



CSCI 544: Applied Natural Language Processing

Units: 4

Fall 2024 - Tue/Thu - 4:00-5:50 PM

Location: SAL 101

Instructor: Swabha Swayamdipta

Office: SAL 238 (request appointment)

Office Hours: Wednesdays @ 8-9 AM over Zoom (link on Brightspace)

Contact Info: swabhas@usc.edu I will do my best to respond within 48 hours. However, I encourage you to use Brightspace to maximize your chances of response. Please do not ask for D-clearance, I'm unable to provide it.

Teaching Assistant: Brihi Joshi

Office: RTH 314

Office Hours: Tuesdays @ 10-11 AM

Contact Info: brihijos@usc.edu

Teaching Assistant: Ting-Yun Chang

Office: PHE 102

Office Hours: Thursdays @ 10-11 AM

Contact Info: tingyun@usc.edu

Teaching Assistant: Matthew Finlayson

Office: TBA

Office Hours: TBA

Contact Info: TBA

Teaching Assistant: Sayan Ghosh

Office: RTH 420 (for now)

Office Hours: Fridays @ 11-12 PM

Contact Info: ghoshsay@usc.edu

Teaching Assistant: Abel Salinas Jr.

Office: SAL 213

Office Hours: Thursdays @ 3-4 PM

Contact Info: abelsali@usc.edu

Teaching Assistant: Ziyi Liu

Office: PHE101 / <https://usc.zoom.us/my/zliu2803>

Office Hours: Wednesday 3-4pm

Contact Info: zliu2803@usc.edu

Graders:

Shahzaib Saqib Warraich

Neeharika Gupta

Xingjian Dong

Smit Shah

Course Description

This course covers both fundamental and cutting-edge topics in Natural Language Processing (NLP) with a focus on Language Models. Natural language processing (NLP) has been revolutionized by the advancement of large-scale language models achieving state-of-the-art performance across a wide variety of tasks. This course will cover the fundamentals of language modeling and related topics in natural language processing, deep learning and machine learning. Students will gain familiarity with the capabilities of large language models as well as get hands-on experience with building and evaluating small-scale language models. The class will also explore the real-world consequences of deploying language models, such as the ethics and harms associated with them.

Learning Objectives

This course is designed to give students an overview of language models in the context of natural language processing. Students will get hands-on experience on developing and evaluating language models trained on (noisy and) real data via class programming assignments. Moreover, students are expected to come away with skills on classical and current NLP practices, as well as communicating their ideas.

Course Notes

Grading type: Letter

Technological Proficiency and Hardware/Software Required

Experience with programming in Python

Required Readings and Supplementary Materials

All reading material will be posted on the course web page at the beginning of the course. All reading material will be freely and publicly available online. The textbooks we will follow in this class include:

- Jurafsky and Martin. "Speech and Language Processing." 3rd Ed. This textbook contains chapters on the fundamentals of natural language processing.

<https://web.stanford.edu/~jurafsky/slp3/>

- Eisenstein. “Natural Language Processing.” This textbook contains an overview of machine learning approaches for NLP.
<https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf>
- Goldberg. “Neural Network Methods for Natural Language Processing.” This textbook provides a deep learning perspective towards NLP.

Optional Readings and Supplementary Materials

Will be provided on the class website.

Description and Assessment of Assignments

Homework Assignments

There will be four coding homework assignments. The assignments must be done individually. Each assignment is graded on a scale of 0-100 and the specific rubric for each assignment is given in the assignment.

Grading inquiries and questions about the grading of the homeworks and the quizzes can be asked (to the TAs) within two weeks from the grading date.

Course Project

An integral part of this course is the course project, which builds on the topics and techniques covered in the class. Students **must** work in teams of five people on their project, with very few exceptions.

Project Timeline:

- Week 6: Project proposals (1 page)
- Week 10: Project status update due (2-3 pages, expanding on proposal)
- Week 13: Project final report (4-6 pages, expanding on proposal and status reports)

Project description: Each project team will select a topic of their choice. The project types can include NLP prototype design, presenting the design of a novel, original NLP application.

Grading breakdown of the course project:

- Proposal: 10% (4% of total grade)
- Status Reports: 20% (8% of total grade)
- Project Presentation: 25% (10% of total grade)
- Final Write-up: 45% (18% of total grade)

Participation

Students are required to attend classes. Attendance will be taken at random on some days. Non-attendance can be the basis for lowering the grade.

Grading Breakdown

Quizzes: There will be five quizzes based on prior lectures. Missed quizzes will receive a zero grade, and there will be no make-up quizzes.

