Proforma for submission of nominations for
IndiaAI Fellowship under the IndiaAI Mission

PART I: DETAILS OF STUDENT & PROJECT

1. Name and Address of the Academic Institute: Indian Institute of Technology Madras, Chennai 600 036, India
2. Student Information
    a. Name of Student: Soumitra Das
    b. Nationality: Indian
    c. Branch: Data Science and Artificial Intelligence
    d. Institute ID Number: DA24M022
    e. Full Time Student (Y/N): Y
    f. For B.Tech. students
        i. CGPA/ Marks Percentage (Till last completed semester):
        ii. Number of AI or related Full time Courses (Till last completed semester):
        iii. Names of the AI or related Full time Courses (Till last completed semester):
    g. Email address: da24m022@smail.iitm.ac.in
    h. Phone Number: 8670234735
    i. Aadhaar Number: 563983309754
    j. Bank Account Number: 39820591047
    k. Bank Name: State Bank of India
    l. IFSC code: SBIN0001319
3. Title of Project: Privacy in AI
4. Project Brief (Impact and Objectives) (Min. 200 Words):

    Recent advancements of AI, day to day improvement of the large language models (LLMs) and agents, the first question that comes in our mind is, Is our personal information safe? Do the models store our personal information? Are our prompts being shared with law enforcement?

    AI systems pose many of the same privacy risks we've been facing during the past decades of internet commercialization and mostly

unrestrained data collection. The difference is the scale: AI systems are so powerful that we have even less control over what information about us is collected, what it is used for, and how we might correct or remove such personal information.

There's the risk of others using our data and AI tools for anti-social purposes. For example, generative AI tools trained with data scraped from the internet may memorize our personal information, preferences, relationship, like, dislike and use it to recommend something we might like or give that information to others so that they can use it against us.

Similarly, our data such as a resume or photograph that we've shared or posted for one purpose being repurposed for training AI systems, often without our knowledge or consent and sometimes with direct civil rights implications.

Another type of problem is biases of AI. For example, Amazon built an AI enabled hiring screening tool, but it was discovered that it is biased against female candidates. Similarly, There was research for identifying criminals on the basis of facial features, but it was also biased against black men, this is a crucial problem for the ai models.

Everyone is aware of Generative AI, but the question is actually how much "generative" is the AI? Is it really creating something on the basis of the knowledge gained from the data or is it just modifying the existing data?

In this project our main goal is to face these types of problems and modify the algorithms so that the model will only take the knowledge from the data not the information. So that the future AI system can be more trustworthy and ethical, and we can be safe from sharing our information online, instead of knowing that the data can be used for the training purpose for some AI model.

5. Project Guide Details
    a. Name: Harish Guruprasad Ramaswamy
    b. Designation: Assistant Professor
    c. Department: Data Science and Artificial Intelligence
    d. Email ID: hariguru@dsai.iitm.ac.in

e. Contact No.: +91 95351 45899

6. Project Co-Guide Details (if any)
    a. Name:
    b. Designation:
    c. Institute & Department:
    d. Email ID:
    e. Contact No.:

I, the undersigned, certify that to the best of my knowledge and belief, the details mentioned above are authentic and correct. I understand that any misstatement or misrepresentation described herein may lead to my disqualification or dismissal by the IndiaAI.

(Student Signature with date)

<To be signed by Project Guide/s attached to the project >

(Signature and Stamp of Project Guide)
Name: Harish Guruprasad Ramaswamy
Designation: Assistant Professor
Date:
Place: