CSCI-499:

Language Models in Natural Language Processing

Units: 4.0
Spring 2024 MonWed 4:00-5:50PM

Location: THH 112
Class Website: http://swabhs.com/2024-spring-csci499-lms-in-nlp/
Class Location: TBD

Instructor: Swabha Swayamdipta
Office: SAL 238
Office Hours: TBD
Contact Info: swabhas@usc.edu. I will reply within 48 hours. Please prepend your email subject with "[CSCI 499]"; e.g. "[CSCI 499] Extra office hours?"

Course Description

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.

Recommended Preparation:

CSCI-270 (Introduction to Algorithms and Theory of Computing (4.0 units) as well as 1 of (CSCI-360 Introduction to AI, CSCI-467 Introduction to Machine Learning, or equivalent experience). Fluency with python programming is highly recommended. Please email the instructor for special circumstances or specific clarifications.

Course Notes

Lecture notes in the form of slides will be made available online after each class.

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.
- **Eisenstein. "Natural Language Processing."** This textbook contains an overview of machine learning approaches for NLP.
- **Goldberg. "Neural Network Methods for Natural Language Processing."** This textbook provides a deep learning perspective towards NLP.

Description and Assessment of Assignments

Grades will be based on homework assignments (40%), a class project (40%), and quizzes (15%) and class participation (5%).

Homework: Students will explore course concepts in detail through homeworks such as coding assignments that require implementing core concepts in PyTorch and training and executing models against real data.

Project: Students must work in teams of two or three (at maximum) to complete a final project on a topic related to the class. This project is expected to include application of ideas in class to any task of their choice. Please come to office hours or email me if you have questions related to choosing a project direction.

Quizzes: In-class quiz sheets will be distributed in class containing 3-5 short questions on topics covered recently. As the quizzes are designed to encourage class participation and attendance, students who do not attend the class will not be allowed to take the quiz later. Quizzes will be timed and will be conducted in some, but not all the classes. Students are expected to turn completed quiz sheets in class.

Class Participation: Students are expected to participate in class discussions by not only asking questions, but also sharing their opinions on the topics under discussion.

Grading Breakdown

Assessment Tool (assignments)	% of Grade
Homework Assignments (4 total)	40
In-Class Quizzes	15
Project Proposal	5
Project Progress Report	10
Project Presentation	15
Project Final Report	10
Class Participation	5
TOTAL	100

Grading Scale

Course final grades will be determined using the following scale

A   95-100
A-   90-94
B+   87-89
B   83-86
B-   80-82
C+   77-79
C   73-76
C-   70-72
D+   67-69

D	63-66
D-	60-62
F	59 and below

Assignment Submission Policy

Assignments should be submitted by email before 11:59pm on the due date.

Grading Timeline

Assignments will be graded within ten days of submission

Additional Policies

Students will have 6 late days they may use on any assignment excluding the Project Final Report. Each late day allows them to submit the assignment 24 hours later than the original deadline. They may use a maximum of 3 late days per assignment. If working in a group for the project, submitting the project proposal or midterm report one day late means that each member of the group spends a late day. We will not allow use of late days for the final project report because we must grade the projects in time to submit final course grades. If a student has used up all their late days and submits an assignment late, they will lose 10% of their grade on that assignment for each day late. We will not accept any assignments more than 3 days late.

Use of Generative AI in this Course

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.

Course Schedule: A Weekly Breakdown

	Topics/Daily Activities	Readings/Preparation	Deliverables
Week 1 Jan 8-12	Introduction to Language Models; N-Gram Models	J&M Chapter 3	HW1 out.
Week 2 Jan 15-19	N-Gram Models Contd. ; Logistic Regression / ML Basics	J&M Chapter 3; 5	
Week 3 Jan 22-26	Word Embeddings	J&M Chapter 6	HW 1 due.
Week 4 Jan 29-Feb 2	Feedforward Neural Networks	J&M Chapter 7	Project proposal due.
Week 5 Feb 5-9	Recurrent Neural Net Language Models	J&M Chapter 9	HW2 out.
Week 6 Feb 12-16	Transformers and Transformer LMs	J&M Chapter 10	
Week 7 Feb 19-23	Masked Language Models	J&M Chapter 11	HW 2 due.
Week 8 Feb 26-Mar 1	Encoder - Decoder Language Models	J&M Chapter 9	
Week 9 Mar 4-8	Flipped Classroom: Project Discussions	-	Project progress report due. HW3 out.
Week 10 Mar 11-15	Spring Break	-	
Week 11 Mar 18-22	Prompting and Instruction Tuning	J&M Chapter 12	HW 3 due
Week 12	LLM Applications: Code Generation / Dialog	Lecture Notes	HW4 out.
Week 13	Generating text from a language model	Lecture Notes	
Week 14	LLMs and Society	Lecture Notes	HW4 due.
Week 15	Project Presentations	-	
FINAL	Exam Week	-	Project final report due.

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 relationship 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).

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, comprises 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.

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 relationship 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).

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.