**MATH 60621A: Natural language processing, Winter 2026**

Mondays 12:00 PM – 3:00 PM, Location TBD

**Instructor** Dr. Eva Portelance (she/they)
eva.portelance@hec.ca
Office hours: TBD

## 1 Overview

This course will cover the fundamental tools and models used for analysing and generating natural language data in a computational setting. We will learn about the core principles behind contemporary natural language processing methods and about the structure of natural language data and how it may be used for downstream tasks such as document search, classification, text generation, and translation.

## 2 Topics covered include:

- The nature of text data
- Text preprocessing
- N-gram language models and classification
- Embeddings and vector semantics
- Recurrent neural language models and LSTMs
- Transformers and large language models
- Encoder models and semantic search
- Encoder-decoder models and language generation tasks
- Finetuning and in-context learning

By the end of this course, students will:

- Understand the nature of text data and how statistical algorithm may be applied to it;
- Learn how both statistical and neural network based language models are structured and how to implement them;
- Learn to think critically about current NLP methodologies and to identify the limitations and reasonable applications of each algorithm covered;
- Develop practical programming skills and concepts that support the above goals.

## 3 Prerequisites

The course expects previous experience with the Python programming language, though no previous experience with machine learning specific libraries is required.

## 4 Course structure

Lectures will be delivered in person, and will not be recorded. Slides for lectures will always be provided before class and students are expected to take their own notes. The first half of each course will be a presentation of new course materials. This presentation will then be followed by a 15 minute break. After break, students will work together in class on each week's practical exercises.

There will be two types of assessments: (1) practical assignments consisting solely of applied coding questions on which students will work either alone or in teams of 2-3 people (each assessed individually); (2) exams covering written theoretical and ethical questions.

Students are encouraged to attend my office hours to ask questions in person, or to discuss course materials and related topics.

## 5 Group work policy

Students can work either alone or in teams of 2-3 people (in fact they are encouraged to) on assignments, if they so desire. If they choose to work with others, they must write down the names of the students with whom they collaborated on their assignment file in the indicated fields. They are still required to individually submit their assignments and if they fail to do so, they will receive a zero grade.

## 6 Late assignment policy

Students will be deducted 15% of their assignment grade per late day, no exceptions. You will be given assignments 3 weeks in advance before their due date — plan accordingly, there is no reason they should be late.

## 7 Email policy

Please allow 2 work days for all email responses. I will never send or respond to emails outside of working hours (8:00-17:00).

## 8 Evaluation

- **Assignments (50%)**: 2 assignments, each worth 25%
- **Midterm exam (25%)**
- **Final exam (25%)**

## 9 Resources

Recommended chapter readings will come from:

- Dan Jurafsky and James H. Martin. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition.* Online manuscript released August 24, 2025. https://web.stanford.edu/ jurafsky/slp3/

## 10 Course schedule

Note that the exact dates below are subject to change, depending on how quickly we make progress through topics in the course.

| Date | Content |
|---|---|
| January 5 | Class introduction and the nature of text data |
| January 12 | Text preprocessing<br>*Recommended readings: Chapter 2, up to and including section 2.7* |
| January 19 | N-gram language models<br>*Recommended readings: Chapter 3* |

| | |
|---|---|
| January 26 | Embeddings and vector semantics<br>*Recommended readings: Chapter 5*<br>*Notes:* Assignment 1 released |
| February 2 | Logistic regression and basic neural networks for classification<br>*Recommended readings: Chapters 4 and 6* |
| February 9 | Recurrent neural language models and LSTMs<br>*Recommended readings: Chapter 13 up to 13.6* |
| February 16 | Transformers<br>*Recommended readings: Chapter 8* |
| February 23 | [Break]<br>**Assignment 1 Due** |
| March X | **Midterm Exam** |
| March 9 | Large language models<br>*Recommended readings: Chapter 7* |
| March 16 | Encoder models and information retrieval<br>*Recommended readings: Chapters 10 and 11*<br>*Notes:* Assignment 2 released |
| March 23 | Encoder-decoder models and generative tasks<br>*Recommended readings: Chapter 13.7-13.9 and Chapter 12* |
| March 30 | Finetuning and in-context learning<br>*Recommended readings: Chapter 9* |
| April 6 | [Easter break] |
| April 13 | Guest research lecture<br>**Assignment 2 Due** |
| April X | **Final Exam** |