Research Interest Statement – Aranya Saha

My research interests lie at the intersection of **Efficient Deep Learning**, **Trustworthy AI**, and **Secure Machine Learning**, with a focus on optimizing and safeguarding multimodal large language models (LLMs) and distributed learning systems. I aim to develop computationally efficient, privacy-preserving, and certifiably reliable AI frameworks that can be safely deployed in real-world, resource-constrained environments.

As deep learning models become increasingly large and multimodal, they raise critical challenges in terms of efficiency, privacy, and trustworthiness. Training or fine-tuning such models often relies on sensitive or proprietary data, making them vulnerable to data leakage, gradient inversion, and membership inference attacks. At the same time, the enormous compute demands of these models hinder their practical and scalable deployment. My research seeks to address these dual challenges by developing techniques that enable efficient and privacy-preserving training, adaptation, and unlearning in large-scale AI systems.

Relevant Background and Expertise

In my undergraduate thesis, I developed an efficient and diagnostically precise **multimodal medical assistant** that integrates visual and textual understanding for skin disease diagnosis. There, I applied and modified Grouped Relative Policy Optimization (GRPO) and Direct Preference Optimization (DPO) to enhance structured reasoning and efficiency in clinical inference. I focused on optimization for lightweight deployment in low-resource medical environments, which parallels the goals of building resource-efficient, privacy-conscious AI systems.

As a **Machine Learning Engineer** at ACI Ltd., I have led projects that bridge research and deployment, developing LLM-powered tools for document processing and tabular data analysis, including a CV Sorter and Insight Explorer. These systems were optimized for scalable inference while maintaining data confidentiality, reinforcing my commitment to building efficient, secure, and reliable AI pipelines.

Future Research Aspirations

My long-term goal is to become a leading researcher in efficient and secure AI, advancing foundational methods that enable trustworthy distributed learning and privacy-preserving large-scale AI systems. I am particularly motivated by recent works such as ***Source-Free Machine Unlearning***, ***Certified Unlearning without Access to Source Data***, and ***AdMiT: Adaptive Multi-Source Tuning in Dynamic Environments***, which introduce principled methods for dynamic model adaptation, privacy protection, and data-free unlearning.

Ultimately, my goal is to design adaptive, efficient, and privacy-preserving multimodal AI frameworks that combine architectural innovation with strong theoretical guarantees of security and reliability. Such systems will not only advance the scientific understanding of efficient deep learning but also ensure the safe and equitable deployment of intelligent models in real-world applications.