Machine Learning in Safety-Critical Domains

In recent years, machine learning algorithms have started influencing every part of our lives, including those that are safety-critical, e.g., robotics, automated vehicles, medical diagnostics, etc. Therefore, we must consider the safety of these systems involving machine learning modules. Unfortunately, there are several characteristics of machine learning that can impact safety or safety assessment.

- Powerful machine learning models like deep neural networks (DNN) are considered non-transparent and behave like a black box. Therefore, it is difficult for humans to understand the reasoning behind predictions made by these models, and assess their reliability.
- A machine learning model typically does not operate perfectly and exhibits some error rate. Furthermore, although error rate of a machine learning model can be estimated with respect to the test data , there is only a statistical guarantee about the reliability of this estimate.
- Supervised and unsupervised learning based machine learning models are trained using a subset of possible inputs that could be encountered operationally. Thus, the quality of the data would significantly impact the safety of the resulting models.
- Machine learning models like DNN are typically trained using local optimization algorithms, and there can be multiple optima. Thus, even when the training set remains the same, the training process may produce a different result. This characteristic makes it difficult to debug models or reuse parts of previous safety assessments.
- Formal verification of machine learning components is a difficult, and somewhat ill-posed problem due to the complexity of the underlying machine learning algorithms, large feature spaces.

In this project, we aim to design and develop techniques to improve the safety of machine learning component itself (e.g., improve model interpretability, design safe fail mechanism) as well as approaches to verify the correctness of systems with machine learning modules.

Expectation from the student:

- Good mathematical background
- Good probability and statistics background
- Good programming skills (python, C,C++)