Applied Machine Learning

Preliminary Meeting (IN2106, IN4192)

Lecturer: Prof. Dr. Stephan Günnemann & Dr. Leo Schwinn

Summer Term 26

Team

- Prof. Dr. Stephan Günnemann & Dr. Leo Schwinn
- Leon Götz, Fabio Rosenthal, Johanna Sommer
- With Pruna AI and BMW

This is a practical course (Praktikum) for **Master**'s students!
Name of module: Applied Machine Learning (IN2106, IN4192)

Website: https://www.cs.cit.tum.de/en/daml/lehre/sommersemester-2026/applied-machine-learning/

Why attend our ML lab course?

1. Opportunity to **implement and apply** state-of-the-art ML algorithms
2. Gain **hands-on experience** working on **real-world data**, solving **real-world tasks** by working on projects offered by our **industry partners** as well as **academic projects**
3. Work on **large-scale problems** with the support of our **GPU computing resources**

Organization – Structure

- Groups of 3 students
- We offer 4 different projects
- Students get access to our GPU servers, each with
  - 4x NVIDIA GPU with 11GB RAM
  - 10-core CPU
  - 256 GB RAM

  $\rightarrow$ Scale up your models and data!
- In some projects, students will have API access to models, like ChatGPT

Organization – Course

- Where and when?
  - Room: Forsoft/Seminar room (5611.EG.038)
  - Date: 13.04.2026 first meeting
  - Time: 2pm to 4pm
- Bi-weekly course meetings (around 2 hours)
  - in person
  - All groups present their work
  - Each group should briefly report their progress and next steps
- Bi-weekly group meetings
  - with advisor and industry partner
  - analyze results, plan next steps
- Regular documentation of your work on wiki
- Code on git (gitlab.lrz.de)

Projects – Industrial and Academic

Redteaming Large Language Models
(with BMW, 1-2 Groups, TBD)

Toward Efficient World Models
(with Pruna AI, 2 Groups)

Redteaming Large Language Models (with BMW)

- Large Language Models (LLMs) are vulnerable to adversarial attacks, such as prompt injection [1] or jailbreaking [2]
- You will attack a large language model to trigger unwanted behavior (e.g., toxic responses) using state-of-the-art attack methods
- At the end of the project, you will give a preliminary assessment of the robustness of the model

[1] Zou, Andy, et al. "Universal and transferable adversarial attacks on aligned language models." arXiv preprint arXiv:2307.15043 (2023).

[2] Chao, Patrick, et al. "Jailbreaking black box large language models in twenty queries." arXiv preprint arXiv:2310.08419 (2023).

Toward Efficient World Models

- World models are generative models that learn to simulate how an environment evolves over time, for example, by predicting future video frames based on past observations and actions (e.g., Google's Genie 3 [1]).
- Students will focus on optimizing world models for faster and more efficient inference, with the goal of enabling high-quality, real-time performance. The goal is to make these models **significantly faster at inference time** with **minimal or no perceptual quality loss**.
- Students are free to explore a wide range of optimization techniques, such as attention and memory optimizations, caching, temporal sparsity, token pruning, quantization, or system-level acceleration.
- Model optimization will be driven by continuous evaluation in iterative cycles, and the final models will be compared using a human evaluation benchmark focusing on realism and temporal consistency.

[1] https://deepmind.google/models/genie/

Registration for the respective topics

Use the following form to state your preference for **all** topics - we will assign based on preference if possible
(1 highest preference, 5 lowest preference)
`https://forms.gle/w6XfU1qxzJVR9bHT8`

Deadline 01.04.2026

If you have any questions, please let us know at:
l.schwinn@tum.de

See you soon!