Information about the Course Final Assessment

To pass the course, you need to fulfill the following criteria:

- Submit "Lab Zero" to Lisam.
- Pass all labs.
- Pass the hand-in assignment.
- Contribute to the final course project and participate actively in the final course seminar. This includes submitting presentation feedback and posing relevant questions with your team during the seminar.

The final grade will be determined by the average grade earned on the Lab assignments.

Final Grade	Average grade achieved from the 3 labs
3	$3 \leq \bar{X} < 3.7$
4	$3.7 \leq \bar{X} < 4.5$
5	$4.5 \leq \bar{X}$

Aim of Lab 3

- Learn the differences between risk categories on the EU AI Act and imposed obligations.
- Apply your knowledge of different fairness metrics to decision-making and prediction algorithms.
- Learn the differences between explainability techniques.
- Apply your knowledge of explainability techniques to explain algorithms predictions.

Assessment

To pass this lab demonstrate your solutions at the lab.

Requirements	Grade
Tasks 1, 2, 3	3
Tasks 4 or 5	4
All tasks	5

Good Luck!

Task 1

The European Union (EU) AI Act is the first comprehensive legislative framework, which ensures that AI systems are developed and utilized in a manner that protects fundamental human rights and is consistent with the principles of responsible AI.

Specify in the following table (check the correct column that applies), how each scenario will be treated by the EU AI Act.

(For this assignment, I encourage you to use LLM chatbots (e.g. ChatGPT or Perplexity) to learn about relevant sections and different discourse of the EU AI Act.)

	Scenario	Will be encouraged to follow ethical principles	Will face penalties	Will be banned outright
1	A spam filter trained on biased datasets disproportionately flags emails written in non-standard English or from specific cultural contexts as spam.			
2	An AI photo editor designed to enhance images, unintentionally lightens darker skin tones due to biased training data.			
3	An AI-powered CV screening tool that uses low-quality datasets and unintentionally discriminates against candidates based on sensitive attributes.			
4	An AI system for diagnosing complex diseases. The system doesn't provide clear explanations for its diagnoses.			
5	An AI system for determining loan eligibility without maintaining detailed documentation. The system fails to provide clear information to loan applicants about how decisions are made.			
6	An AI system that uses webcams to monitor employees' facial expressions and infer their emotional states throughout the workday.			
7	An AI system that uses facial recognition technology to categorize individuals based on their sensitive attributes to tailor services.			
8	An AI-powered system that assigns scores to citizens based on their social media activity and financial records to determine access to job opportunities.			

Task 2

Download the data for task 2.

1) Load the data using “numpy.load”.
2) The data represents the results of an algorithm credit score prediction for loan approval.
3) The columns of the data represent: [“Protected attribute”, “True credit worthiness”, “Algorithm prediction”].

4) Calculate equal opportunity and equalized error rates by calculating the elements of a confusion matrix. Does the algorithm predict credit scores fairly?

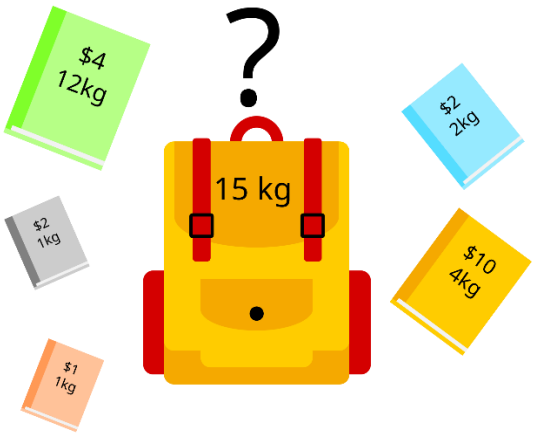
5) Which fairness metric do you find most suitable to consider for this specific case?

Task 3

Write a python code that solves the Rook’s problem for 10 rooks using evolutionary algorithms. Explain in your code how you represented the data, and how you mapped different algorithm steps to model your problem.

Task 4

Write a python code that solves the following Knapsack[1] problem. Make sure your implementation allows for more than one selection of each item as a solution.



[1] https://en.wikipedia.org/wiki/Knapsack_problem

Task 5

Download the data for task 5.

Write a Python code that fits an explainable boosting machine model to the data. Load the data using the following command and consider the last column with the label “target” as the y values.

df = pandas.read_json('UCI Heart Disease Prediction.json')

a) Visualize features importance using model’s global explanation.
b) Visualize a local explanation for the datapoint in row 200.