Research Group Social Computing
Department of Informatics
Technical University of Munich

Preliminary Meeting of the XAI Lab Course WS2023

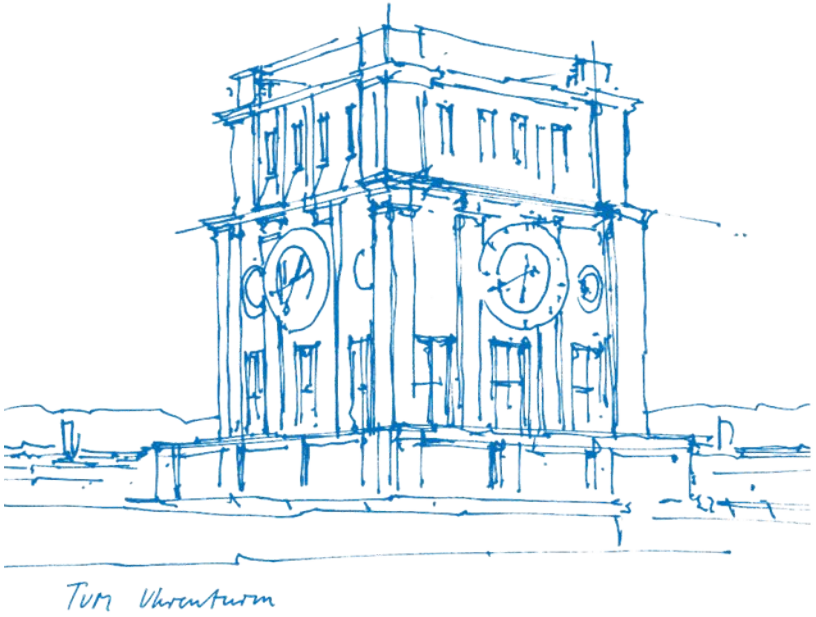
Master Lab Course - Explainable AI for Machine Learning (IN2106, IN4249)

Edoardo Mosca, M.Sc, Tobias Eder, M.Sc. M.A.,
Miriam Anschütz, M.Sc., Daryna Dementieva, M.Sc,
Fabienne Marco, M.Sc..

Prof. Dr. Georg Groh

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

13.07.2022



Outline

1. Requirements
2. Registration
3. Procedure
4. Projects
    - Explainable AI for ML
    - Ethical aspects of ML
    - Text simplification and German Leichte Sprache

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, Pandas, Numpy, and SciPy
- Basic knowledge about artificial neural networks
- Basic knowledge about natural language processing

Optimal:

- Practical experience with Deep Learning frameworks, such as PyTorch, Tensorflow, Theano, Keras, etc.

Registration

- Until **27 Jul**, fill out the registration form

- Your entries are considered when ranking the interested students for the course.
- From **22** to **27 Jul**, you also have to register for the course on the matching system.
- Around the **05 Aug**, you are (probably) notified by the matching system about the status of participation.
- We will get in touch with you in September 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 conduct a final project presentation and report at the end of the semester.

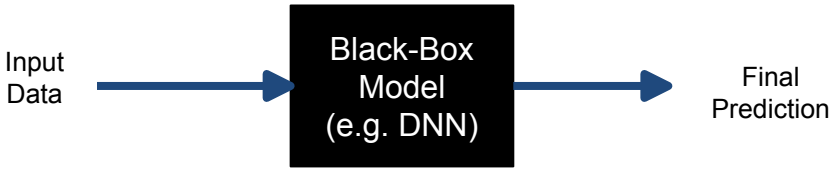
Everything else will be announced at the beginning of the semester.

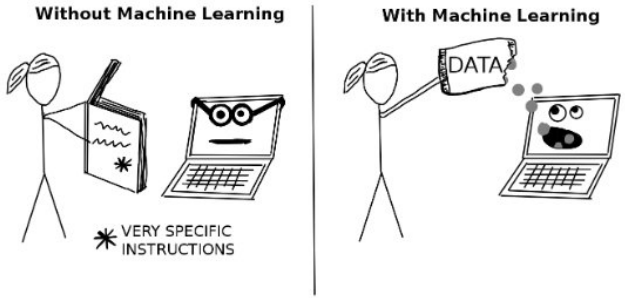
Research Group Social Computing
Department of Informatics
Technical University of Munich

Projects– Explainable AI for Machine Learning

Edoardo Mosca, M.Sc.

