Call for Proposal for E-SESH

1.Introduction:

This project aims at developing an Artificial Intelligence Enabled Multi-purpose System that will be able to provide a simple, informative, and easy-to-use interface supporting multiple AI-based modules such as Emotion and Sentiment Analysis, Sarcasm detection, and Hate Speech Detection.

2.Project Description:

a) **Upgradation of C-DOT Search Star Application**

C-DOT text-analytic browser-based application 'Searchstar' collects filtered data from Twitter, passes them through multiple analytical engines simultaneously and displays the analysed inferences in user-friendly dashboards. It integrates the Sentiment and Emotion models for Hindi and English language, developed as part of 'SESH' project, for sentiment and emotion analysis respectively on collected tweets. Following enhancements are proposed in this project proposal to make 'Searchstar' client-ready:

i. Up-gradation of software packages (Python, Php, Apache kafka, MongoDB, Neo4j, OS)
ii. Integration of SESH Hate-Speech model including design and development of basic dashboard.
iii. Dashboard design for inferences from Emotion Analysis
iv. Migration to new software architecture to support multiple simultaneous queries and synchronization between models before query's topic deletion
v. Integration of SESH -Sarcasm and Hate-Speech models including output dashboards design and development.
vi. Secure Connection, TLS/SSL security for web APIs
vii. VAPT testing for Searchstar, Certification

b) **Feedback based fine-tuning**

Dedicated feedback interface to be developed for collecting user's comments and suggestions

User-feedback facility will have mechanism to collect:
1. Quality of predicted output

2. Output Classes predicted correctly or not
3. Suggested correct output
4. Evaluate the user feedback periodically and update the model

c)   Sentiment Analysis on Code-mixed Inputs

Understanding the sentiment of the public about a particular event can help the government to make strategic decisions. It can be an essential indication for online information operations in identifying issues of concern or potential bad user behaviours. The presence of sarcasm - a linguistic expression intended to indicate the opposite of what is said with the goal to offend or ridicule - in online content is a substantial impediment to sentiment analysis performance. Sarcasm detection tool can be of relevance to security agencies/LEAs for anything from improving sentiment analysis in general to refining automated monitoring of social media for national security threats through an understanding of sarcastic false positives.

d)   Hate Speech Detection on Multi-modal Inputs:
In the recent past, social media platforms have helped people connect and communicate with a wider audience. But this has also led to a drastic increase in cyberbullying. It is essential to detect and curb hate speech to keep the sanity of social media platforms.

e)   Named Entity Recognition

The existing NER system will be updated. This will require creating new datasets, and build models using Conditional Random Field and/or Transformer based models.

f)   Topic Modelling

The existing topic modelling algorithm will be adapted for the task.

3. Roles and responsibilities of CDOT:

C-DOT team will conduct algorithm, design and code reviews. It will assess/test the deliverables and monitor the project.

4. Roles and responsibilities of participants:

For getting the product 'client-ready', the required modules can be utilized independently for task-specific applications. These shall be put in a pipeline to enhance the overall system's performance.

The roles & responsibilities of participants are as follows:

1. Data collection
2. Data annotation
3. Upgradation of CDOT SearchStar Application and testing.
4. Sentiment Analysis module upgrade and testing
5. Hate-Speech Detection module upgrade and testing
6. Named Entity Recognition module upgrade and testing
7. Topic Modelling module upgrade and testing
8. Secure Connection, TLS/SSL security for web APIs
9. Security enhancement for VAPT and testing
10. Release generation and integrated testing
11. VAPT & Certification
12. Hosting C-DOT SearchStar application as a service.
13. Sentiment Analysis on Code-Mixed Inputs.
14. Hate Speech Detection on Multi-modal Inputs,
15. Incremental dataset creation/ upgradation of Sentiment, Hate-Speech, NER, Topic Modelling.
16. Maintenance of the system
17. TOT to be given to CDOT team.

Timeframes(Note) :
Primary SearchStar hosting as a service to be done in 6 months.

5. Ownership of outcomes:

CDOT will own the IPR of the developed solution.

6. Format of response:

Companies / organizations / institutions / individuals developing enabling technologies / modules / components / subsystems / products in this area are required to respond to this EOI in the format provided in
https://www.cdot.in/cdotweb/web/ccrp.php

7. Submission Procedure:

A separate response shall be emailed to connect-ccrp@cdot.in, not later than three weeks from date of issue of this EoI.

8. The Next Steps:

On receipt and evaluation of responses, C-DOT will make an assessment of indigenous technologies available for achieving objectives of the project.

A Product Requirement Specifications of the final product to be built collaboratively with Partners will be evolved through an open process of consultations with all concerned stake holders. A formal Request For Proposal (RFP) shall be issued for selection of collaboration Partners for the project.