



CSCI 544: Applied Natural Language Processing

Units: 4

Fall 2026 - Tue/Thu - 4:00-5:50 PM

Location: SAL 101

Instructor: Xiang Ren

Office Hours: Friday 9-10am, GCS SB5 (Under Level 2)

Contact Info: xiangren@usc.edu **This is for personal urgent matters only.** I will do my best to respond within 48 hours. For course related questions, please use Piazza to maximize your chances of response. Please do not ask for D-clearance, I'm unable to provide it.

Teaching Assistant:

Huijuan Wang (huijuanw@usc.edu)

Shicheng Wen (wenshich@usc.edu)

Xu Wang (xwang032@usc.edu)

Hirona Arai (hjarai@usc.edu)

Office: GCS SB5 (Under Level 2)

Office Hours: Friday 11am-12pm

Graders:

TBD

Catalogue Description

Natural language processing; language modeling; deep neural networks for language processing

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

By the end of this course, students will be able to:

- **O1:** Apply key pieces of modern natural language processing pipelines, such as recurrent and Transformer-based sequence-to-sequence models.
- **O2:** Critically analyze contemporary research on large language models, including architectures, training and data paradigms, inference-time methods, and evaluation.
- **O3:** Explain and reason about the core principles, assumptions, and trade-offs underlying modern LLM-based NLP systems across topics.
- **O4:** Identify open research problems at the frontier of LLM research and lead scholarly discussions through presentation and contextualization of recent top-tier research papers.
- **O5:** Design, execute, and clearly communicate an original graduate-level research project on applied NLP or large language models.

Prerequisite(s): There are no formal prerequisites, however you should have a good understanding of the following (though we will review): **Linear algebra** (vector and matrix math basics); **Probabilities:** random variables, discrete and (some) continuous distributions, Bayes' Theorem, Chain Rules;

Programming: python using a Linux environment, mostly without a Jupyter notebook

Co-Requisite(s): N/A

Concurrent Enrollment: N/A

Recommended Preparation: Fluency with Python programming

Course Website

https://inklab.usc.edu/cs544_f26/

Course Notes

Grading type: Letter

Contact Method

On Piazza, or in class/office hours. Please **do not email** (unless urgent personal matters).

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 (20% of total grade)

Through coding homework assignments, students will apply key pieces of modern natural language processing pipelines, such as recurrent and Transformer-based sequence-to-sequence models (Learning Objective **O1**). Homework coding assignments that will involve implementing core concepts using frameworks like PyTorch, then training and executing corresponding machine learning models on real natural language processing data. These coding assignments will be graded based on code correctness in terms of producing expected output, as well as through a written report documenting the code design choices and results of the training and evaluation experiments associated with the assignment.

There will be **two** 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.

Semester Course Project (55% of total grade)

Students will design and carry out a research project that aims to answer a question in natural language processing (Learning Objective **O4**). Students **must** work in **teams of six people** on their project.

Project description: The focus of the class project can be research-focused or application-focused. The project involves identifying a communication or exploration need that language could resolve, data sources available to inform the problem and method, and the techniques needed to approach it. The grading distribution for deliverables of the course project as well as expectations for each deliverable are detailed below.

A **research-focused project** will develop models and analyze data of an existing problem in NLP, or formulate a new problem altogether.

An **application-focused project** will train (possibly only fine-tuning) and deploy NLP models to new application areas, while not necessarily developing any novel research question to be answered. Students will leverage tools, concepts, and techniques presented in the class.

The **deliverables** include:

- Project proposal (1-page single space)
- Project midterm report (4-page single space)
- Project midterm presentation (timed, 2-slide brief update, with questions)

- Project final presentation (timed, with time for questions)
- Project final report (8 pages single space for the main document, up to 15 with appendix/figures).

The *project proposal* (1 page, single-space, 10pt font size) should outline the type of project (research-focused or application-focused), and then answer the following questions clearly in a sentence and/or a few paragraphs each, as appropriate: What are you trying to do? Articulate your objectives using absolutely no jargon. How is it done today, and what are the limits of current practice? What is new in your approach and why do you think it will be successful? Who cares? If you are successful, what difference will it make? What are the risks? What could go wrong, and how will you pivot early on if that happens? How much will it cost? That is, what resources will you need in terms of time and computation? Are these reasonable for a semester and what access you have? Identify two milestones along the way to your finished project.

The *project midterm presentation* (2min presentation with 2 slides + 1min FAQ) will be given by your team in class and graded on clarity, novelty, and presentation quality (by instructor and TA). 1 member from the team can do the presentation representing the whole team. Your presentation should convey:

- **Slide 1:** An overview of the problem and your method (and differentiation with prior work)
- **Slide 2:** the main takeaway you have reached with your project so far; and an overview of what remains to be done before the final report.

The *project midterm report* (4 pages, single-space, 10pt font size) should cover these questions in detail: What are you trying to do? Articulate your objectives using absolutely no jargon. Who cares? If you are successful, what difference will it make? Additionally, it should cover related work in detail, as well as document the proposed method and what is new in that approach, as well as how you plan to conduct experiments, what hypotheses these experiments test, and how you will evaluate the quality of the results. If you have preliminary results to report, they should be discussed here. Additionally, you should revisit your milestones from the proposal and document your progress with respect to these, and propose updated milestones for the remainder of the semester.

