

EECS 487: Introduction to Natural Language Processing

Winter 2025

Meeting Times and Locations

Lectures: Mondays and Wednesdays, 9-10:30am, 1013 Dow Building [Lu Wang]

Labs/Discussions: Fridays, 3:30-5:30pm, 1060 FMCRB [Frederick Zhang, Kaijian Zou, Gordon Lim]

Course Staff

Professor Lu Wang

Email: wangluxy@umich.edu

Website: <https://web.eecs.umich.edu/~wangluxy/>

Office hours: Wednesdays 4:15-5:15pm, BBB 3812

GSI Frederick Zhang

Email: xlzfzhang@umich.edu

Website: <https://web.eecs.umich.edu/~xlzfzhang/>

Office hours: Tuesdays 5-6pm, BBB 1637

GSI Kaijian Zou

Email: zkjzou@umich.edu

Office hours: Thursdays 11-12pm, BBB 1637

IA Gordon Lim

Email: gbtc@umich.edu

Office hours:

- Wednesdays, 3:30–5:00 PM (Virtual) - umich.zoom.us/j/97983202883 (Passcode: 410461)
- Fridays, 1:30–3:30 PM, FMCRB Atrium (Area outside 1060 FMCRB)

Note: Office hours start the week of **Jan 13**.

Discussion Forum (Piazza)

Piazza URL: <https://piazza.com/umich/winter2025/eecs487/home>

Signup URL: <https://piazza.com/umich/winter2025/eecs487>

Course Objectives

The purpose of this course is to provide a broad introduction to the fundamental concepts, tasks, and techniques of natural language processing, and its recent advances based on large language models and applications for interdisciplinary subjects. The successful student will finish the course with specific modeling and analytical skills, knowledge of the most important language concepts and machine learning schemes, and a broad understanding of natural language processing models and practice. The course will serve to prepare the student for further study of NLP and AI in general, as well as to inform any work involving the design of computer programs for substantial application domains.

Prerequisites

This course is designed for junior/senior undergraduate students majoring in computer science, information science, linguistics, and other related areas. Students who take this course are expected to be able to write code in some programming languages (e.g., Python is recommended) **proficiently**, and finish courses in algorithms, probability, and statistics. Linear algebra is optional, but highly recommended. It would be beneficial if the students have prior knowledge of supervised machine learning.

Textbook

Dan Jurafsky and James H. Martin, Speech and Language Processing, 2nd Edition, Prentice Hall, 2009.

Dan Jurafsky and James H. Martin, Speech and Language Processing, 3rd edition draft, web.stanford.edu/~jurafsky/slp3/.

Other Reference:

Jacob Eisenstein, Introduction to Natural Language Processing, The MIT Press, 2019

Interacting with the Instructors

Lectures

Lectures are an important part of this course. While we will generally follow the course text, lecture is an opportunity to emphasize certain topics, to present additional examples, to clarify murkier parts of the text, and to cover complementary technical material.

Lab/Discussion Section

The course includes a weekly lab/discussion section for digging more deeply into aspects of the course material. At some times, the GSI and the IA will work through examples of problems to help students cement the concepts and build strategies for attacking homework and exam problems. At other times, the GSI and the IA will highlight how the foundational course concepts are being extended and applied in contemporary problems and systems. Students will acquire hands-on experience through individual and/or pair programming.

Office Hours

The instructors will have regularly scheduled office hours each week. You are encouraged to make use of these to discuss aspects of the course including lecture material and the homework problems. In cases where you cannot make office hours, contact the course staff to arrange an appointment; don't wait until the last minute though!

Online Resources

Canvas

We will use the Canvas system. If you are registered for the course, you should be able to access all of the important course materials there, including handouts, syllabus updates, etc.

Piazza

We will use Piazza as an online forum for the course. Piazza has various benefits to students who, despite legitimate efforts to understand course material, are confused by course concepts or assignment expectations, and can seek clarification online. If you are asking a question about the assignment that others might have already asked, please first check to see if the question has already been answered in Piazza. It also provides students with greater mastery an opportunity to deepen their understanding by explaining course material to their fellow students. This course wholeheartedly supports such uses of Piazza.

Piazza does, however, have downsides. Some students are tempted to turn to Piazza prematurely, before making a legitimate effort to understand the material from the lecture, discussion, and textbook sources. While doing so might get an answer to a specific question sooner, it robs a student of the opportunity to form a deeper understanding of the material and to develop personal resourcefulness skills that are valuable throughout a career.

Thus, please be thoughtful in your use of Piazza:

- Check to see if a question you have has already been asked before asking it.
- Do not make your (partial) solutions to a problem visible to other students, either in a question (“this is how I’ve solved it so far...”) or in a response (“here is how I did it...”). You can create private postings that only the instructors can see for questions of this type.

- Always be respectful in how you address others on Piazza, just like you should be in all other encounters at the University.
- The instructors will endeavor to check Piazza periodically, but you should not expect responses from them in less than 24 hours.

Assignments

All assignments will be posted and submitted on Gradescope. There will be four homework assignments in this course, due at approximately two to three-week intervals. These may involve some programming or other software work, and also will involve mathematical analysis, critical thinking, and other demonstrations of mastery of the material. Do not leave problem sets (especially any programming tasks) to the last minute, as network servers sometimes go down, and working through the problems usually takes longer than you think (even when you take this fact into account). We give you two to three weeks for a reason!

Late Policy

To be considered on time, assignments must be turned in by 11:59pm on the due date. We will use the submission timestamp to validate turn-in time. Even if you update only a portion of your assignment after the turn-in time, the entire assignment will have a late timestamp and be considered late.

Assignment or report turned in late will be charged 20 points (out of 100 points) off for each late day (i.e. every 24 hours). Each student has a budget of **8 days** throughout the semester before a late penalty is applied. You may want to use it wisely, e.g., save for emergencies. Each 24 hours or part thereof that a submission is late uses up one full late day. Late days are not applicable to final presentation. Each group member is charged with the same number of late days, if any, for their submission (e.g., a proposal or a report). There is no need to inform the instructors if late days are used; timestamp of the last submission on Gradescope will be used for automatic grade calculation.

Collaboration

Adherence to the Engineering Honor Code (<http://www.engin.umich.edu/students/honorcode/code>) is required in all assignments and exams. All problem sets (homework assignments) are to be completed on your own. You are encouraged to discuss ideas and techniques broadly with other class members, but all written work, whether in scrap or final form, are to be generated by you working alone unless otherwise expressly stated in the homework assignment. You are not allowed to sit together and work out the details of homework or exam problems with anyone. You are not allowed to discuss the problem set or exam with previous class members, nor anyone else who has significant knowledge of the details of the questions. Nor should you compare your written solutions, whether in scrap paper form or

your final work product, to those of other students (and vice versa). You are also not allowed to possess, look at, use, or in any way derive advantage from the existence of solutions prepared in prior years, whether these solutions were former students' work products or copies of solutions that had been made available by instructors. Violation of this policy is grounds to initiate an action that would be filed with the Dean's office and would come before the College of Engineering's Honor Council. If you find any ambiguity about this policy, it is your responsibility to contact the course staff for clarification.

Quizzes

For lectures: There will be a quick quiz roughly every week. It usually happens at the beginning of a lecture, covering the content from two previous lectures. The quizzes are designed to cover the key points of each topic. Please see the course schedule for a detailed plan.

For lab sections: There will be a quick quiz or assignment for some of the sections. It usually happens at the end of a section, covering that section's content. The questions should be easy to program and answer as long as the student engages with the section. Please see the course schedule for a detailed plan.

Exam

There will be an open-book midterm exam. You can bring any book, notes, and a laptop, but you are not allowed to use the Internet nor communicate with others via any messaging tool. No cellphone is allowed. That said, though, a sure way to run out of time on an exam is to depend on looking everything up during the exam. One strategy that tends to work well is to prepare a "cheat sheet" on which you can write anything that you expect to find useful. The midterm will be held the day and time indicated in the course schedule. Please ensure that you do not miss the exam.

Course Project

Teams of 3 or 4 students will work together on a course project based on their own selection throughout the semester. Sample project ideas will be provided. The team will work together to provide a 1-page proposal, a 2-page progress report, and finally a project presentation and a 4-page final report over the course of the semester.

Grading

Grades will be determined based on homeworks, quizzes, a course project, an exam, and participation:

- Homeworks (44%): four assignments, each of 11%
- Lecture quizzes (8%): ten quick in-class tests, each of 1%; two with lowest grades will be dropped, and **no make-up**
- Lab quizzes/assignments (4%): ten quick in-class tests, each of 0.5%; two with lowest grade will be dropped, and **no make-up**
- Project (16%): team of 2 to 3 students, proposal (2%), reports (4%+6%, mid-term progress and final), presentations (4%, with 1% as bonus if selected as best project by peer students)
- Midterm exam (25%): open-book
- Participation (3%): participating in-class discussions, interaction on Piazza (answering questions, sharing notes, etc)

Schedule for Lectures, Lab/Discussion and Deliverables

The schedule is given on the Canvas website. Be forewarned that adjustments/revisions to the schedule might be made over the course of the term.