Research Group Social Computing
Department of Informatics
Technical University of Munich

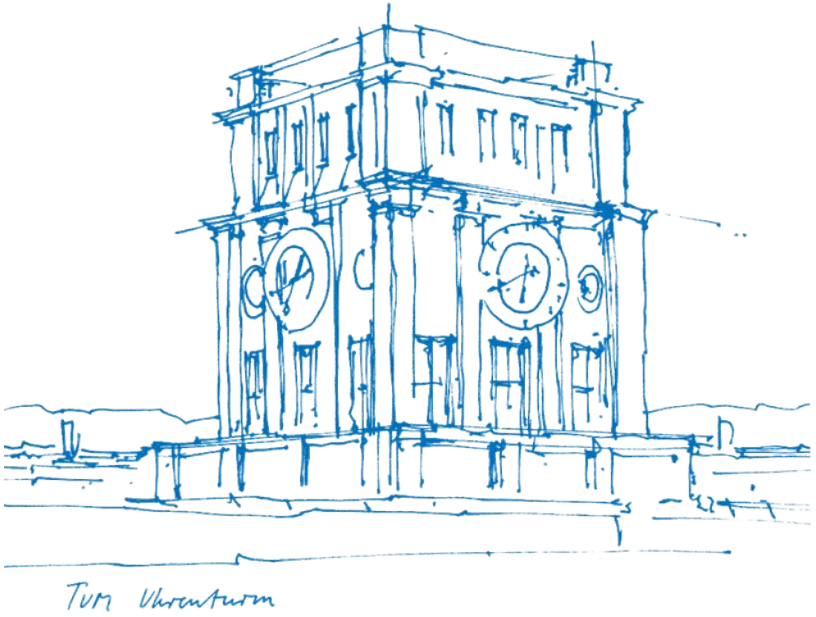
Preliminary Meeting of the NLP Lab Course WS 2024/25

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

Miriam Anschütz, Alexander Fichtl,
and many more

Prof. Dr. Georg Groh

Research Group Social Computing, Department of Informatics,
Technical University of Munich



02.07.2024

Outline

1. Requirements
2. Registration
3. Procedure
4. Project examples
    - Text simplification and summarization
    - Ethical AI
    - Evaluating correctness of generated text
    - Biomedical knowledge enhancement
    - 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 **16 July**, fill out the registration form

- Your entries are considered when ranking the interested students for the course.
- From **12** to **16 July**, you also have to register for the course on the matching system.
- End of **July**, you will (probably) be notified by the matching system about the status of your participation.
- We will get in touch with you in August 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:

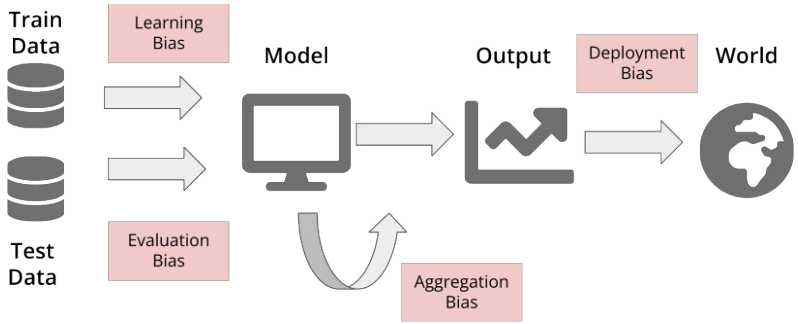
- There will be one kickoff meeting at the beginning of the semester.
- There are going to be bi-weekly consulting and progress report sessions.
- You have to be part of a poster session and hand in a report at the end of the semester.

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

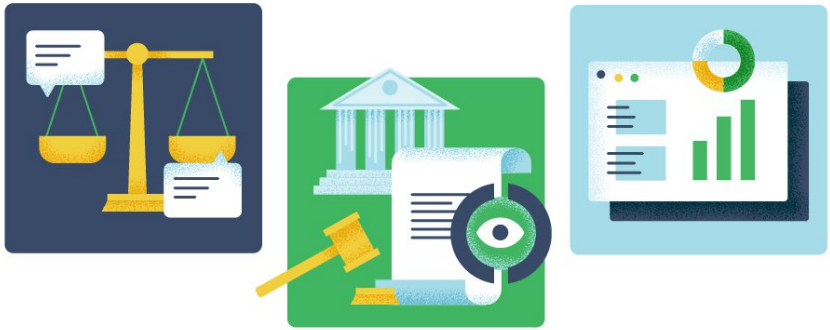
Research Group Social Computing
Department of Informatics
Technical University of Munich

Projects– Ethical AI and Natural Legal Language Processing

Tobias Eder, M.Sc.



- NLP models can be used in a myriad of use-cases
- We experiment with different use cases of NLP that focus on large textual data analysis, multi-modal sentiment and emotion recognition or legal argument mining
- Apart from the technical challenges of implementing these systems we also look at issues of data bias and fairness



Research Group Social Computing
Department of Informatics
Technical University of Munich

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.

