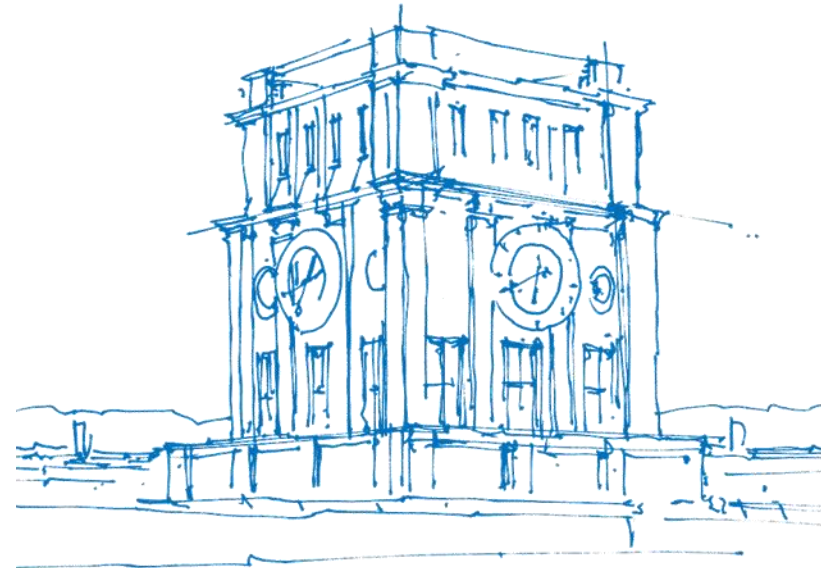# Preliminary Meeting of the NLP Lab Course SS2024

Master Lab Course - Machine Learning for Natural Language Processing Applications (IN2106, IN4249)

Miriam Anschütz, Tobias Eder,
and many more

Prof. Dr. Georg Groh

Research Group Social Computing, Department of Informatics,

Technical University of Munich

31.01.2024

Outline

1. Requirements
2. Registration
3. Procedure
4. Project examples
   - Explainable AI
   - Text simplification and summarization
   - Evaluating correctness of generated text
   - Green and efficient AI

Requirements

Minimum:
- Master student in computer science, data engineering, or "alike"
- Good enough English skills
- Basic programming and machine learning knowledge

Important:
- Hands-on experience in Python, especially Pandas and Numpy
- Basic knowledge about artificial neural networks
- Basic knowledge about natural language processing

Optimal:
- Practical experience with Deep Learning frameworks, such as PyTorch, Tensorflow, Huggingface, etc.

Registration

- Until **14 February**, fill out the registration form

- Your entries are considered when ranking the interested students for the course.
- From **09** to **14 February**, you also have to register for the course on the matching system.
- End of **February**, you will (probably) be notified by the matching system about the status of your participation.
- We will get in touch with you in March for the following steps.

Procedure

Project teams:

- You are going to work in teams of 2 or 3 people on one project topic.
- You can choose with whom to work with the project topic.
- Every project member has to report and work equally (no dirty business!).

Procedure:

- There will be one kickoff meeting at the beginning of the semester.
- There are going to be bi-weekly consulting and progress report sessions.
- You have to be part of a poster session and hand in a report at the end of the semester.

Everything else will be announced at the beginning of the semester.

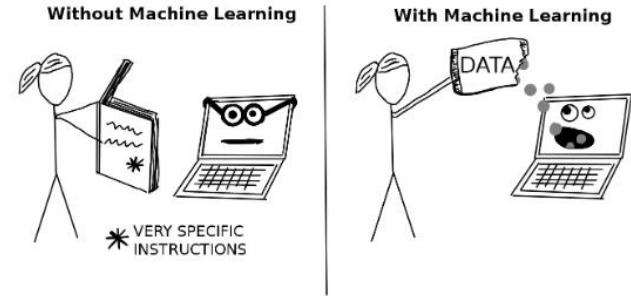# Projects– Explainable AI for Machine Learning

Simon Malberg, M.Sc.

Learning from data is powerful, but at what cost?

