

# Assignment 4 - Project

## Advanced Topics in Neural Networks

November 22, 2024

### Project - 40 points

Choose one of the following projects:

1. Noisy CIFAR-100:  
<https://www.kaggle.com/t/087d976c4de44155b82ee6312a2f5ad1>
2. LLM-based Subject Tagging:  
<https://sites.google.com/view/llms4subjects/>
3. Mu-SHROOM, the Multilingual Shared-task on Hallucinations:  
<https://helsinki-nlp.github.io/shroom/>

Create teams of 1 or 2 students and participate in the chosen competition. Elaborate a research report in which you present the problem and related work, the solution you elaborated and your results.

### Evaluation

You will be evaluated based on the following components. **All 3 components must be present to be graded**

5 points: You have to present your progress at least once in Week 10, Week 11 and Week 12.

30 points: The research report. The evaluation takes into consideration:

1. **Quality of the report**
2. **The Ablation Study**
3. **The Benchmark Performance**

5 points: Oral presentation of 10-15 minutes in Week 12, Week 13 or Week 14. It is not mandatory to create slides for the presentation. During the presentation, you have to answer the following questions:

- Which aspect(s) influenced the best result, and why?
- Mention how could you improve the results, provided you had more time.

### Quality of the report

- **Low quality:** for a simple description of related work, the new solution, and the results.
- **Medium quality:** for detailed related work, well-written solution description, presentation of results, and an analysis of them.
- **High quality:** an academic publication level of writing and structure in all aspects: related work, method description, theoretical analysis, results, and results analysis.

### The Ablation Study

- **No ablation study:** no comparison between a baseline and the proposed solution.
- **Simple ablation study:** for basic comparison between the baseline and new solution.
- **Good ablation study:** for more detailed comparisons on various metrics such as speed, compute, memory usage.
- **Complex ablation study:** for extended comparisons in more scenarios, types of data, data regimes, evaluations of robustness, interpretability, and other more complex studies.

### Benchmark performance

This section is graded based on the leaderboard.