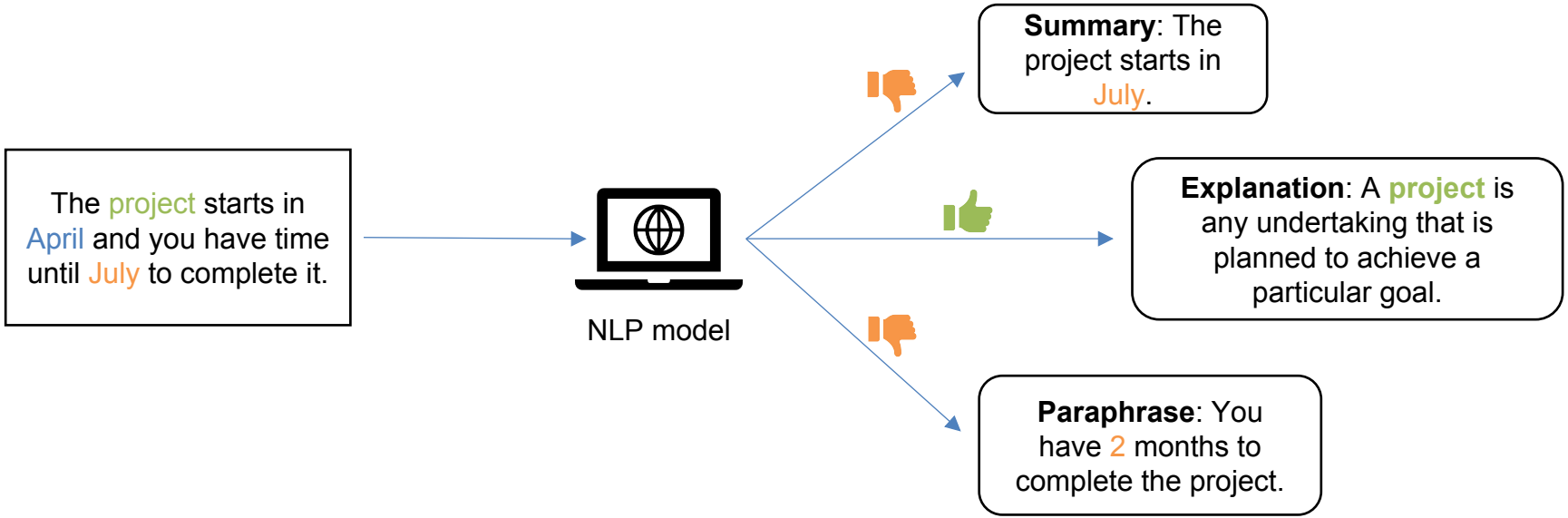
Extractive summary:
'Register in the Matching-System', 'enter your course preferences'

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

Research Group Social Computing
Department of Informatics
Technical University of Munich

Projects – Evaluating correctness of generated text

Miriam Anschütz, M.Sc.



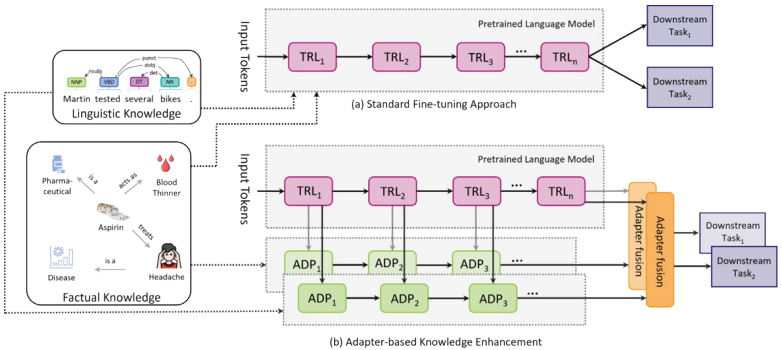
Research Group Social Computing
Department of Informatics
Technical University of Munich

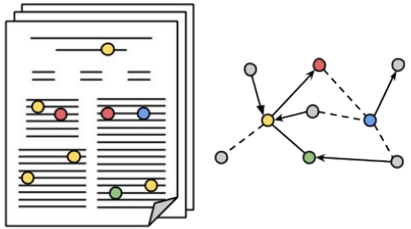
Projects – Biomedical Knowledge Enhancement of LMs

Alexander Fichtl, M.Sc.

- 60% of medical doctors in Germany spend 3 hours or more per day on documentation and administrative tasks
- Many tasks in bio-medicine can be supported by LMs
- Domain specific LMs profit from explicit knowledge injection

➢ You will be using biomedical knowledge graphs to enhance domain specific LMs that can potentially address these issues





BLURB

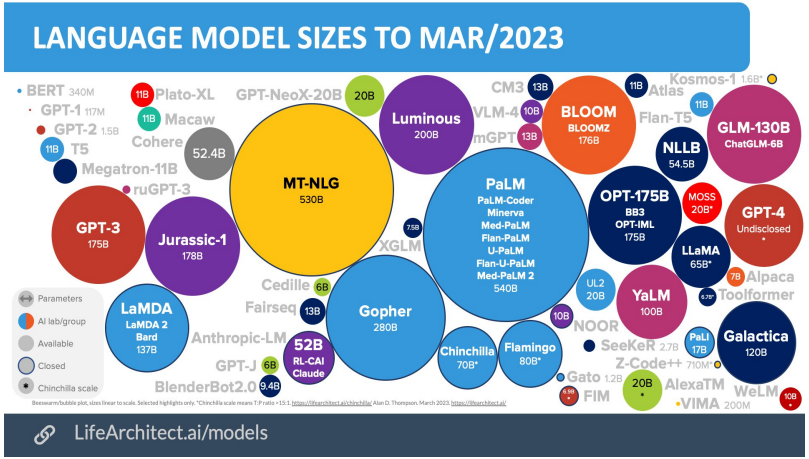
Biomedical **L**anguage **U**nderstanding and **R**easoning **B**enchmark

Research Group Social Computing
Department of Informatics
Technical University of Munich

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: