More than 16GB vRAM and 32 GB CPU RAM compute resource for 100 users using it simultaneously (2 points). The solution should be able to work in limited internet connectivity.
Annexure IIA Para 6: Evaluation Mechanism Sr No 2 Criterion: Performance	• **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) ○ Able to do with more than 80% accuracy (10 points) ○ Between 60-80% times (5 points) ○ Below 60% (0 points)	• **Accuracy of Data Extraction:** Measures the extraction performance of the solution for extraction of textual data from document files at sentence level on test data-set shared by MoSPI. (10 points) ○ Able to do with more than or equal to 85% accuracy (10 points) ○ More than or equal to 75% and less than 85% (6 points) ○ More than or equal to 60% and less than 75% (3 points)
Annexure IIA Para 6: Evaluation Mechanism Sr No 2	• **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	• **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)

Para No.	Original	May be read as
Criterion: Performance	shared by MoSPI (10 points) o Able to generate accurate results more than 70% times (10 points) o Able to generate accurate results between 50-70% times (5 points) o Able to generate accurate results less than 50 % times (0 points)	o Able to generate accurate results more than or equal to 90% times (10 points) o Able to generate accurate results more than 70% time and less than 90% time (5 points)
Annexure IIA Para 6: Evaluation Mechanism Sr No 2 Criterion: Performance	**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) o 5 sec (10 points) o Between 5-10 sec (5 points) o Above 10 sec (0 points)	**Query Response Time:** 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) o 5 sec (10 points) o Between 5-10 sec (5 points)
Annexure IIA Para 6: Evaluation Mechanism Sr No 2 Criterion: Performance	**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) o 30 sec (5 points) o Between 30-60 sec (3 points) o Above 60 sec (0 points)	**Document Processing Time:** The solution will be able to process more than 99% of documents in the test data-set shared by MoSPI within (5 points) o 30 sec (5 points) o Between 30-60 sec (3 points)
Annexure IIA Para 6: Evaluation Mechanism Sr No 2 Criterion: Performance	• **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) o Yes (5 points) o No (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) o Yes (5 points)
Annexure IIA Para 6: Evaluation Mechanism	**Indian Official language support:** The solution will be able to support Indian official	**Indian Official language support:** The solution will be able to support

Para No.	Original	May be read as
Sr No 2 Criterion: Performance	languages in addition to English language (5 points) • 3 or more languages (5 points) • 1-2 language (3 points) • 0 languages (0 points)	Indian official languages in addition to English language (5 points) • 3 or more languages (5 points) • 1-2 language (3 points)
Annexure IIA Para 6: Evaluation Mechanism Sr No 2 Criterion: Project Implementation and Support	**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): • 4 months (5 points) • 4-6 months (3 points) • More than 6 months (0 points)	**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): • 4 months (5 points) • 4-6 months (3 points)
Annexure IIA Para 6: Evaluation Mechanism Sr No 2 Criterion: Project Implementation and Support	**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): • 0 months (0 points) • 3 months (2 points) • 6 months (4 points) • 12 months (5 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): • 3 months (2 points) • 6 months (4 points) • 12 months (5 points)
Annexure IIB Para 6: Evaluation Mechanism Sr No 1 Criterion: Technological Approach	• **AI/ML Capability:** The proposed solution is leveraging the state-of-the-art AI/ML capabilities (Max. 10 points). o Able to process textual data available as images, text and scanned documents, and take text and voice instructions for querying (10 points) o Able to process textual data available as text, scanned	• **AI/ML Capability:** The proposed solution is leveraging the state-of-the-art AI/ML capabilities (Max. 10 points). o Able to process tabular data available as images, text and scanned documents, and take text and voice instructions for querying (10 points) o Able to process tabular data available as text, scanned documents and take only text

Para No.	Original	May be read as
	documents and take only text instructions for querying (7 points) o Able to process textual data available as text documents and take only text instructions for querying (5 points)	instructions for querying (7 points) o Able to process tabular data available as text documents and take only text instructions for querying (5 points)
Annexure IIB Para 6: Evaluation Mechanism Sr No 1 Criterion: Technological Approach	**Software Requirement:** Open-source software tools required for development of the solution are accounting for (5 points) o More than 90% of all software tools required (5 point) o Between 60-90% of all software tools required (3 point) Below 60% of all software tools required (1 point)	**Software Requirement:** Open-source software tools required for development of the solution are accounting for (5 points) o More than or equal to 90% of all software tools required (5 points) o More than or equal to 70% but less than 90% of all software tools required (3 points) o More than or equal to 60% but less than 70% of all software tools required (2 points)
Annexure IIB Para 6: Evaluation Mechanism Sr No 1 Criterion: Technological Approach	**Compute Infrastructure Requirement:** The solution will be able to run on cloud/on-premise hardware with capacity of (5 points) o Max of 64 GB CPU RAM for 100 users using it simultaneously (5 points) o 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).	**Compute Infrastructure and Internet Connectivity Requirement:** The solution will be able to run on cloud/on-premise hardware with capacity of (5 points) o Max of 32 GB CPU RAM for 100 users using it simultaneously (5 points) o Max of 16GB vRAM and 32 GB CPU RAM or more than 32 GB CPU RAM for 100 users using it simultaneously (3 points) o More that 16GB vRAM and 32 GB CPU RAM compute resource for 100 users using it simultaneously (2 points). The solution should be able to work in limited internet connectivity.
Annexure IIB Para 6: Evaluation Mechanism Sr No 2	• **Accuracy of Data Extraction:** Measures the extraction performance of the solution for extraction of textual data from document	• **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)

