IndiaAI CyberGuard AI Hackathon

1. Introduction:

IndiaAI, an Independent Business Division (IBD) under the Digital India Corporation (DIC) of the Ministry of Electronics and IT (MeitY), is the implementation agency of the IndiaAI Mission, which aims to democratize AI's benefits across all strata of society, bolster India's global leadership in AI, foster technological self-reliance, and ensure ethical and responsible use of AI.

As part of this Mission, the IndiaAI Application Development Initiative (IADI) will promote the development, deployment, and adoption of AI applications in critical sectors that have the potential to catalyze large-scale socio-economic transformation.

With this mandate, IndiaAI is working in partnership with the Indian Cybercrime Coordination Centre (I4C), established by the Ministry of Home Affairs to leverage AI to drive efficiencies in the prevention, detection, investigation, and prosecution of cybercrime, thereby enhancing citizen welfare. I4C provides a comprehensive framework for Law Enforcement Agencies (LEAs) to tackle cybercrime effectively and manages the National Cybercrime Report Portal (NCRP).

As a part of the initiative, IndiaAI in partnership with I4C is launching the IndiaAI CyberGuard AI Hackathon. The Hackathon is a strategic effort to leverage advanced technologies such as Artificial Intelligence (AI) to improve cybersecurity and address the escalating threat of cyber fraud and crimes. With approximately 6,000 cases reported daily through the NCRP, this initiative seeks to develop robust AI-driven solutions capable of managing and mitigating these threats. The hackathon provides a platform for participants to design and implement sophisticated AI models that can enhance the effectiveness of cybersecurity measures and combat the growing sophistication of cybercrime.

2. Guide to the Hackathon Participation

I) **Stage 1:**

a) Participation requirement:

- Participants may participate as individuals or in groups as per the eligibility criteria detailed below.

b) **Event Process and Schedule**

- All participants are required to register at XX to access the application form.
- A team leader will have to individually register and apply for the challenge on the IndiaAI portal by clicking the submit link.
- After the initial sign-up, the Team Leader should list all the Team Members under the Management Team and complete all organization details.
- Team Leaders will also create an account on GitHub/Open Forge to upload Source Code. The Next, the Team Lead shall paste the link for their GitHub/Open Forge source code in the application form.
- Participants may access the Datasets and the Data Dictionary upon registration.
- Additionally, the Team Leader will have to answer all additional questions, including uploading documents, and click 'Submit'.
- The Hackathon under Stage 1 would take place virtually over 20 days from the start of registration to the final date for submission of developed prototypes. Any edits to the Source Code post the final date of submission will lead to immediate disqualification of the application.

c) Problem Statement:

- **Problem Statement : Development of an NLP Model for Text Analytics and Classification**
- Objective:
  - To develop an NLP model that categorizes complaint based on victim, type of fraud and other relevant parameters used for text classification and preparing the final model.
- Outputs:
  - Text Preprocessing: Tokenization, stop word removal, stemming, and text cleaning to prepare data.
  - Model Development: Selection of a suitable NLP model for text classification
  - Accuracy Measurement: Evaluate the model based on metrics such as accuracy, precision, recall, and F1-score.

II. Stage 2:

a) Participation requirement:

- Top 20 Solutions will be selected for an in-person round.
- Upto 3 members per team will be provided travel and accommodation support to participate.

b) Event Process and Schedule:

- The second round will be organised as an in person 3 day event in New Delhi. Details on the location and logistics will be communicated via email to the shortlisted participants.
- Participants would receive a dataset from the NCRP database.
- Participants need to work on the given problem statement detailed below.

c) Problem Statement:

Problem Statement: Development of an NLP Model to guide citizens in filing cybercrime reports on the National Cyber Crime Reporting Portal (NCRP) correctly through a real time analysis of the description and incident supporting media files uploaded by the citizen.

- Objective:

  To develop an NLP model that categorizes fraud based on victim descriptions, focusing on improving model accuracy.

- Outputs:
  - Text Preprocessing: Tokenization, stop word removal, stemming, and text cleaning to prepare data.
  - Model Development: Selection of a suitable NLP model (e.g., Logistic Regression, LSTM, or BERT) for multi-class fraud categorization.
  - Accuracy Measurement: Evaluate the model based on metrics such as accuracy, precision, recall, and F1-score.

- Result:

Participants will deliver a working NLP model with an emphasis on achieving the highest possible accuracy within the 3-day timeframe.

III. Stage 3:

a) Participation requirement:

- The Top 3 teams will be selected post evaluation of the results from Stage 2.

b) Event Process and Schedule:

- The teams will pilot and deploy their prototype based on the datasets and development environment provided in collaboration with IndiaAI, I4C, and a MeitY empanelled CSP.
- The teams will be free to choose any of the empanelled CSPs as per their choice. The solution will then be presented to the Jury along with a presentation made by the team.
- The winning team will be awarded INR 25 lakhs based on their working prototype. The winner will also get a contract to deploy their solution for use by the Government of India and its associated entities for a period of 4 years.

