Graduate School Event

Thesis Defense: On the Utility and Effects of Efficiency in Artificial Neural Networks

Eugenia Iofinova (Alistarh Group)
Alistarh Group

Host: Samara Ren

As neural-network-based models grow both in size and popularity, interest has grown in making the models smaller and more efficient to train. To that end, many methods have been proposed to prune models by reducing their number of nonzero parameters. Additionally, parameter-efficient fine-tuning, in which a much smaller number of parameters than the total contained in the model is updated during training, has become very popular, especially in the space of Large Language Models. At the same time, the increasingly routine deployment of machine learning in real-world applications has spurred a drive to make them more trustworthy - in the sense of, among other things, being unbiased, interpretable, and editable. In this thesis, we examine the interplay between efficiency and trustworthiness.First, we analyze the effects of model pruning on bias in computer vision models, demonstrating that increased sparsity leads to greater bias, largely as a function of increased model uncertainty in marginal cases. Based on this observation, we propose several bias mitigation techniques. Then, we demonstrate that example-specific model pruning can improve model interpretation methods while improving pruning efficiency to make example-specific model pruning feasible in real time. Then, we investigate the effectiveness of parameter-efficient and data-efficient model personalization via fine-tuning, demonstrating that it is highly feasible with very small computational and data resources. Finally, we consider efficiency in editing model knowledge using a custom synthetic data framework, demonstrating that parameter-efficient, low-rank fine-tuning frequently outperforms full-rank fine-tuning, and, additionally, restricting fine-tuning to specific model blocks frequently improves results. Together, the results in this thesis provide new insights and techniques for combining trustworthiness and efficiency during neural network inference and training.

Monday, April 20, 2026 10:00am - 11:00am

Office Bldg West / Ground floor / Heinzel Seminar Room (I21.EG.101) and Zoom

This invitation is valid as a ticket for the ISTA Shuttle from and to Heiligenstadt Station. Please find a schedule of the ISTA Shuttle on our webpage: https://ista.ac.at/en/campus/how-to-get-here/ The ISTA Shuttle bus is marked ISTA Shuttle (#142) and has the Institute Logo printed on the side.