Para No.	Original	May be read as
Criterion: Performance	files on test data-set shared by MoSPI (10 points) o Able to do with more than 80% accuracy (10 points) o Between 60-80% times (5 points) o Below 60% (0 points)	o Able to do with more than or equal to 85% accuracy (10 points) o More than or equal to 75% and less than 85% (6 points) o More than or equal to 60% and less than 75% (3 points)
Annexure IIB Para 6: Evaluation Mechanism Sr No 2 Criterion: Performance	• **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) o Able to generate accurate results more than 90% times (10 points) o Able to generate accurate results between 70-90% times (5 points) o Able to generate accurate results less than 70 % times (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) o Able to generate accurate results more than or equal to 70% times (10 points) o Able to generate accurate results more than 50% times and less than 70% times (5 points)
Annexure IIB Para 6: Evaluation Mechanism Sr No 2 Criterion: Performance	• **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) o 5 sec (10 points) o Between 5-10 sec (5 points) o Above 10 sec (0 points)	• **Query Response Time:** 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) o 5 sec (10 points) o Between 5-10 sec (5 points)
Annexure IIB Para 6: Evaluation Mechanism Sr No 2 Criterion: Performance	• **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) o 30 sec (5 points) o Between 30-60 sec (3 points) o Above 60 sec (0 points)	• **Document Processing Time:** The solution will be able to process more than 99% of documents in the test data-set shared by MoSPI within (5 points) o 30 sec (5 points) o Between 30-60 sec (3 points)
Annexure IIB	**Scalability of solution on large volume of dataset:** The solution	**Scalability of solution on large volume of dataset:** The solution

Para No.	Original	May be read as
Para 6: Evaluation Mechanism Sr No 2 Criterion: Performance	will be able to scale for ministry's all documents available on the website (5 points) o Yes (5 points) • No (0 points)	will be able to scale for ministry's data file more than 1000 documents (5 points) o Yes (5 points)
Annexure IIB Para 6: Evaluation Mechanism Sr No 2 Criterion: Performance	**Indian Official language support:** The solution will be able to support Indian official languages in addition to English language (5 points) • 3 or more languages (5 points) • 1-2 language (3 points) • 0 languages (0 points)	**Indian Official language support:** The solution will be able to support Indian official languages in addition to English language (5 points) • 3 or more languages (5 points) • 1-2 language (3 points)
Annexure IIB Para 6: Evaluation Mechanism Sr No 2 Criterion: Project Implementation and Support	**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): • 4 months (5 points) • 4-6 months (3 points) • More than 6 months (0 points)	**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): • 4 months (5 points) • 4-6 months (3 points)
Annexure IIB Para 6: Evaluation Mechanism Sr No 2 Criterion: Project Implementation and Support	**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): • 0 months (0 points) • 3 months (2 points) • 6 months (4 points) • 12 months (5 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): • 3 months (2 points) • 6 months (4 points) • 12 months (5 points)

In RFP, under Heading “Eligibility Criteria”, subsection “Note”, the following two points may be read in addition to the existing points: -

One agency can apply for more than one problem statement.

Separate proposal (both Technical and Financial) needs to be submitted for each problem statement.

Some more queries have been received through email, which are provided below: -

Sr No	Query posed	Reply
1.	Scope of Search Functionality: Could you please specify whether the AI search solution needs to support: • Searches within a single PDF/Document file only. • Searches across multiple files simultaneously. • Or both functionalities?	The search engine should be able to search within a set of documents having multiple files. It should be able to search within any document file, like MS Word or pdf (searchable or non-searchable). Test datasets or their reference will be provided by Ministry. Please refer to “Accuracy of Data Extraction” and “Accuracy and Relevance of Results” under “Performance” heading of Sr No. 2 of the Product Evaluation Criteria of Annexure IIA and IIB.
2.	Document and File Types: • Could you provide a detailed list of the types of documents and formats (e.g., PDFs, CSVs, Excel files, images, etc.) the solution should support? • If the solution is expected to handle consistent file types or specific formats, we would appreciate any predefined characteristics or standards that should be adhered to for optimal performance.	