

Student Assistant for the Security of Machine Learning

Machine learning (ML) is currently progressing in various areas, e.g., image recognition, autonomous driving, medical applications as tumor diagnoses, security tasks as network intrusion detection.

With the increasing application of ML systems, a number of security, privacy, and functional challenges are posed on the design and implementation of the underlying algorithms and systems. Security threats include trojaned Neural Networks causing them to misbehave in certain situations, or attacks that disturb the training process of Neural Networks. In addition, other attacks threaten the privacy of ML, e.g. by reconstructing the used training data from a trained ML model. Further, also fairness aspects need to be considered to prevent the ML model from discrimination against humans, e.g., in judicial applications.

For the research projects where we investigate open challenges in ML and develop effective mitigation approaches, we are looking for excellent student assistants, who are motivated to be part of ongoing research in these areas.

Depending on your background and personal preferences, you will be assigned to one of the projects, targeting the before mentioned areas. Your tasks include:

- Perform experiments for centralized, as well as collaborative ML to analyze the security of ML
- Implement attack on ML, threatening the security, privacy, and fairness
- Implementing defenses against the aforementioned attacks
- Develop new attack- and defense algorithms to strengthen the security of ML

Requirements

- Good knowledge in computer security and privacy, as well as Deep Learning
- Experience with Python
- Recommended: Experience with ML libraries in Python, e.g., Pytorch or Tensorflow
- Good analytical capabilities
- Motivation and capability to perform independent work as well as willingness to work in a team

Contact Persons

If you are interested in this highly trendy topic, contact Phillip Rieger (phillip.rieger@trust.tu-darmstadt.de) or Hossein Fereidooni (hossein.fereidooni@trust.tu-darmstadt.de) via mail to learn more about it. Please include a brief overview of your study background and a transcript of records.