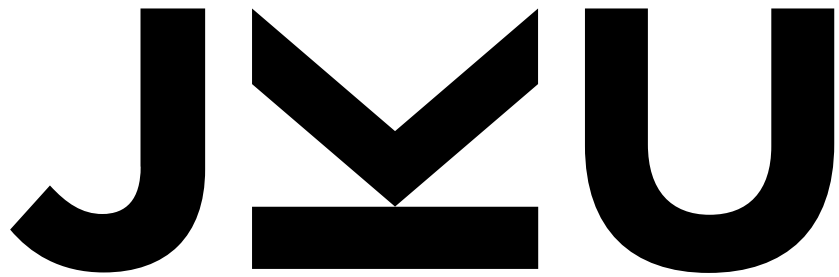


JOHANNES KEPLER
UNIVERSITY LINZ

# UE MLPC 2025: DATA EXPLORATION & CLASSIFICATION

Tara Jadidi, Florian Schmid, Paul Primus

2025-04-28

*Institute of Computational Perception*

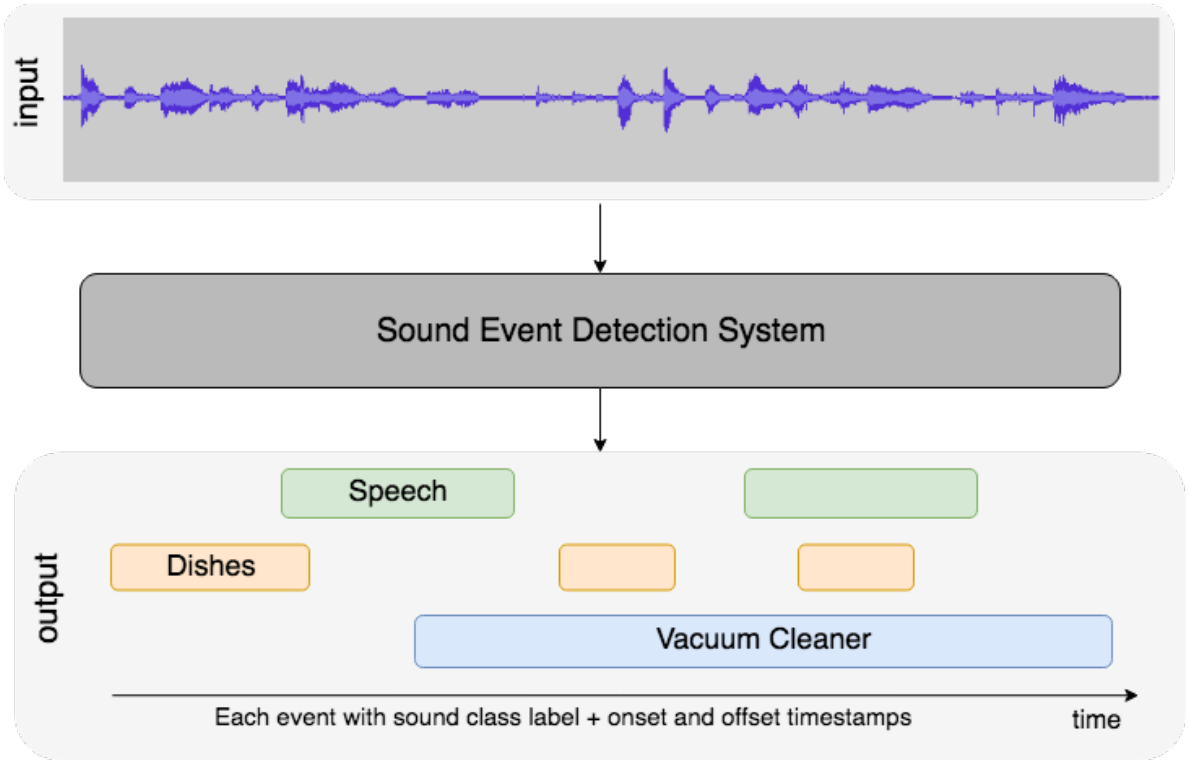
JOHANNES KEPLER
UNIVERSITY LINZ

# Menu for Today

- State of the Project: Where are we?
- Task 2: Selected Presentations
- Task 3: Classification Experiments

