# IN-STK-5000, Open-Ended Project

Christos Dimitrakakis `chridim@ifi.uio.no`

October 28, 2020

Before you start, make sure you have

- Joined one of the project groups.
- Forked the code in `https://github.com/olethrosdc/ml-society-science`
- All questions should go through Padlet or email if urgent

Firstly, visit `https://ai.googleblog.com/2020/02/ml-fairness-gym-tool-for-exploring-long.html` and `https://github.com/google/ml-fairness-gym` and start working on one of the dynamic environments there. In particular, I recommend you study the credit score environment. Using the environments requires the open-AI gym `https://gym.openai.com/`

As a first step, you should place the variables into the following categories:

- Features $x_t$, that e.g. describe the patient $t$ in a medical database
- Actions $a_t$ that describe whatever intervention has been taken, if applicable. There is always at least one action, even if not included in the dataset. E.g. for a diagnostic database, the action would be 'perform the diagnosis'.
- Outcomes $y_t$ that describe what happens after the action has been taken.
- The utility function $U$ that describes what $y_t$ that describe what happens after the action has been taken.

**Simulation selection and analysis (Nov 6)** Here you should describe the simulation, and analyse the data it produces with a simple, default policy for making decisions. You should in particular try to

- Try out a simple, random agent
- Identify causal relations
- Identify sensitive variables and possible fairness metrics.
- Point out privacy issues
- Discuss the utility function

**Improved policies (Nov 20)** Here you should come up with an algorithm that produces improved policies in some sense. Possible things to do include:

- Define a parametric or non-parametric policy for making decisions.

- Perform an analysis of this policy in terms of performance.
- Start a preliminary analysis on fairness and privacy of that policy.

**Final report (Dec 6)** In the final report, make sure to communicate everything related to reproducibility, fairness and privacy of your methodology and the policy you have derived. The final report should include an improved version of your policy and you should concisely discuss

- Reproducibility
- Fairness
- Privacy