

CS 4120 *Natural Language Processing*

Time: MW 2:50-4:30

Location: Churchill Hall 103

Instructor: Terra Blevins

Description

Natural Language Processing is a subfield of Artificial Intelligence that uses methods from Computer Science, Computational Linguistics, Statistics, and Machine Learning to automatically analyze, categorize, understand, and generate natural language with computers. Unlike other kinds of language (say, programming languages), natural language is often ambiguous, nuanced, and subjective. In this course, we will learn about: (i) the linguistic phenomena that make NLP hard for computers to approach; (ii) the main problems that are addressed with NLP; and (iii) the methods used to understand and generate natural language with computers.

This course is hands-on and combines theoretical understanding of NLP methods with practical experience implementing and applying them to real problems. We will examine core tasks, including text classification, sequence labeling, language modeling, and text generation, alongside the linguistic properties that make each task non-trivial. Starting from classical statistical approaches and linear models, we will build toward neural networks (MLPs, RNNs, and Transformers) and modern large language models. Throughout, we will also consider the ethical implications and potential for bias in NLP systems.

Class Staff

Name	Office Hours	Location
Terra Blevins	M 12:30-2:30	Snell Library 009
William Seward	TBA	TBA
Jennifer Justin	TBA	TBA

Please contact the course staff via Piazza or during their office hours. If you have any questions about the course material, you should post a public question – we will redirect you there if you message us directly on general topics that would be beneficial to the whole class.

Prerequisites

Courses: CS 3100/CS 3500/DS 3500

Skills: Students must be familiar with Python to complete the programming assignments. As modern NLP relies heavily on statistical methods and machine learning, *students must be comfortable with basic mathematical concepts from Linear Algebra, Probability, and Calculus*. While we will briefly review relevant concepts, this is not a rigorous treatment of these subjects, and students are encouraged to proactively fill any gaps.

Readings

Readings will be from the following books (all have freely available drafts online):

- [Speech and Language Processing 3rd Edition](#) (SLP), *Dan Jurafsky and James H. Martin* [pdf](#)
(Note that assigned chapters are based on the Jan 2026 version linked here, unless otherwise indicated in the schedule below to be a separate PDF on Canvas).
- [Dive into Deep Learning](#) (DDL), *Aston Zhang, Zachary C. Lipton, Mu Li and Alexander J. Smola* [pdf](#)
- [Natural Language Processing](#) (NLP), *Jacob Eisenstein* [pdf](#)

Note that while most readings are assigned from SLP, [Natural Language Processing](#) (NLP) also covers many topics in this course and can be considered complementary to the SLP readings, if you'd like an additional perspective. Similarly, DDL gives more context on deep learning aspects of the class, if you want to explore those areas further. DDL Chapter 2.3, 2.4, and 2.6 also provide an overview of many mathematical concepts needed for this course.

Coursework and Grading

Homework assignments: These involve implementing various models discussed in class and will be evaluated on standard benchmarks. The assignments are due Wednesdays at 11:59 pm. Each student will receive **3** late days to be used on Homework Assignments. If all late days have been used, Homework Assignments can still be turned in for up to 60% of full credit within three days after the submission deadline. Assignments not submitted within three days of the deadline will receive a score of 0.

Notebooks: In some of the lectures, we will work through exercises in Jupyter notebooks — these must be submitted by midnight on the lecture day. There are no late submissions for notebooks; however, the lowest two notebook grades will be dropped from the final grade. These will be posted on Canvas by the day of the appropriate lecture.

Exams: The midterms will be written in-class and will consist of theoretical and practical questions, including the calculations needed for specific algorithms. Each midterm is non-cumulative, covering different topics from the class.

Final Project: The final project is an opportunity to test how different types of NLP models handle the same task, which is chosen by the group. For this project, you must work in groups of three to four (we suggest that you try to form groups as soon as possible). Further instructions will be posted in Week 2.

Your final grade will be calculated as follows. Assignments in *Homeworks*, *Midterms*, and *Notebooks* are equally weighted unless otherwise stated.

- Homework Assignments (35%)
 - HW1 - Linear Classifiers
 - HW2 - Feedforward NNs
 - HW3 - Transformers
- Notebooks (5%)
- Midterms (35%)
 - Midterm 1
 - Midterm 2
- Final project (25%)

Regrade Considerations: If you are confused or concerned about feedback on course assignments, you may submit any requests for grading reconsideration **within 7 days after the feedback was released**. For homeworks, the regrade requests must be submitted through Gradescope. You may contact the course staff via Piazza to request a regrade for the midterm exam.