Learning from data is powerful, but at what cost?





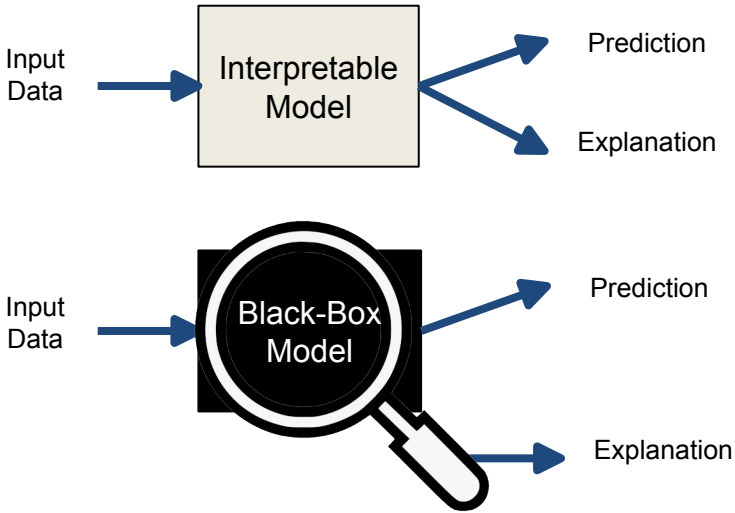
- Models are harder to debug and comprehend
- Models can be biased and unfair
- Models are less accepted by society
- Models can't be deployed in high-stake scenarios

Research Group Social Computing
Department of Informatics
Technical University of Munich

Projects– Explainable AI for Machine Learning

Edoardo Mosca, M.Sc.

Why becomes as important as **what**. Two possibilities:



Methodology:

- Usage of s.o.t.a methods from Explainable AI and other fields *(BERT, SHAP, ..)*
- Applications in various domains such as NLP and social media.

Potential Topics:

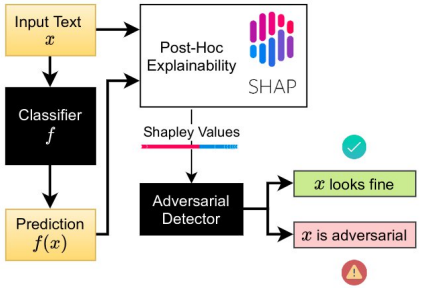
- Knowledge-augmented explanations
- Human-in-the-loop systems
- NLP-tailored explanations
- Building interpretable latent spaces

Research Group Social Computing
Department of Informatics
Technical University of Munich

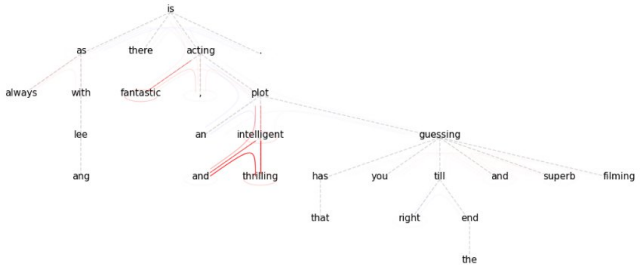
Projects- Explainability Projects from previous years

Edoardo Mosca, M.Sc.

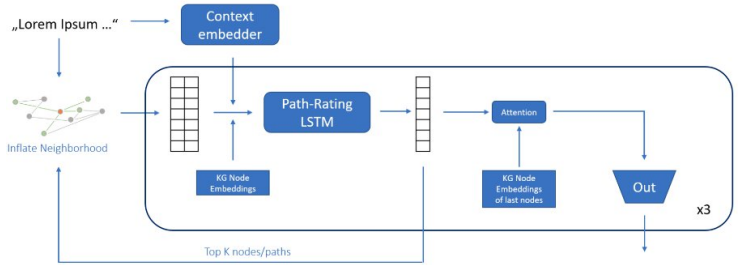
Detecting adversarial examples via SHAP explanations