# The Project Vision

- Goal: Train models on a general-purpose dataset that can detect a set of arbitrary sound events with their respective onsets and offsets



# WHERE ARE WE?

# Project Schedule

| | | Date/Deadline |
|---|---|---|
| **Meeting 1** | Introduction, explain Tasks 0 and 1 | March 10 ✔ |
| **Task 0** | Form teams | March 24 ✔ |
| **Task 1** | Data Annotation | March 24 ✔ |
| **Meeting 2** | Release dataset, explain Task 2 | April 7 ✔ |
| **Task 2** | Data Exploration | April 24 ✔ |
| **Meeting 3** | Discuss results, explain Task 3 | April 28 ◀ |
| **Task 3** | Classification Experiments | May 22 |
| **Meeting 4** | Present results, release test data, explain Task 4 | May 26 |
| **Task 4** | The Challenge | June 18 |
| **Meeting 5** | Final presentations | June 23 |

# TASK 2: DATA EXPLORATION

# Data Exploration

Goals:

- Analyze the quality of your temporal and textual annotations.
- Find meaningful clusters in the text and audio feature space.
- Connect audio and text features.
- Draw conclusions for the next phase of the project.

# SELECTED PRESENTATIONS

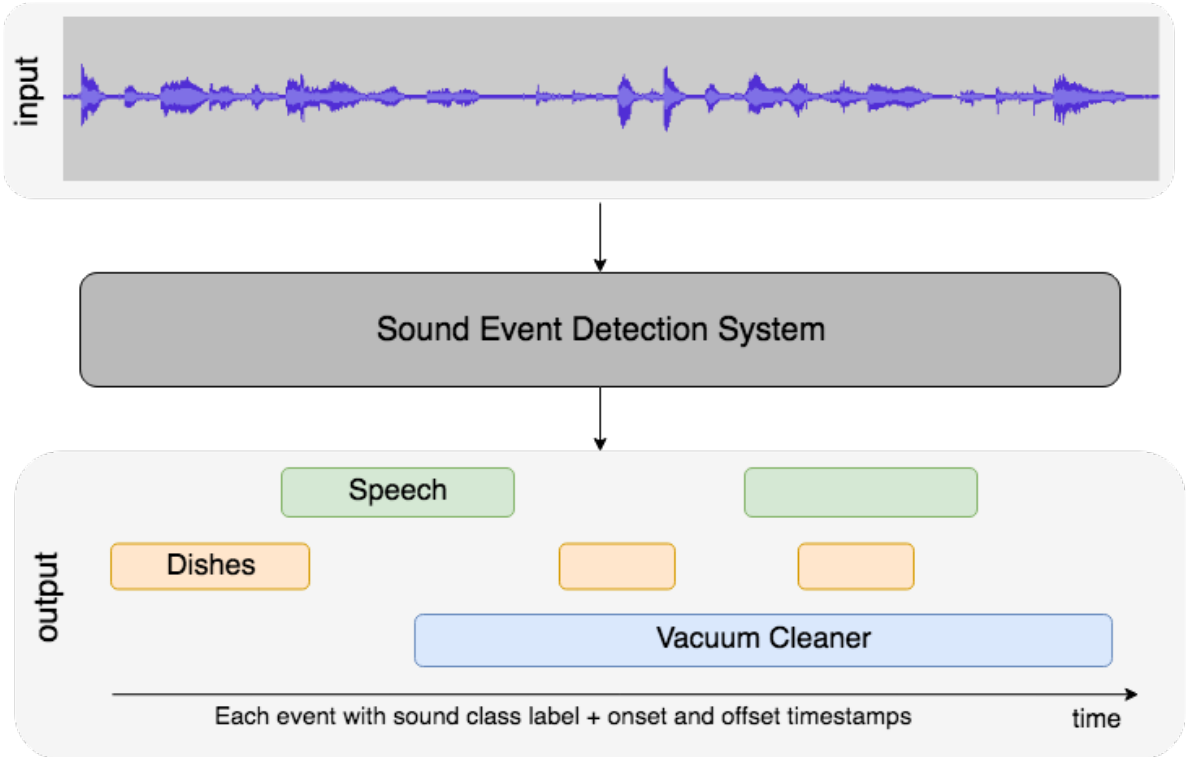
# Presentations

1. **Team Ban**: Annotation Quality
2. **Team Toothpaste**: Audio Features
3. **Team Vegetable**: Text Features
4. **Team Expansion**: Case Study & Conclusion