Schedule

Weeks	Topic	Date	Lectures	Materials + Readings	Assignments Due
Week 1	Introduction + Course Logistics	09/09 (W)	1. Introduction to NLP		
Week 2	NLP Data	09/14 (M)	3. Text Cleaning and Preprocessing	SLP 2.1-2.2 ; Normalization pdf ; Notebook 1	Notebook 1
		09/16 (W)	3. Data and Annotations	NLP 1, 4.5	
Week 3	Linear Models for Text Processing	09/21 (M)	4. Math Review; Naïve Bayes	Naïve Bayes pdf ; DDL 2.3, 2.4, 2.6 (math review)	
		09/23 (W)	5. Logistic Regression and Perceptrons	SLP 4-4.8 Notebook 2	Notebook 2
Week 4	Neural Networks for Text Classification	09/28 (M)	6. Multilayer Perceptrons and FFNNs	SLP 6-6.4; 6.6-6.6.2 Suggested: SLP 6.6.3-6.7 Notebook 3	Notebook 3
		09/30 (W)	7. Recurrent Neural Networks	SLP 13-13.1; 13.3-13.6 Notebook 4	Notebook 4 Final Project Proposal
Week 5	Word Embeddings and Tokenization	10/05 (M)	8. Word Embeddings	SLP 5 Notebook 5	Notebook 5
		10/07 (W)	9. Embeddings pt 2 + Tokenization	SLP 6.5, 2.3-2.4, 2.8	HW1 - Linear Classifiers
Week 6	NLP Evaluation and Biases	10/12 (M)	No Class (Indigenous Peoples Day)		
Week 6 cont.		10/14 (W)	10. Evaluation, Bias, and Fairness	SLP 4.9-4.10; 4.12 Bolukbasi+ Notebook 6	Notebook 6
Week 7	Structured Prediction Tasks	10/19 (M)	11. Sequence Tagging	SLP 17 Notebook 7	Notebook 7

		10/21 (W)	12. Syntactic Parsing	SLP 18-18.7, 19-19.1	
Week 8	Midterm 1 and Project Check-ins	10/26 (M)	Midterm Exam 1: Weeks 2-6		
		10/28 (W)	No class: Project Check-in Meeting with TA Mentor		
Week 9	Language Models	11/02 (M)	13. N-gram Language Models	SLP 3-3.6 Notebook 8	Notebook 8
		11/04 (W)	14. Neural Language Models	FF LM pdf ; SLP 13.2; 7-7.2	HW 2 - Feedforward NNs
Week 10	Attention	11/09 (M)	16. Attention Mechanisms	SLP 8.1 Suggested: DDL 11 Notebook 9	Notebook 9
		11/11 (W)	No Class (Veterans Day)		
Week 11	Translation and Transformers	11/16 (M)	15. Text Generation and Translation	SLP 7.4; 12	
		11/18 (W)	17. Transformer Architectures	SLP-8.2-8.7 Notebook 10	Notebook 10
Week 12	Midterm 2	11/23 (M)	Midterm Exam 2: Weeks 7-11 (through Attention Mechanisms)		
		11/25 (W)	No Class (Fall Break)		
Week 13	Transformers for NLP	11/30 (M)	18. Pretraining Transformers	SLP 7.5-7.7; 8.8; 9	
		12/02 (W)	19. In-context Learning; Post-training	SLP 7.3; 10	HW3 - Transformers
Week 14	Modern NLP	12/07 (M)	20. Information Retrieval	SLP 11	
Week 14 cont.		12/09 (W)	21. Interpreting NLP Models	SLP 8.9	
Week 15	Final Project Due	12/14 - 12/16	No Class (Final Exam Week)		12/14: Final Project Deliverables!

Course Policies

Academic Integrity: Please read the [Northeastern Academic Integrity Policy](#). All students are required to adhere to this policy during the course. Note that while students are encouraged to discuss course materials, no plagiarism/copying is allowed. **Assignments and code that you turn in should be written entirely on your own.**

You are expected to use the internet as a place for online resources, such as documentation, not as a place to get solutions to your assignments. This includes the use of LLMs or other AI models: while you may use them as a resource, it is not acceptable to have them generate code or other assignment solutions. **If any chatbots, LLMs, AI coding agents, or other sources are used to complete the assignment, you must *fully* disclose where and how this used.** Failure to do so is a violation of the course's Academic Integrity policy.

Academic Integrity violations will result in a 0 on the assignment in question.

Classroom Environment: To create and preserve a classroom atmosphere that optimizes teaching and learning, all participants share the responsibility of creating a civil, non-disruptive forum for the discussion of ideas. Students are expected to conduct themselves at all times in a manner that is respectful towards all participants and does not disrupt teaching or learning. The instructor reserves the right to interrupt conversations that deviate from these expectations. Repeated unprofessional or disrespectful conduct may result in a lower grade or more severe consequences.

Title IX: Northeastern's Title IX Policy prohibits Prohibited Offenses, which are defined as sexual harassment, sexual assault, relationship or domestic violence, and stalking. The Title IX Policy applies to the entire community, including male, female, and transgender students, faculty, and staff.

If you or someone you know has been a survivor of a Prohibited Offense, *confidential* support and guidance can be found through [University Health and Counseling Services](#) staff and the [Center for Spiritual Dialogue and Service](#) clergy members. By law, those employees are not required to report allegations of sex or gender-based discrimination to the University.

Alleged violations can be reported non-confidentially to the Title IX Coordinator within **The Office for Gender Equity and Compliance** at titleix@northeastern.edu and/or through NUPD (Emergency 617.373.3333; Non-Emergency 617.373.2121). Reporting Prohibited Offenses to NUPD does **NOT** commit the victim/affected party to future legal action.

Faculty members are considered "mandatory reporters" at Northeastern University, meaning they are required to report all allegations of sex or gender-based discrimination to the Title IX Coordinator.

Please visit <https://www.northeastern.edu/titleix> for a complete list of reporting options and resources both on- and off-campus.

Students with Disabilities: Students who have disabilities who wish to receive academic services and/or accommodations should visit the [Disability Resource Center](#) at 20 Dodge Hall or call (617) 373-2675. If you have already done so, please provide your letter from the DRC to me early in the semester so that I can arrange those accommodations