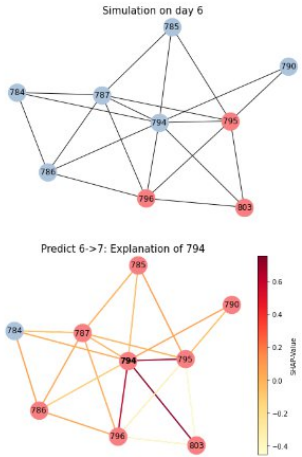
NLP-targeted explanations for words interaction



Knowledge-graph-enriched explanations



Explainable GCN to predict COVID19-spread



Projects– Ethical Aspects of Machine Learning

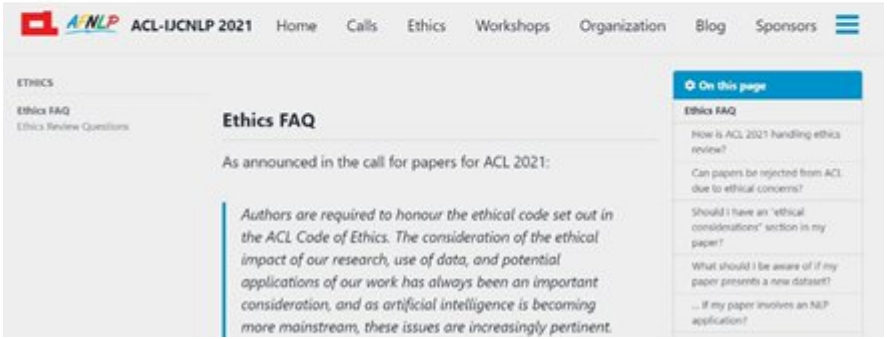
Tobias Eder, M.Sc. M.A.

Ethical Aspects in ML are becoming a big part of AI research.

Some problems include:

- Accountability
- Privacy
- Bias in automated systems
- Disparate impact and fairness
- Possibilities for recourse and influence

This years lab course will focus on various tasks from NLP and other domains. Using explainability methods OR how to leverage those tasks themselves for explaining.

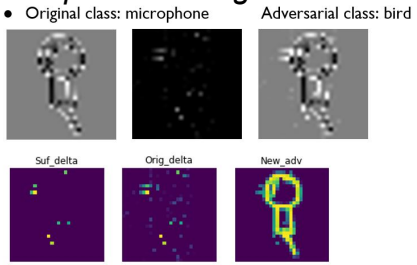


Research Group Social Computing
Department of Informatics
Technical University of Munich

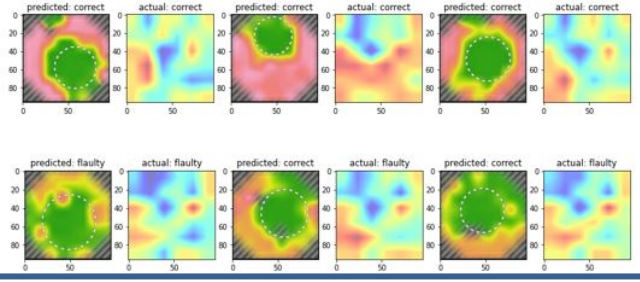
Projects - Explainability Projects from the previous year

Tobias Eder, M.Sc. M.A.

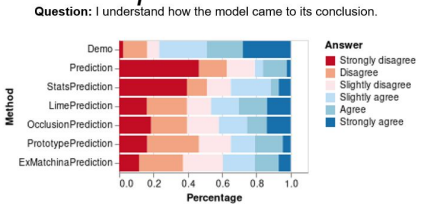
Contrastive explanations through pertinent negatives



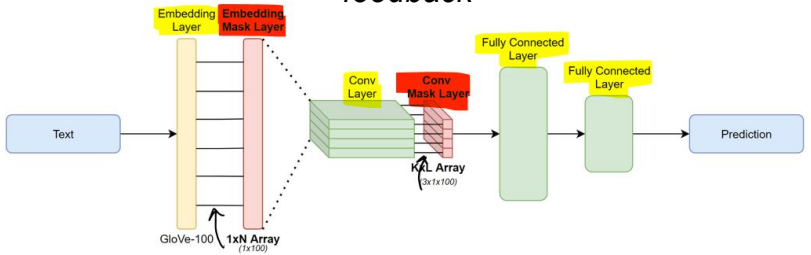
Improving error-detection rates through explanations in welding spots



User-study on the trustfulness of explanations



Better explanations with human-in-the-loop feedback



Projects– Text simplification

Miriam Anschütz, M.Sc.