3. Expected Deliverables from Participants

- **Model Code and Documentation (OpenForge/GitHUb)**
  - Clear and well-documented code used to build, train, and test the submitted model.
  - Explanation of the key methodology and steps taken in model development.
- **Project Report (PDF)**
  - Discuss the significant findings from your NLP analysis (Summarize in 500 words)

    For instance:

    - Sentiment trends over time
    - Commonly recurring themes/topics
    - Present visualization as per need
    - Text classification accuracy and key drivers behind correct/incorrect predictions

- What is the Evaluation of the model using the provided metrics?
- What is your implementation plan (Summarize in 200 words)

  (e.g., changes to the system, further analysis needed, deployment plans)
- Cite all relevant work used, libraries, and others along with plagiarism declaration.

- **Presentation (Only Stage 3)**
  - Summary of your approach, findings, and recommendations.
  - Visual aids to support your presentation (graphs, charts, etc.).
  - Appendices if any

4. Submission Guidelines

1. Fill submission form with all fields.
2. The write-up should follow these formatting guidelines:
   - **Submission Format**: PDF
   - **Font**: Times New Roman
   - **Font Size**: 12 pt
   - **Margins**: 1-inch margins on all sides
   - **Line Spacing**: 1.5 line

5. Evaluation Process:

The evaluation process of the Hackathon would be overseen by a distinguished panel of jury members comprising experts from the fields of machine learning, data science, and cyber security. The jury would rigorously assess each submission based on predefined criteria to ensure a fair and comprehensive evaluation. The evaluation will ensure equitable weightage is given to both the Technical and General parameters.

Stage 1

- **Initial Screening**: Submissions would undergo an initial screening to ensure compliance with submission guidelines and basic functionality.

- **Technical Evaluation**: The jury would conduct a detailed technical evaluation of the models.

Stage 2

- Based on the initial evaluation, up to 20 teams would be shortlisted for the second round.
- In second round, post submission of the deliverables, top 3 teams will be selected for a presentation with the jury.

Stage 3

- The three teams will pilot their solutions in a contained real-world environment to the Jury to illustrate the efficacy of their solutions. Basis the results of the pilot program the winning team of the hackathon will be selected.

a) Decision Making

- Prizes would be awarded to the top three teams whose models meet the jury's satisfaction regarding usability as viable solutions. A special prize would be awarded to women only teams, if any.
- If no solution meets the required standards, consolation prizes would be awarded based on the jury's discretion.
- The jury's decision would be final and binding.

6. Awards/Outcomes:

- **Opportunity to Build for the Nation**: Contribute to developing innovative solutions that address critical challenges faced by the country, making a direct impact on society.
- **National Recognition**: Gain visibility and recognition from government officials, industry leaders, and peers for your contributions and innovative ideas.
- **Networking Opportunities**: Connect with like-minded innovators, potential collaborators, and key stakeholders in the tech and innovation ecosystem.

- **Exposure to Real-world Challenges**: Work on pressing issues faced by the nation, providing practical experience and a deep understanding of real-world problems.

- **Support for Implementation**: Wining solution will get potential support in scaling and implementing the solution at a national level, bringing your ideas to life.

- **Prizes and Incentives:**

  - **First Prize**: Upto INR 25 Lakhs

  - **Second Prize**: Upto INR 7 lakhs

  - **Third Prize**: Upto INR 3 lakhs

  - **Special Prize of INR 5 lakhs for All-Women Teams** (in addition to the top three prizes)

7. Eligibility:

- **Indian Company:** The team may be an Indian company registered under the Companies Act. An Indian company must have 51% or more shareholding by Indian citizens or persons of Indian origin.
- **Start-up:** Alternatively, the team can qualify as a start-up according to the latest notification by the Department of Industrial Policy and Promotion (DIPP), which can be accessed at Startup India.
- **Academic/R&D Organizations**: All Institutions falling in the following categories would be eligible to participate in the programme and receive funding:
  - Indian Institutes of Technologies (IITs)
  - National Institutes of Technologies (NITs)
  - Indian Institutes of Information Technology (IIITs)
  - Indian Institutes of Science Education and Research (IISERs)
  - Central Universities/Deemed Universities under Central/State Government
  - Colleges/Institutions of National Importance/Eminence
  - R&D Organizations/Institutions

- Private Universities/ Private Deemed Universities/Private Colleges-Institutions should be approved by AICTE and/or the Institution should be accredited by NAAC (National Assessment and Accreditation Council of UGC).

- **Autonomous Bodies:** Autonomous bodies, including public sector organizations, are eligible to participate.
- **Others**: Indian students or researchers associated with educational institutions, or working professionals associated with Indian startups and companies can participate in their individual capacity or as teams. The participant must be the citizen of India. However, Teams must ensure they are registered before they are selected for the Stage 3.

8. Evaluation Parameters:

I. General:

	Parameter	Description
1	**Approach Towards Problem Solving**	Product Idea, Degree of Innovation, Simplicity of Final Solution, Uniqueness of Idea, Novelty of Approach,
3	**Solution Technical Feasibility**	Product features, Scalability, Interoperability, enhancement & expansion, Underlying technology components & stack and futuristic orientation
4	**Product Roadmap**	Potential Cost to Build Product, Regulatory compliance, System Integration plan
5	**Team Ability & Culture**	Prior Experience, Team Leader's Effectiveness (i.e. Understanding of subject matter, Ability to guide, Ability to present idea), Ability to Market Product, Growth Potential of Organization
7	**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
8.	**Adherence to Data Policies and Cyber Security Guidelines**	Adherence to applicable Government of India policies, guidelines, regulations on Data Governance and Cyber Security

II. Technical

	Parameter	Description
1.	**Data preparation**	○ Has explored the data and removed unnecessary columns. ○ Participants has checked if there is any skewness in the data and tried to mitigate it. ○ Has performed stratified train-test split successfully to create train & test datasets.
2.	**Model Building**	○ Participants has performed the required cross-validation and has built different models on raw-data. ○ After evaluation on the raw dataset, Model hyperparameters are tuned using correct principles and the approach is explained clearly. ○ A reasonable number and variety of different models are attempted, and the best one is chosen based on key performance metrics.
3.	**Model Evaluation**	○ Model evaluation is conducted using an appropriate metric. ○ Model evaluation results are on par with the best possible models on this data set.

4.	**Code readability and conciseness**	○ The code is well commented, and analysis is explained in the report format with findings from the dataset. ○ Efficient, concise code is written.
5.	**Technical Robustness**	○ Accuracy: The proportion of correctly classified instances (both true positives and true negatives) out of the total instances. ○ Precision: The proportion of true positive instances out of the instances predicted as positive. ○ Recall (Sensitivity or True Positive Rate): The proportion of true positive instances out of the actual positive instances. ○ F1 Score: The harmonic mean of precision and recall, providing a single metric that balances both concerns. ○ AUC-ROC (Area Under the Receiver Operating Characteristic Curve): Measures how well the model distinguishes between classes. ○ Confusion Matrix: A table providing a detailed breakdown of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). ○ Other Metrics (Optional): Log Loss and Balanced Accuracy of the model. ○ Additional Criteria: Any other metrics as decided by jury members.

9. Timeline:

S.No.	Activity	Timeline

1	Launch Date	17th October 2024
2	Last Date for Online Submission	7th November 2024
3	Announcement of Results of First Round	9th December 2024
4	Round 2 (New Delhi)	18th December 2024
5	Pilot Program and Announcement of Winner	8th January 2024

10. Intellectual Property Rights;

The models submitted or awarded will become the property of IndiaAI and I4C, including all intellectual property rights to their underlying methodologies and innovations, and the participants shall be deemed to have given their no objection/consent for the same and shall also remain bound by the terms of a Non-Disclosure Agreement (NDA) with respect to such work. The participants agree to provide a No Objection Certificate as an author in favour of IndiaAI and I4C for the purposes of IPR registration and ownership rights, as and when required by IndiaAI and I4C.

11. General Terms:

1. All participants have to be eligible (See Eligibility Criteria) to participate.
2. If individual innovators are associated with any institution, they must provide a No Objection Certificate (NOC) from their employer, stating that the company will have no rights to the prize money and intellectual property rights (IPR). Additionally, individuals must inform their employer of their participation in the Hackathon.
3. Any edits to the source code post the last date of submission will lead to immediate disqualification.
4. The participants will ensure code is free from viruses, malware.

5. The participants will not use this contest to do anything unlawful, misleading, malicious, or discriminatory.
6. The solution must not violate/breach/copy any copyrighted or patented, concepts in the AI market.
7. The solutions must not violate any data protection and governance regulations and policies.
8. The solution must be in adherence with related cybersecurity standards and guidelines of the Government of India.
9. The developed solution/product will be deployed in the chosen Cloud Environment and used for Union/State/UT government entities.
10. Solutions must adhere to ethical principles and guidelines for the development, deployment and use of AI technologies, including fairness, transparency, accountability, and non-discrimination.

12. Plagiarism and Ethics

1. Participants are expected to uphold the highest standards of ethics and integrity throughout the Hackathon.
2. All work submitted must be original and developed by the participant or their team.
3. Plagiarism, or the use of someone else's work without proper attribution, is strictly prohibited and would result in immediate disqualification.
4. Participants must ensure that their solutions are created from scratch and not copied from existing projects or code repositories.
5. Moreover, the use of any external resources or pre-trained models should be clearly cited, and proper permissions should be obtained where necessary. Adherence to these ethical guidelines ensures a fair and competitive environment for all participants.
6. By registering for this Hackathon, participants are giving an undertaking to adhere to all plagiarism and ethical guidelines set forth by the IndiaAI.