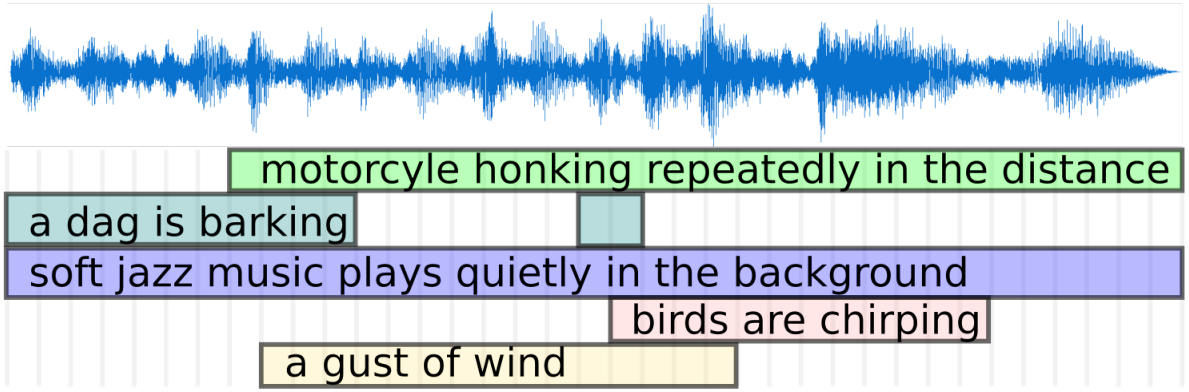
# TASK 3: CLASSIFICATION

# The Project Vision

- Goal: Use our dataset to train sound event detection models for selected event categories.



# The MLPC2025 Data Set

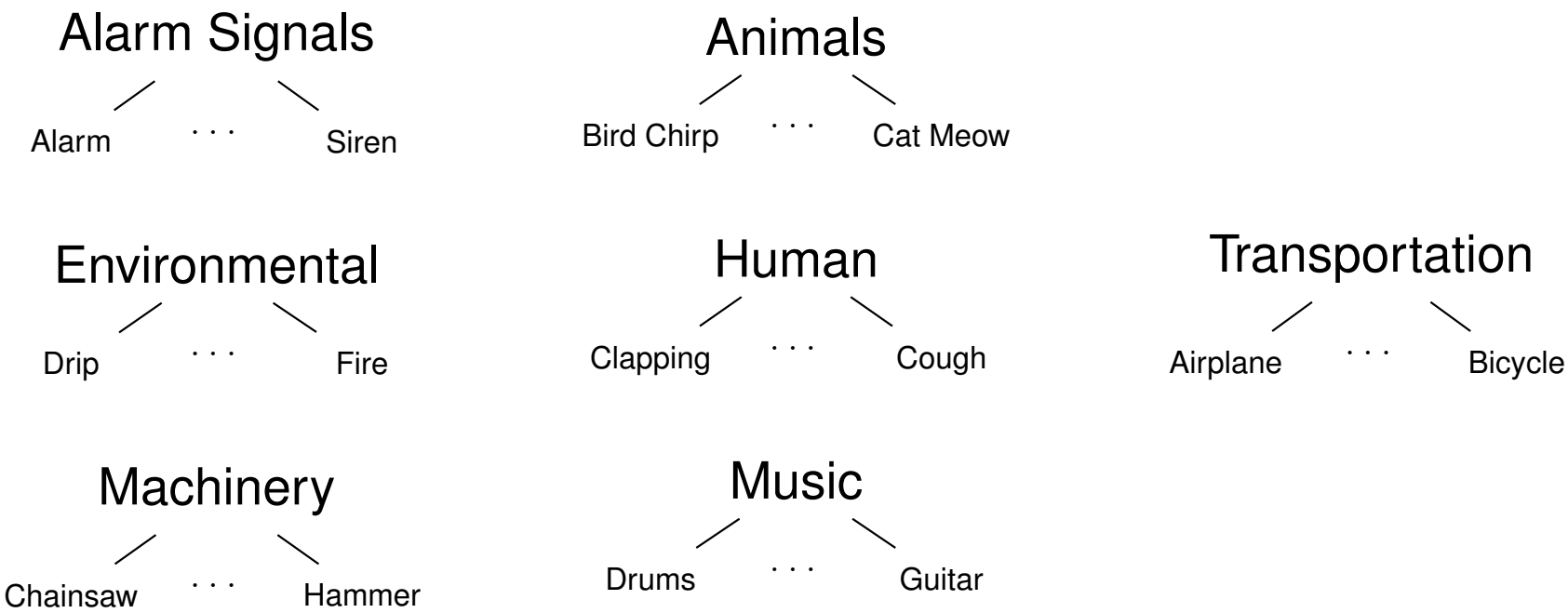



To train classifiers, we need to map each free-text annotation to a predefined class.

- "motorcycle honking repeatedly" $\rightarrow$ motorcycle, honk
- "a dag is barking" $\rightarrow$ dog barking
- ...

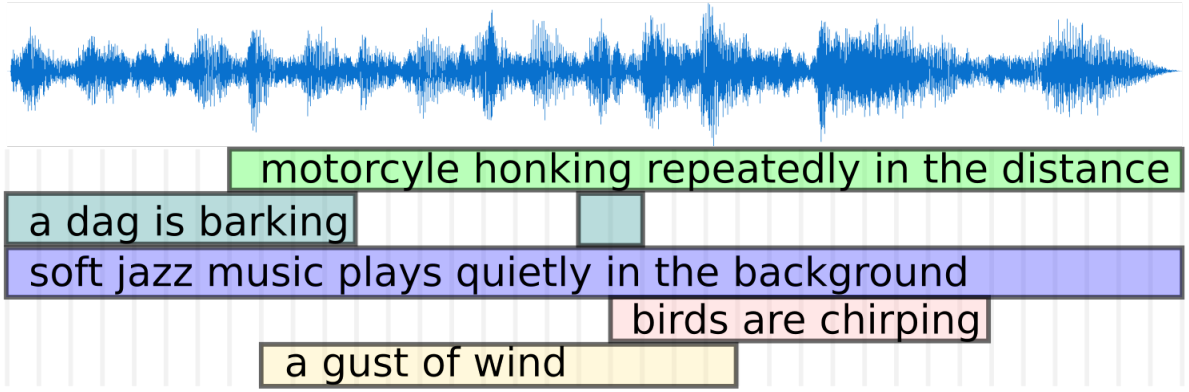
# Step 1: Define Classes

Based on your textual annotations we've identified 58 sound classes for this phase of the project:



# Label Mapping

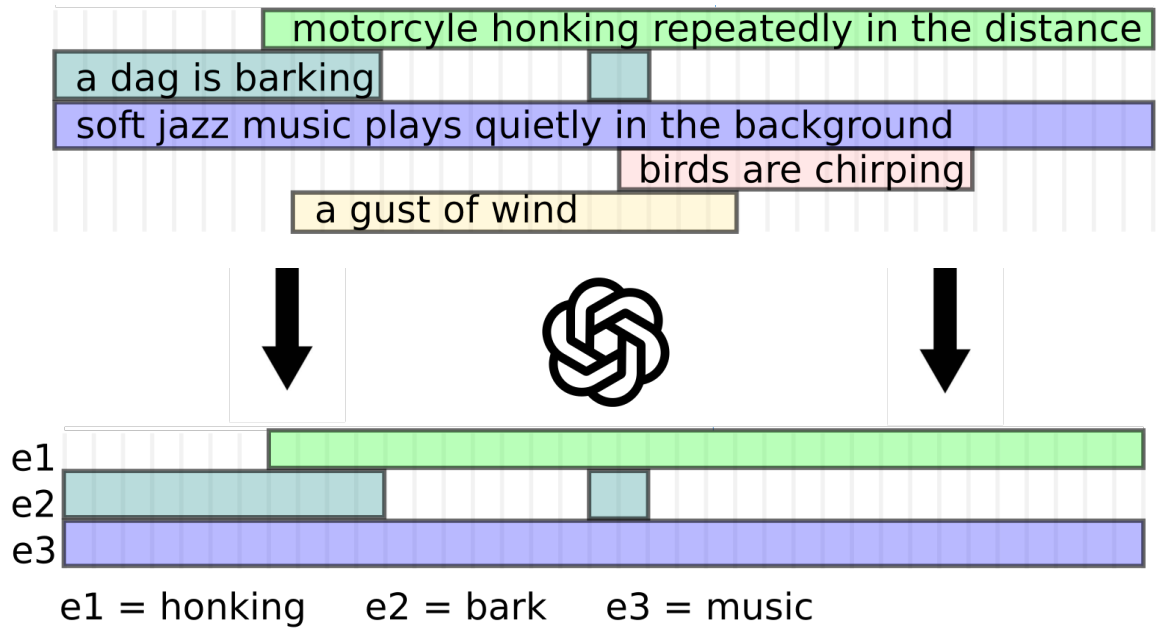
How can we map these free text annotations to a fixed set of class labels?



# Step 2: Map Text to Class Labels (1/3)

We used a LLM to map the text annotations to the 58 classes.

- Each annotation was mapped to one or multiple classes.
- Annotations that cannot be mapped were ignored.



# Step 2: Map Text to Class Labels (2/3)

One label file for each audio recording:

- Keys are the 58 class names.
- Values are 2D-arrays of shape $(\text{time\_steps} \times \text{annotators})$.
  - The sequence is aligned with the sequence of audio features (120ms frames).
  - Individual values correspond to the number of times this class was mentioned by one annotator.

Python snippet for loading labels:

```
labels = np.load(filename)
print("Classes:", list(labels.keys()))
print("Bird Chirp Labels:", labels["Bird Chirp"].mean(-1))
```

More details in the tutorial session next week!

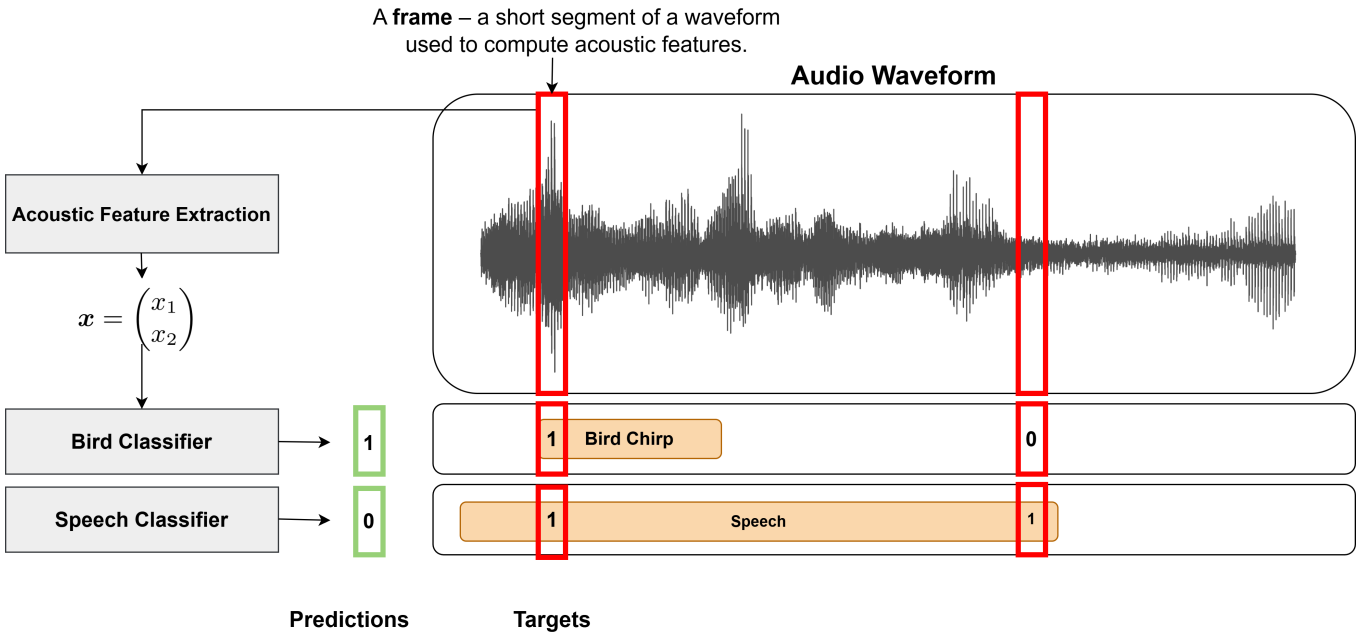
# Step 2: Map Text to Class Labels (3/3)

New dataset:

```
MLPC2025_classification/
|- audio/
|  |- 0.mp3
|  \- ...
|- audio_features/
|  |- 0.npz
|  \- ...
|- labels/
|  |- 0_labels.npz
|  \- ...
|- metadata.csv
|- annotations.csv
|- README.md
```

# Training Classifiers in a Nutshell

Train a separate binary classifier for each of the 58 categories.




- Each frame is represented as a feature vector and can be used as a training example.
- The corresponding label can be found in the label file.

# What to Investigate (1/6)

## Labeling Function[1]

1. How accurate are the labels provided by the labeling function?
2. Which audio features are useful to detect the classes of interest?
3. Do feature vectors that belong to the same class form clusters in the feature space?

[1] Find the detailed questions in the task description!

# What to Investigate (2/6)

## Data Split[2]

1. How did you split the data for your experiments?
2. How did you avoid data leakage across the data splits?
3. How did you obtain an unbiased estimate of the final performance?

[2] Find the detailed questions in the task description!

# What to Investigate (3/6)

## Audio Features[3]

1. Which subset of audio features did you select?
2. Did you apply any preprocessing steps?

[3] Find the detailed questions in the task description!

# What to Investigate (4/6)

## Evaluation[4]

1. Which evaluation criterion did you choose to compare hyperparameter settings and algorithms?
2. What is the baseline performance?

[4] Find the detailed questions in the task description!

# What to Investigate (5/6)

## Experiments[5]

1. For at least three different classifiers, systematically vary the most important hyperparameters and analyze their fitting behavior.
2. What is the final performance of the best model found for each model class?

[5] Find the detailed questions in the task description!

# What to Investigate (6/6)

## Analyzing Predictions[6]

1. Use the spectrogram and the sequence of predictions to visualize the classifier output.
2. Listen to the audio recordings and inspect the corresponding predictions of the classifier.
3. Can you identify problematic conditions that cause the classifier to mispredict classes? Can you think of simple postprocessing steps to improve the prediction performance?

[6] Find the detailed questions in the task description!

# Classification Task: Report

Compile a short technical report that addresses all the previous questions.

- One report per group.
- Template is available on Moodle.
- max. 7 pages (including tables, figures)
  - max. 5 pages of text (excluding tables, figures)

# Classification Task: Slides

In addition to the detailed report, compile a short presentation

- Cover **selected** aspects and sub-questions.
- Your topic is determined by the first letter of your group name (see table below).
- max. 4 slides + 1 title slide

| First letter of group name | Topic |
|---|---|
| A, C, E, M, Q | Analysing Predictions |
| B, F, I, L, N, P | Experiments |
| D, G, J, R, T, U, W | Data Split & Evaluation |
| H, K, O, S, V, Y, Z | Labeling Functions & Audio Features |

Table: Assignment of topics for the slides based on the first letter our your group name.

# Classification Task: Submission

- Submit your report and slide deck as two separate PDF files via Moodle by May 22nd (Thursday), 23:59.
- Selected groups will be asked to present their results in class on May 26th (Monday).
- Presenters will be informed on May 25th (Sunday).
- At least one team member must be available to present in-person or via Zoom.

# Classification Task: Grading

- Completing all tasks **is mandatory** to pass the course!
- The report is worth 37 points and the slides 3 points
- Grading criteria for the report in the task description on Moodle.
- Submitting a day late will cost you $\frac{1}{3}$ of the total points:
  - Up to May 22nd, 24:00: 100 %
  - May 23rd 00:00–24:00: 66.66%
  - May 24th 00:00–24:00: 33.33%
  - Afterwards, we will not accept submissions.

# Classification Task: Group Restructuring

If there are inactive Team Members in your group, or if you are looking for a new team member, please contact tara.jadidi@jku.at until this Friday (2nd of May).

Best practices for teamwork:

- Coordinate early
- Distribute tasks
- Define deadlines
- Check in regularly

# Classification Task: Summary

- Completion of Task 3 is mandatory.
- Answer all aspects and the corresponding questions in your **written report**.
- Use the LaTeX template, stick to the page limit (7 pages, 5 pages text max.) and include a statement.
- Create a **slide deck** which tackles the selected aspects assigned to your group. 4 slides + 1 title slide max.
- Upload both until May 22th to get up to 40 points.