

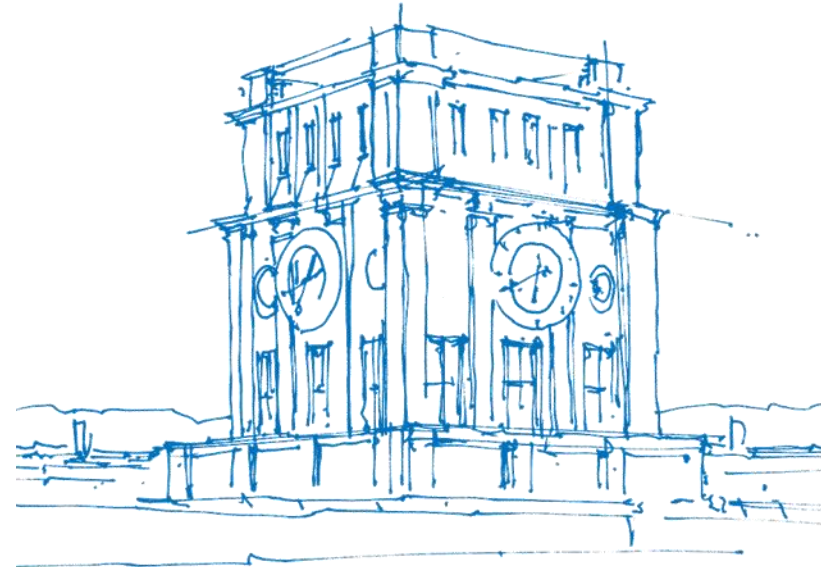
# Preliminary Meeting of the Ethical AI Lab Course SS2021

Master Lab Course – Ethical AI: Problems and Applications (IN2106, IN4249)

Tobias Eder, M.A. M.Sc.,  
Prof. Dr. Georg Groh

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

05.02.2021



*TUM Uhrenturm*

# Outline

1. About this course
2. Your profile
3. Topic examples
4. Applying for a spot

# About this course



This is a brand-new lab course!

We try to connect the practical side of machine learning and AI research with the discussion about the ethics within and around the use of such systems.

Why is this relevant?

Modern machine learning research has started embracing ethics as a core part of their mission statement.

Ethics is no longer just an afterthought when designing and building systems, but integral to their development.

# About this course

## A lab course about Ethics? Isn't that too theoretical?

- Ethical problems are at the core of all the topics we deal with in the lab course.
- It doesn't mean that the topics are focused (solely) on an ethical discussion of the material.
- Instead, we want to focus on fields of research and applications that have an ethical underpinning of some sort.
- They should help us ask questions about ethics in the context of AI and conceptualize systems with their societal benefits and risks in mind.



# About this course

To get some assumptions out of the way:

This is a practical lab course

– that means the focus is on the building and trying out!

Topics will be quite varied

– and for some of them the ethical context might not be immediately obvious

Ethics is not supposed to be the spoilsport of research

– this course is not focused on finger-wagging and warning against the dangers of AI

But we are still interested in engaging with the ethical aspects of the projects

# Semester plan

## Project teams:

- You are going to work in teams of ~3 people on one project topic.
- Topics are going to be implementation work, paired with research. It's all about trying out new things and getting them to work.
- You can either team up ahead of time for a project or get paired with students according to your topic preferences.
- Every project member has to take an active role in the team (no freeloading!).

## Procedure:

- There will be one kickoff meeting at the beginning of the semester to give an intro and present available topics. You can also suggest your own topic beforehand if there is something specific you would like to work on.
- After the kickoff meeting, participants can give their preferences on available topics and will form groups.
- There are going to be bi-weekly consulting and progress report sessions afterwards.
- You have to give a final project presentation and submit a project report at the end of the semester.

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

# Your profile

## Minimum:

- Master student in Computer Science, Data Engineering, or "alike"
- Some programming proficiency in at least one language and knowledge about machine learning basics

## Important:

- Hands-on experience in Python, Pandas and Numpy
- Basic knowledge about artificial neural networks
- Keen on doing both engineering work and looking at the broader impact of such systems

## Optimal:

- Experience with at least one Deep Learning framework, such as PyTorch, Keras, Tensorflow, etc.
- Some experience tackling practical machine learning problems

