# Master project at NXP Semiconductors

# Security for Machine Learning

Nowadays, more and more problems are solved by machine learning (ML) algorithms. One aspect of such an algorithm is the machine-learning approach used. Popular examples are neural networks and support vector machines. These approaches are generic in nature and are tailored to solve a particular problem via its model. Such a model is obtained via a training algorithm that uses labeled training data.

In this project we focus on classification problems where it is the task to decide for input data, such as an image, to which category from a small set of categories it belongs. For such problems, the labeled training data consists of input data together with its correct class.

An example use-case for such a classification problem is anomaly detection, where the ML algorithm has to decide whether the behavior of a device is suspicious, which may indicate that it is under attack. Another example is image recognition, which can be used for autonomous driving.

A well-known problem for ML-algorithms is their vulnerability to adversarial examples. This means that an adversary carefully constructs or modifies an input of the ML algorithm with the objective that the algorithm misclassifies the input. Such adversarial examples can be a serious threat for ML applications. Imagine, for instance, that a stop sign is misclassified as a yield sign in case some graffiti is sprayed on it.

In the literature, several techniques have been presented to construct adversarial examples as well as to defend against them. Recently, a repository has been opened, called Clever Hans, with open source software implementing attacks as well as defenses.

The end-result of this master project is to evaluate the efficiency and effectiveness of the techniques offered in this repository and to study possibilities to improve them.

Your profile:

Computer science student

Knowledge of security

Good Python skills

The project will take six months, including the writing of a final thesis report. It will be carried out onsite at the High Tech Campus in Eindhoven, The Netherlands, due to the required high-intensity knowledge transfer and supervision, as well as the availability of specific software tools.

Working in this project at NXP Semiconductors in Eindhoven implies working in a stimulating, multidisciplinary environment at the forefront of technology, with knowledgeable colleagues, and an excellent infrastructure.

Contact information:

Wil Michiels

High Tech Campus 46

5656AE Eindhoven

NXP Semiconductors

Wil.michiels@nxp.com