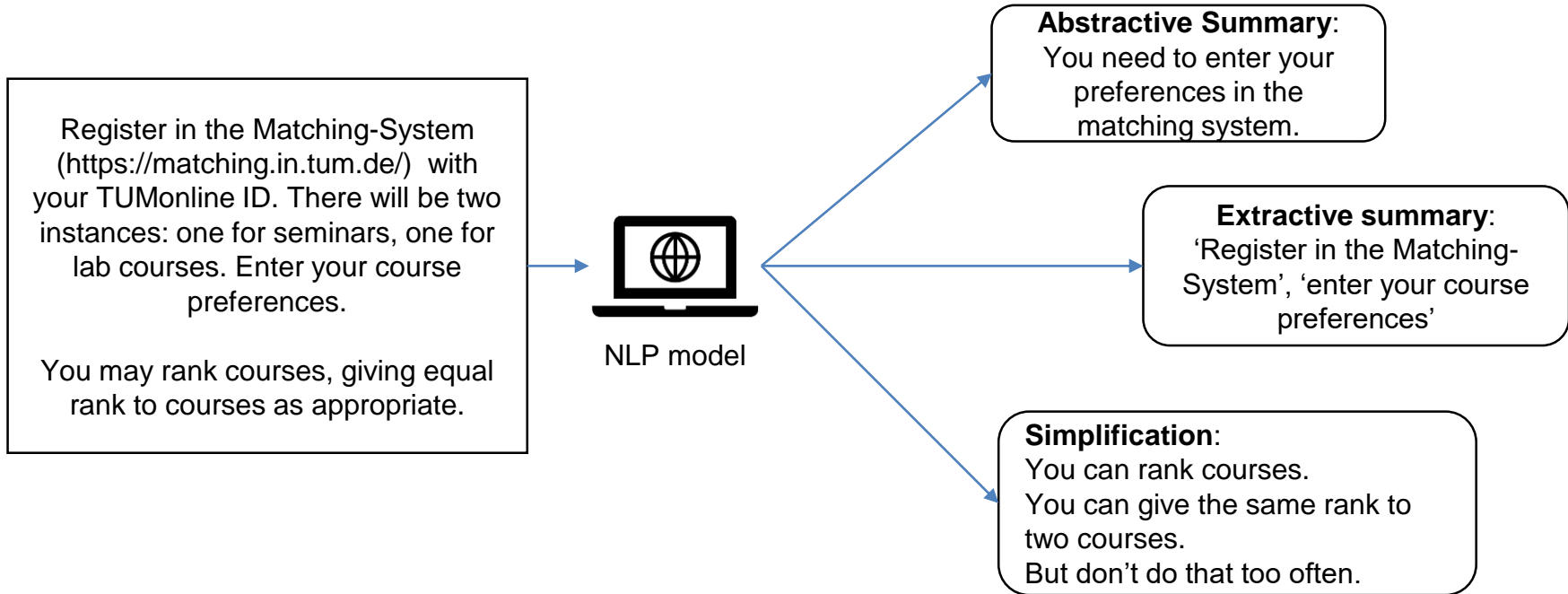Input Data → Black-Box Model (e.g. DNN) → Final Prediction
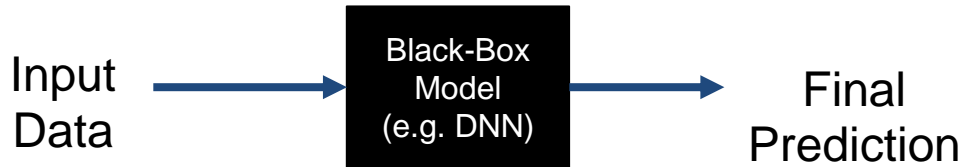
- Models are harder to debug and comprehend
- Models can be biased and unfair
- Models are less accepted by society
- Models can't be deployed in high-stake scenarios

Projects – Text summarization and simplification

Miriam Anschütz, M.Sc.; Ahmed Mosharafa, M.Sc.

Register in the Matching-System (https://matching.in.tum.de/) with your TUMonline ID. There will be two instances: one for seminars, one for lab courses. Enter your course preferences.

You may rank courses, giving equal rank to courses as appropriate.

NLP model

Abstractive Summary: You need to enter your preferences in the matching system.

Extractive summary: 'Register in the Matching-System', 'enter your course preferences'

Simplification:
You can rank courses.
You can give the same rank to two courses.
But don't do that too often.

Projects – Evaluating correctness of generated text

Miriam Anschütz, M.Sc.

Projects – Green and Efficient AI

Jeremias Bohn, M.Sc.

- Language Model size has exploded over the last years
- Increasing size comes with many disadvantages:
    - Longer training times
    - Increased memory consumption
    - Higher energy consumption (and thus higher CO2 emissions
    - Independent researchers struggle to contribute, leading roles are taken by big corporations
- We try to focus on reducing model sizes, making training and inference more efficient and less resource-demanding, and reducing power consumption

Questions?

Registration form: