Applied MSc in Data Science & AI

Applied MSc in Data Engineering & AI

Deep Learning with Python

Course Project (2025)

Instructor: Benoit Mialet

Project Overview:

The aim of the course project is to expose students to real-world scenarios in the field of Deep Learning and build their practical expertise in tackling and solving Deep Learning problems. **The primary objective is not to achieve the best performance, but rather to apply a sound scientific approach to a meaningful problem—regardless of the final results.**

The project is a group project. Since this is a mixed course (DS/DE), students are **highly encouraged** to form mixed groups to benefit from the skills of their teammates when working on different components of the project.

Additionally, groups will have the choice to work on 1 of 2 projects:

- Project 1: Solve a real-world Computer Vision task
- Project 2: Solve a real-world Natural Language Processing task

N.B.: choosing a Transformer model is not mandatory. Students may choose another Neural Network since it remains a Deep Learning project.

Note that there is no advantage in picking one project over the other from a grading perspective. The objective is to get students to work on something interesting to them and close to their hearts. All projects are equally rewarding!

Project 1 Summary:

The aim of this project is to familiarize students with the different sub-tasks that exist in the realm of Computer Vision. The objective of this project is to select at least one sub-task in Computer Vision and solve it with a model of the students' choice. A fine-tuning step is required to adapt the model.

Students have complete control over the datasets and the model selection. **However, the methodology should be justified and reproducible.** This is a chance to get students to work in the field of their liking and practically handle complex models.
The sub-tasks can be found here: https://huggingface.co/tasks.

Project 2 Summary:

The aim of this project is to familiarize students with the different sub-tasks that exist in the realm of Natural Language Processing. The objective of this project is to select one sub-task in Natural Language Processing and solve it with a model of the students' choice. A fine-tuning step is required to adapt the model.

Students have complete control over the datasets and the model selection. **However, the methodology should be justified and reproducible.** This is a chance to get students to work in the field of their liking and practically handle complex models.
The sub-tasks can be found here: https://huggingface.co/tasks.

Project guideline

- Each project must include a working solution. Ideally, the solution should be deployed in a Docker environment or on the cloud (e.g., AWS), but this is not mandatory. It may also be delivered as a Jupyter notebook available online (via a shared URL), or a Git repository that can be cloned and executed locally. Students are free to choose their preferred implementation method.
- A demo interface is encouraged, using tools such as Gradio or any other user-friendly

interface. The solution should be lightweight: it must run on a personal computer without requiring a GPU, and should be usable on smaller devices if possible.

- Each project must include a report. Guidelines for this report are given at the end of this document.
- No oral presentation is required

Project Evaluation:

All projects will be evaluated using the following criteria. The final grade will be over 100.

- Project approach, choices and methods (dataset(s), model(s)…)
- Project architecture and code
- Project report
- Project reproducibility (requirements file with necessary packages, README file for running the project, environment, model code)
- Project hosting and deployment (Github, Docker, AWS, Hugging Face Hub or any other method)

Project Timeline:

The deadline for the project is **determined by the Direction of Studies team and will be 2025-11-23.** Additionally, you are free to set a meeting with the instructor to discuss possible approaches, problems or other points pertaining to the project.

Hugging Face organization

If they wish to, students may join the HF DSTI organization, created by Pr. Hanna Abi Akl, and share their datasets (Dataset format) and models for Open-Source community. Uploaded contents may be enriched by students from next cohorts. Don't upload any private, sensitive or personal content here! Joining or not the group will of course have no impact on the grading.

School of Engineering

Report Guidelines

Each group is required to submit a written project report (15–25 pages, excluding appendices). It should include the following sections, but it can be adapted:

1. Introduction (What & Why):

- What is the purpose of your project. Why is it important or interesting?
- Briefly review existing work or methods relevant to your project (academic publications, industrial practices, public datasets, etc.).

2. Methods (How):

- General Approach: Describe your overall strategy to address the problem. Which methodology did you choose (e.g., supervised learning…) and why? Briefly mention external libraries or tools you used.
- Data Description: Present the dataset(s) used. Justify the relevance of the data for your specific objective. Explain how the data was collected, selected, or even constructed from scratch. Mention any preprocessing steps taken (cleaning, etc.).
- Modeling: Describe the models or algorithms you used. Explain your choice. There can be several of them. Detail the training/validation/test setup.

3. Results

- Present your findings using commented tables, graphs, visualizations. Include performance metrics (e.g., accuracy, F1-score, etc.) appropriate to your problem.
- Compare different models or approaches when relevant (and before/after training).

4. Discussion

- Analyze and interpret your results: Summarize the main takeaways from your results. Why do you think you obtained these results? Did anything unexpected happen? What were the main limitations you encountered?
- Propose ideas for improvement: How could you improve the model, data, or approach? What would you do with more time, data, or computational resources?

5. Conclusion

- Recap the objectives and main results. Reflect on the potential impact of your work.

6. Appendix (optional)

- You may include additional figures, tables, code snippets, or detailed mathematical derivations in the appendix. Include a list of references for any external sources used.