Not everybody understands everyday language, e.g. ...

- Functional an-alphabets
- Mentally disabled people
- Visual or hearing impaired people

=> **We need simplified language!**

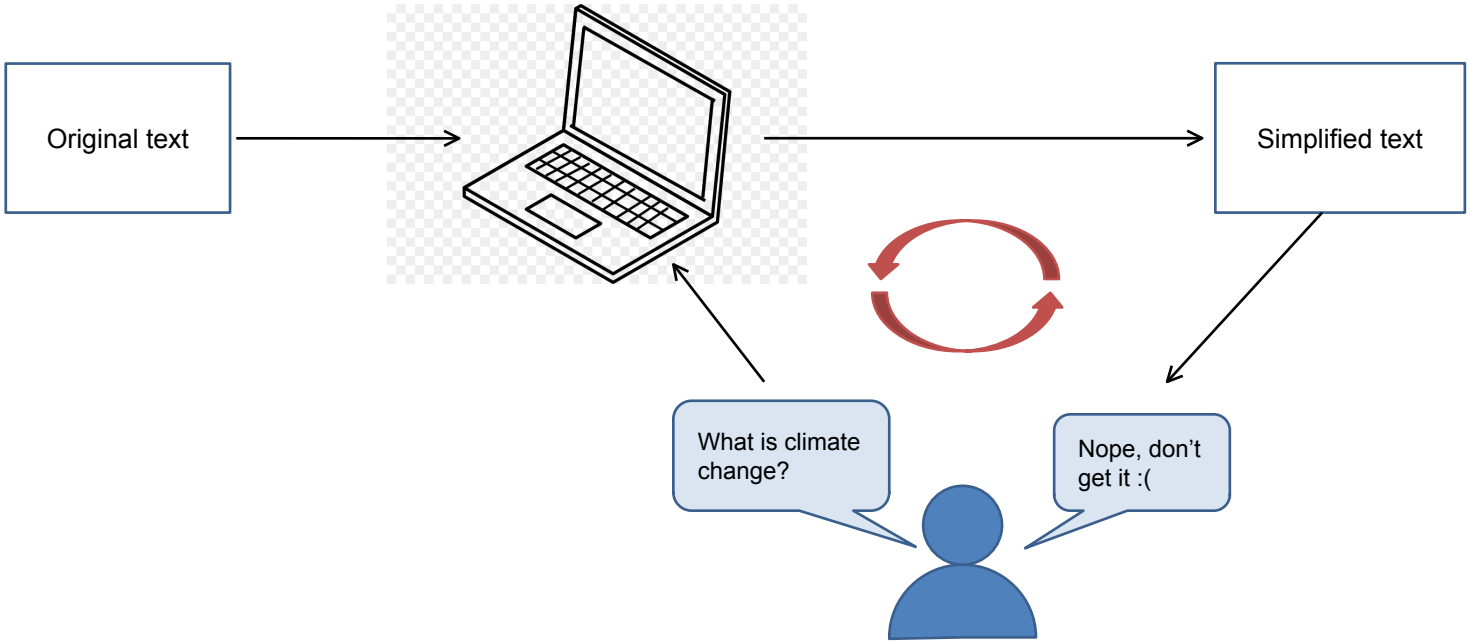
However, manual translation is very costly. Therefore, we need reliable models to simplify texts.

Standard	Plain language / einfache Sprache	Easy language / Leichte Sprache
Climate change poses an immense threat to humanity.	Climate change is dangerous for people.	The climate is changing. It is getting warmer and warmer. This is dangerous for people.
Der Klimawandel stellt für die Menschheit eine immense Bedrohung dar.	Der Klima-Wandel ist für die Menschen gefährlich.	Das Klima ändert sich. Es wird immer wärmer. Das ist für die Menschen gefährlich.

Research Group Social Computing
Department of Informatics
Technical University of Munich

Projects – Incorporating human feedback

Miriam Anschütz, M.Sc.



Questions?