Homework: There will be four coding homework assignments based on the topics of the class.

Midterm Exam: The midterm exam will contain a mixture of multiple choice and long form questions, covering about the first half of the material covered in the class.

Final Exam: The final exam at the end of the semester covering all of the material covered in the class will contain a mixture of multiple choice and long form questions.

Class Project: Each student will do a group class project based on the topics covered in the class. Students will propose their own project, do the research and build a proof-of-concept, create a video demonstration of the proof-of-concept, and present the project in their report.

Paper Presentation: The project teams will present a scientific publication related to their project to the class. All members of the team are expected to identify the central points of the research, and present that research to the class, as well as answer questions from the instructor and fellow students.

Assessment Tool (assignments)	% of Grade
Quizzes	10%
Homework	20%
Paper Presentations	5%
Class Project	40%
Midterm	10%
Final	15%
TOTAL	100%

Grading Scale

Course final grades will be determined using the following scale:

Letter grade	Corresponding numerical point range
A	93-100
A-	90-92.9
B+	87-89.9
B	84-86.9
B-	81-83.9
C+	78-80.9
C	74-77.9
C-	71-73.9
D+	68-70.9
D	65-67.9
D-	62-64.9
F	Below 62

Assignment Submission Policy

Homework assignments are due at **2:59 PM** on the due date and should be submitted via Brightspace. Late submissions without prior approval, e.g., due to medical conditions, will not be graded.

Grading Timeline

Grades will be available within 2-2.5 weeks.

Late Day Policies

Students are allowed a maximum of 6 late days total for all assignments (but NOT the quiz sheets). You may use up to 3 late days per assignment. Using one late day for a project assignment involves each of the teammates using a late day each. Partial late days are not permitted. For every extra late day beyond the allowed late days, the student / team will lose 20% of the grade for the assignment.

Academic Integrity

The University of Southern California is foremost a learning community committed to fostering successful scholars and researchers dedicated to the pursuit of knowledge and the transmission of ideas. Academic misconduct is in contrast to the university's mission to educate students through a broad array of first-rank academic, professional, and extracurricular programs and includes any act of dishonesty in the submission of academic work (either in draft or final form).

This course will follow the expectations for academic integrity as stated in the [USC Student Handbook](#). All students are expected to submit assignments that are original work and prepared specifically for the course/section in this academic term. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s). Students suspected of engaging in academic misconduct will be reported to the Office of Academic Integrity.

Other violations of academic misconduct include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see the [student handbook](#) or the [Office of Academic Integrity's website](#), and university policies on [Research and Scholarship Misconduct](#).

Policy for the use of AI Generators

Generative AI permitted but limited as follows: In this course, you are permitted to use artificial intelligence (AI)- powered programs to help you, but only on assignments that explicitly indicate a permitted use of AI. However:

- You should also be aware that AI text generation tools may present incorrect information, biased responses, and incomplete analyses; thus, their answers may not meet the standards of this course.
- To adhere to our university values, you must cite any AI-generated material (e.g., text, images, and other content) included or referenced in your work and provide the prompts used to generate the content. Using an AI tool to generate content without proper attribution will be treated as plagiarism and reported to the Office of Academic Integrity. Please review the instructions in each assignment for more details on how and when to use AI Generators for your submissions.

Policy on Plagiarism and Collaboration

Please ask me if you are unsure about what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

Collaboration. In this class, you are expected to submit work that demonstrates your individual mastery of the course concepts.

Group work. Unless specifically designated as a 'group project,' all assignments are expected to be completed individually.

Work by others or Generative AI. Plagiarism includes the submission of code or assignments written by, or otherwise obtained from someone else or generative AI.

If found responsible for an academic violation, students may be assigned university outcomes, such as suspension or expulsion from the university, and grade penalties, such as an "F" grade on the assignment, exam, and/or in the course.

Course Content Distribution and Synchronous Session Recordings Policies

USC has policies that prohibit recording and distribution of any synchronous and asynchronous course content outside of the learning environment.

Recording a university class without the express permission of the instructor and announcement to the class, or unless conducted pursuant to an Office of Student Accessibility Services (OSAS) accommodation. Recording can inhibit free discussion in the future, and thus infringe on the academic freedom of other students as well as the instructor. ([Living our Unifying Values: The USC Student Handbook](#), page 13).

Distribution or use of notes, recordings, exams, or other intellectual property, based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study. This includes but is not limited to providing materials for distribution by services publishing course materials. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relation to the class, whether obtained in class, via email, on the internet, or via any other media. ([Living our Unifying Values: The USC Student Handbook](#), page 13).