The *project final presentation* (6min presentation with 5 slides + 2min FAQ) will be given by your team in class and graded on clarity, completeness, and presentation quality (by instructor and TA). *2-3 members max from the team can do the presentation representing the whole team.* Your presentation should convey: the main takeaway you have reached or aim to reach with your project; a brief motivation which can include some background; the key insight that overcomes the problems presented in the motivation with what was done before this project; the technical details behind the insight; and an overview of what remains to be done before the final report.

The *project final report* (8 pages, single-space, 10pt font size, with up to 15 pages appendix and reference) should answer all the following questions in detail, and is expected to include revised content from the midterm report based on feedback your team received at that time. What were you trying to do? Articulate your objectives using absolutely no jargon. Who cares? If you were successful, what difference would it make? How is it done today, and what are the limits of current practice? What is new in your approach and why did you think it would be successful? How did you conduct experiments? What hypotheses did those experiments test? How did you evaluate the quality of the results? What are the results of your experiments? What

did you discover? Were you able to support your hypotheses? What are the main takeaways of your project? What would you do next if you wanted to keep working in this space? What new questions can you formulate given the work this semester?

The project is total **55% of final grade** with the following breakdown:

Grading breakdown of the course project:

- Proposal: 5% of total grade
- Midterm update presentation: 5% of total grade
- Midterm Report: 15% of total grade
- Project Final Presentation: 10% of total grade
- Final Report: 20% of total grade

Paper review (10% of total grade):

Students will *write a research paper review* to explain concepts underlying natural language processing in their own words (Learning Objective **O2**). The course explores topics through a series of assigned readings in the form of book chapters. Also, the course project would require a literature review. Students will select one reading option and submit a **two-page summary of all the reading**. Reviews will be assessed based on answering a small set of questions, to be released at the time of the paper assignment, clearly and correctly. In most cases, each question will warrant at minimum a paragraph to answer.

Participation (10% of total grade)

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.

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

Grading Breakdown

Assessment Tool (assignments)	% of Grade
Quizzes & Participation	15%
Homework	20%
Paper Review	10%
Class Project	55%
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 day tokens total per student. These tokens can be expended on homeworks, the paper review, and project deliverables (**NOT quizzes or presentations or final project report**) to extend the deadline, one day at a time, for a student without incurring a late penalty.

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.

For project teams, Late Day Token expenditures are **on a per-student basis** (i.e., if a team of 2 turns in their midterm report one day late, a member expending a Late Day Token will receive a 0% late penalty, while a member not expending a Late Day Token will receive a 5% late penalty).

There are **no refunds for late days**: unused late days cannot be converted into credit of any kind.

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.

Week	Date	Class Topics	Readings	Work Due
1	Aug 25	Introduction to LMs and Course Overview		
	Aug 27	n-gram Models	J&M, Chap 3	
2	Sep 1	n-gram Language Models (Smoothing) + Logistic Regression	J&M, Chap 3	Group Formation Google Form Deadline
	Sep 3	Logistic Regression (contd.)	J&M, Chap 5	
3	Sep 8	Word Embeddings	J&M, Chap 6	
	Sep 10	Word Embeddings (contd.)	J&M, Chap 6 Additional: word2vec Explained	Project Proposal Due
4	Sep 15	Feedforward Neural Nets	J&M, Chap 7	HW1 Release
	Sep 17	Backpropagation	J&M, Chap 7	
5	Sep 22	Recurrent Neural Networks	J&M, Chap 8	
	Sep 24	Seq2Seq and Attention	J&M, Chap 8	
6	Sep 29	Transformers - Building Blocks	J&M, Chap 9	
	Oct 1	Transformers (contd.)	J&M, Chap 9	
7	Oct 6	Guest Lecture - PyTorch for Transformers (TA)		
	Oct 8			Fall Recess
8	Oct 13	Pre-training and Finetuning Transformers		HW1 Due
	Oct 15	Tokenization and Generating from LMs	J&M, Chaps 10, 11	HW2 Release
9	Oct 20	Project Midterm Presentation I	J&M, Chaps 2.5, 13	
	Oct 22	Project Midterm Presentation II	J&M, Chaps 13	

10	Oct 27	Large Language Models - Evaluation	J&M, Chaps 10, 12	Project Midterm Report Due
	Oct 29	Large Language Models - In-context Learning, Scaling Law	J&M, Chaps 12	
11	Nov 3	LLMs - Post-Training (SFT, RLHF)	J&M, Chaps 12	
	Nov 5	Guest Lecture on LLMs		
12	Nov 10	Guest Lecture on LLMs		HW2 Due
	Nov 12	Guest Lecture on LLMs		
13	Nov 18	Guest Lecture on LLMs		
	Nov 20			Project Presentations I;
14	Nov 24			Project Presentations II;
	Nov 26			Thanksgiving
15	Dec 1			Project Presentations III
	Dec 3			Project Presentations IV
17	Dec 15			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.