# What will we study in this course

Welcome to F21NL - Introduction to Natural Language Processing (NLP) course!

We hope you will enjoy learning from the content found herein, as well as from each other.

Here is a summary of the course:

The *main learning outcomes* are divided into four parts:

1. The theory behind some **foundation problems in NLP**
2. Several popular **Deep Learning** architectures for **Language Modelling**
3. An overview of some **core NLP Applications**
4. **Practical implementation** of models mostly using Deep Learning architectures on downstream tasks

The *topics* we will cover will span 3 different modules:

1. **Theory** Learning Outcomes
    I. **Foundations in NLP**
        1. Problems and Phenomena of Interest in NLP
        2. Machine Learning for NLP Primer
        3. Distributional Semantics and Language Modelling
    II. **Deep Learning Architectures in NLP**
        1. Language Modelling with Neural Networks
        2. Sequence-to-sequence models
        3. Self-attention and Transformers
        4. Pre-trained Language Models
    III. **NLP Applications**
        1. NLU and NLG tasks (e.g., Text Summarisation, Question Answering)
        2. Vision and Language tasks (e.g., Caption Generation, Visual Question Answering)
2. **Practical** Learning Outcomes
    1. Use PyTorch and PyTorch-lightning to train models
    2. Explore popular pre-trained models

The *logistics* of the course are as follows:

Every week consists of 2 *Units*:

- Theory Unit
    - 2-hour **Classroom Session** on **1-2 topics**: *Each* topic typically consists of a few core concepts that we will be exploring in more detail following the slide deck of **Lecture Notes** and a few pages from the relevant chapter from the **handbooks (https://ebsglobal.instructure.com/courses/28967/pages/course-handbooks-ay23-24)** . The Lecture Notes will typically be available a few days before class. Some topics might take more than one session.
- Practical Unit
    - 1 Python **Lab Sheet** mostly using Google Colab
    - 2-hour **Lab Session** that will work on Lab Sheet interactively. Labs will follow the theory topics covered in the classroom and will cover all necessary frameworks required to tackle the coursework. No prior knowledge of the frameworks taught is required of course!

Finally, there will be 2 different types of assessments:

- 2 **Coursework Assignments** (40%)
    - 1 Programming assignment on Word Embeddings, and implementation of downstream tasks and their evaluation (15%; **Handed-out**: 24/09/2024 - Week 3, **Deadline**: 15/10/2024 - Week 6)
    - 1 Programming assignment on Machine Translation, including extensive training and evaluation experiments. (25%; **Handed-out**: 22/10/2024 - Week 7, **Deadline**: 26/11/2024 - Week 12)
- 1 **On-campus invigilated closed-book online Final Exam on Canvas** (60%)
    - Quizzes and practical exercises covering key concepts in the course;
    - Design questions of NLP applications that require critical thinking

If you have any questions about the course at any point, please do get in touch with us (Yannis, Alessandro) either via email, during office hours, or in the discussion boards.

We are looking forward to an exciting Semester full of NLP, Language Modeling, and PyTorch coding!

-- Course Team (Yannis Konstas, Alessandro Suglia)