Course Evaluations

Course evaluation will occur at the end of the semester university-wide. In addition, we will be conducting a [mid-semester evaluation](#) right before the midterm exam.

Course Schedule

Note: Exam dates are final. Some Quiz and HW dates are subject to change.

	Date	Topics	Deliverables
Week 1	Aug 27	Intro to LMs; N-gram Models	
	Aug 29	N-gram Models	
Week 2	Sep 3	Naive Bayes; Logistic Regression	HW1 Release
	Sep 5	Naive Bayes; Logistic Regression (contd)	Quiz1
Week 3	Sep 10	Word Embeddings	Group Formation Deadline
	Sep 12	Feedforward Neural Nets	Quiz2
Week 4	Sep 17	Backprop	HW1 Due HW2 Release
	Sep 19	Recurrent Neural Nets	
Week 5	Sep 24	Sequence-To-Sequence and Attention	Quiz3
	Sep 26	Transformers - Building Blocks I	Project Proposal Due
Week 6	Oct 1	Transformers - Building Blocks II	HW2 Deadline HW3 Release
	Oct 3	Pre-training and Finetuning Transformers	Quiz4 Mid-semester Eval.
Week 7	Oct 8	[Guest Lecture] PyTorch for Transformers	
	Oct 10	Fall Break	
Week 8	Oct 15	Midterm Exam	
	Oct 17	Generating from LMs	
Week 9	Oct 22	Prompting LLMs	HW3 Deadline HW4 Release
	Oct 24	Preference-Tuning LLMs	Paper Selection Deadline Project Status Report Deadline
Week 10	Oct 29	Ethics and Advanced Topics	Quiz5
	Oct 31	Paper Presentations	
Week 11	Nov 5	Paper Presentations	
	Nov 7	Paper Presentations	
Week 12	Nov 12	Paper Presentations	HW4 Deadline

	Nov 14	[Guest Lecture] Llama-3	
Week 13	Nov 19	Project Presentation	
	Nov 21	Project Presentation	
Week 14	Nov 26	Project Presentation	
	Nov 28	Thanksgiving	
Week 15	Dec 3	Project Presentation	
	Dec 5	Final Examination	
FINAL	Dec 17	Project Final Report Due by 6:30pm	

Statement on Academic Conduct and Support Systems

Academic Integrity:

The University of Southern California is a learning community committed to developing successful scholars and researchers dedicated to the pursuit of knowledge and the dissemination of ideas. Academic misconduct, which includes any act of dishonesty in the production or submission of academic work, compromises the integrity of the person who commits the act and can impugn the perceived integrity of the entire university community. It stands in opposition to the university's mission to research, educate, and contribute productively to our community and the world.

All students are expected to submit assignments that represent their own original work, and that have been prepared specifically for the course or section for which they have been submitted. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s).

Other violations of academic integrity include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), collusion, knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university. All incidences of academic misconduct will be reported to the Office of Academic Integrity and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see [the student handbook](#) or the [Office of Academic Integrity's website](#), and university policies on [Research and Scholarship Misconduct](#).

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

Students and Disability Accommodations:

USC welcomes students with disabilities into all of the University's educational programs. [The Office of Student Accessibility Services \(OSAS\)](#) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems:

[Counseling and Mental Health](#) - (213) 740-9355 – 24/7 on call

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

[988 Suicide and Crisis Lifeline](#) - 988 for both calls and text messages – 24/7 on call

The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline is comprised of a national network of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

[Relationship and Sexual Violence Prevention Services \(RSVP\)](#) - (213) 740-9355(WELL) – 24/7 on call

Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

[Office for Equity, Equal Opportunity, and Title IX \(EEO-TIX\)](#) - (213) 740-5086

Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

[Reporting Incidents of Bias or Harassment](#) - (213) 740-5086 or (213) 821-8298

Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

[The Office of Student Accessibility Services \(OSAS\)](#) - (213) 740-0776

OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

[USC Campus Support and Intervention](#) - (213) 740-0411

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

[Diversity, Equity and Inclusion](#) - (213) 740-2101

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

[USC Emergency](#) - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

[USC Department of Public Safety](#) - UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call

Non-emergency assistance or information.

[Office of the Ombuds](#) - (213) 821-9556 (UPC) / (323-442-0382 (HSC)

A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

[Occupational Therapy Faculty Practice](#) - (323) 442-2850 or otfp@med.usc.edu

Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.