Deep Learning Tutorial - Proposal

Team members: Dinah Rabe
Johannes Halkenhäußer
Victor Möslein
Benedikt Ströbl

Outline of the preferred topic: Adversarial Machine Learning

Tentative Title	***Adversarial Machine Learning – the Concept, the Issues, and What We Need to Do About Them.***
Outline	For our tutorial and introduction to the topic of Adversarial Machine Learning (AML), we would like to structure our project along the following questions: **1.** What is AML and why does it cause policymakers to report a strong need for countermeasures to protect ML systems, especially in industrial applications and critical infrastructure? **2.** What are the most popular types of attacks in AML and how do they work on a theoretical level? **3.** How can we implement an exemplary attack on a Regression model? What can countermeasures to such an attack look like and how are they implemented? **4.** Is AML already part of current AI legislation and, if so, what is covered? What does the outlook in this subfield of AI look like from a technical perspective? What do scholars suggest? From the points above, we would focus on covering 1,2, and 4 during our presentation when introducing the key concepts of AML. During this presentation, we would also prepare resources explaining the key types of attacks and give an outlook into the technical, political, and legislative dimensions of AML. For interested fellow students that want to dive deeper into the issue, we will provide a curated list of resources that we think are helpful for guided self-studying.
Resources	• Adversarial Machine Learning-Industry Perspectives (Kumar et al. 2020) • Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning (Jagielski et al. 2018) • Adversarial Robustness - Theory and Practice (Kolter et al. 2018)

Outline of the alternative topic: Uncertainty in Neural Networks (Bayesian NNs)

Tentative Title	***An Introduction to Bayesian Neural Networks***
Outline	We motivate the policy relevance with Sun et al (2019) and explain that uncertainty can stem from stochasticity in the data generation environment and uncertainty in model set-up (epistemic uncertainty). We then relate a bayesian approach to the problem and explain how to introduce priors into a deep neural network. We suggest a process how to identify appropriate prior (distributions) and how they affect models. We go from demonstrating this in a very simple neural network and move on to more complex architectures. The tutorial will showcase how changing priors will change the weights and biases and consequent posterior distributions. These posterior distributions will allow us to showcase the uncertainty in the model compared to point estimates generated by "regular" NNs.
Resources	• Using Bayesian deep learning to capture uncertainty for residential net load forecasting. (Sun et al. 2019) • Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. (Hüllermeier et al. 2021). • Online University Course: https://www.cs.tufts.edu/comp/150BDL/2019f/schedule.html • Online Tutorial: https://jorisbaan.nl/2021/03/02/introduction-to-bayesian-deep-learning.html