Research Plan Outline

My future research will be in the realm of Privacy-Preserving Machine Learning (PPML) and Explainable AI (XAI) and will tackle critical challenges in both. Specifically, I plan to:

1. Design Privacy-Preserving XAI Models: I will examine how to integrate privacy-preserving techniques, such as differential privacy and federated learning, into explainable AI models so that the models are not just secure and privacy-preserving but also interpretable and transparent.

1. Create Explainable Deep Learning Systems: My vision is to create deep models that are not only accurate but also explainable. I will focus on making these models more useful and accessible to end-users, especially in applications involving high stakes such as healthcare, finance, and assistive technologies.

1. Enhance Model Transparency and Trust: I will study the application of concept bottleneck models, attention, and layer-wise relevance propagation in order to enhance transparency for complex models. This will help make AI decisions more understandable to non-technical users and establish trust in these machines.

1. Application to Actual Problems: The models developed will be tested in real-life scenarios, particularly in sensitive and privacy-critical environments. Real-world applications, such as increased accessibility for differently-abled users, improved decision-making in healthcare, and transparency and fairness in AI-based financial systems, are the fields I will target.

By doing this work, my objective is to contribute to the development of trustworthy, secure, and simple-to-use AI systems that can be employed in practical applications without harming performance or privacy.