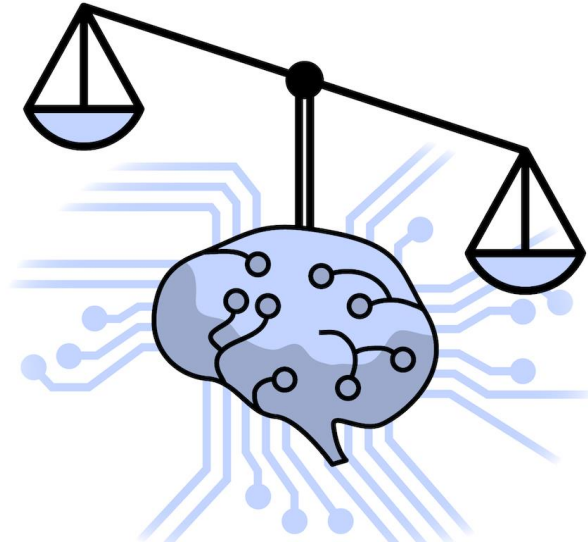
## Topic example – Optimizing Fairness

Deep Learning is often focused on minimizing a training loss, such as RMSE.

This type of optimization has previously been shown to potentially lead to unintended biases in models.

In what ways can systems be optimized for fairness?

Is fairness at odds with a systems general performance?





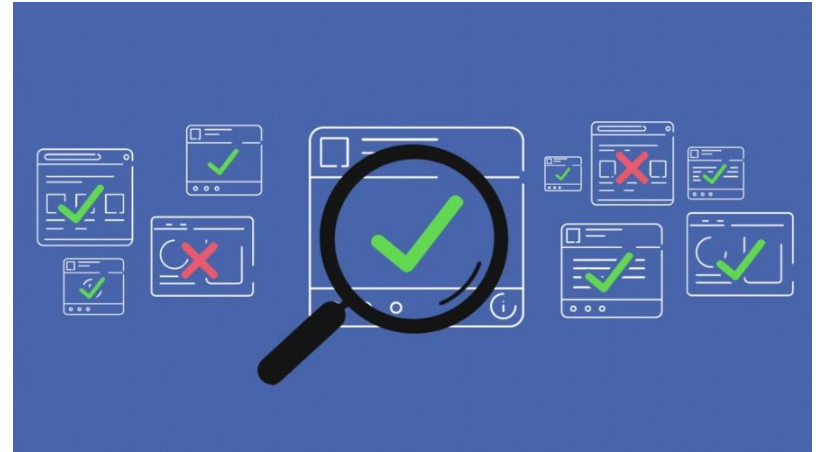
# Topic example – Automated fact-checking and Misinformation

Fact-checking is a complex task for NLP models combining natural language understanding with knowledge representation problems.

A big part of the problem comes from dealing with novel information and systematizing known facts.

How well defined is the fact-checking task? What are alternatives to costly knowledge bases?

Can we get a system to truly ‘reason’ about information?

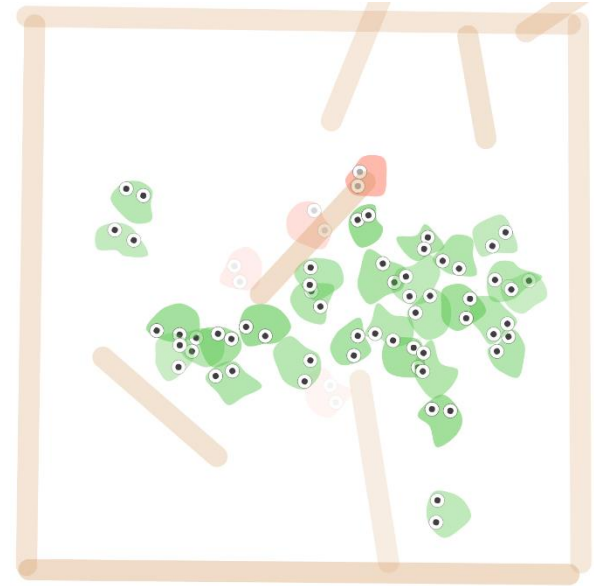


# Topic example – Cooperation and Competition between RL Agents

How can we make agents compete and cooperate in controlled environments?

How do we enforce certain ((un)-ethical) behavior in agents?

Under which circumstances can ‘moral’ behavior emerge from evolutionary algorithms?



source: [blog.otoro.net](http://blog.otoro.net)

# Applying for a spot

- Until **16 Feb**, fill out the registration form:  
<https://wiki.tum.de/display/socialcomputing/Application+Ethical+AI%3A+Problems+and+Applications+SS2021>
- This will be the basis for us doing the matching from our side.
- Until **16 Feb**, you also have to register for the course on the matching system!
- Around the **25 Feb**, you are (probably) notified by the matching system about the status of participation.
- If you get assigned to the course you will be then sent an E-Mail with further steps.

# Questions?