

Generative Models on Text

Large Language Models

Organization and Overview

Marion Di Marco

April 14, 2026

General Information

- Joint Seminar for Bachelor and Master students
 - first part of the seminar: lectures
 - second part of the seminar: student presentations
- Lecturers
 - Professor Alexander Fraser
(Professor for Data Analytics & Statistics)
 - Marion Di Marco (PostDoc)
- Course websites:
https://mariondimarco.github.io/teachingsose26_gmt/
https://mariondimarco.github.io/teachingsose26_llm/
Moodle

Topics – Outline

- Organization and Introduction to Linguistic Concepts (today!)
- N-Gram models
- Neural LMs and word2vec
- Contextual Word Representations
- Sequence-to-sequence models
- Transformers
- BERT
- RLHF: Training LMs to follow instructions with human feedback

- Student presentations

Student Presentations

- Students present research papers in class (individually or in groups)
- Research papers from NLP conferences like EMNLP and ACL
 - typically 8-9 pages long
- Prepare a presentation of ca. 15-20 minutes
 - present and explain the main ideas of the paper
 - answering questions and discussion in class
- Written assignment
 - summarize the main ideas of the paper
 - discussion and your reactions to the paper:
 - what questions came up when discussing the paper,
 - what remained unclear, further ideas, ...
- Bonus 0.3 on the final grade for participating in discussion of student presentations!

GenAI Tools – How can you use them?

- You are **NOT** allowed to use it to write your paper!
ChatGPT does not perform well on this task. There are many problems, including “hallucinations”, which lead to involuntary plagiarism
- Instead, use GenAI for
 - Translations
 - Shortening text
 - Summaries
 - Improving grammar
- Never use GenAI to write entire paragraphs!
- If you use it to improve sentences, always check the output and make sure that it does make sense!
- ChatGPT can be wrong in a very convincing way! You, as the author, always carry full responsibility
- Overall: GenAI can be used to get inspiration, but write the paper on your own!
- More details on where to use GenAI and where not to use GenAI:
<https://ceur-ws.org/GenAI/Policy.html>

CIT Information Engineering Programs: Policy on Generative AI

- You **must** acknowledge all use of generative AI (ChatGPT, Gemini, etc)
- You **must** explicitly say what the generative AI was used for (e.g., summaries, improving grammar)
- You **must** add a sentence stating that you alone take full responsibility for all content at the end of this acknowledgement

Standard Disciplinary Sequence for violating these rules

We operate on a two strikes system:

- The first strike will result in a declaration in your school record which all professors can access.
- The second strike will result in exmatriculation

Extraordinary Disciplinary Measure

For particularly egregious misuses of Generative AI, such as falsification of results and/or research data, you will be exmatriculated immediately – these are attacks on the core scientific system and **cannot** be tolerated even for a first strike.

Material and Reading

- *Speech and Language Processing (3rd ed. draft)*,
Dan Jurafsky and James H. Martin (2024)
<https://web.stanford.edu/~jurafsky/slp3/>
- Research papers
- **Preparing for the next lecture:** the reading material for the upcoming lectures will be announced on the course website