

# CS187: Introduction to Computational Linguistics and Natural-Language Processing

- [Administrative information](#)
  - [Staff](#)
  - [Prerequisites](#)
  - [Website](#)
  - [Enrollment limitation](#)
  - [Texts](#)
- [Coursework](#)
  - [Course content](#)
  - [Workload](#)
  - [Pedagogy](#)
  - [Class meeting schedule](#)
  - [Labs](#)
  - [Lectures](#)
  - [Readings](#)
  - [Problem sets](#)
  - [Exams](#)
  - [Office hours](#)
  - [Discussion forum](#)
  - [Grading](#)
  - [Submitting coursework](#)
- [Course policies](#)
  - [Academic integrity and collaboration policy](#)
  - [Auditing policy](#)
  - [Simultaneous enrollment policy](#)
  - [Late enrollment policy](#)
  - [Pass/fail policy](#)
  - [Course climate](#)
  - [Mental health](#)
  - [Accommodations for students with special requirements](#)
  - [Accommodations for students with severe allergies](#)

**This is a provisional version of the syllabus and calendar. Expect changes over time.**

**Interested in enrolling? Please fill out [the precourse survey](#).**

## Administrative information

### Course

CS 1870 (formerly CS187; Harvard College/GSAS course number 117372)

### Term

Fall 2024-25

### Meeting times and locations

Mondays, Wednesdays, and occasional Fridays, 11:15am-12:30pm ET, [SEC, 150 Western Avenue, room LL2.223](#).

### Instructor

Stuart M. Shieber ([full contact information](#))

### Course description

Natural-language-processing applications are ubiquitous: [Alexa](#) can set a reminder, or play a particular song, or provide your local weather if you ask; [Google Translate](#) can make documents

readable across languages; [ChatGPT](#) can be prompted to generate convincingly fluent text, which is often even correct. How do such systems work? This course provides an introduction to the field of computational linguistics, the study of human language using the tools and techniques of computer science, with applications to a variety of natural-language-processing problems such as these. You will work with ideas from linguistics, statistical modeling, machine learning, and neural networks, with emphasis on their application, limitations, and implications. The course is lab- and project-based, primarily in small teams, and culminates in the building and testing of a question-answering system.

## Staff

- [Stuart Shieber](#), instructor
- [Yonatan Belinkov](#), adviser
- TBD

## Prerequisites

Programming ability and computer science knowledge at the level of CS51; knowledge of discrete mathematics, including basic probability, statistics, and logic at the level of CS20; calculus at the level of Math 21a/b; some familiarity with Python programming and with standard software tools (programmers' editors, Unix command line, git, etc.).

## Website

<http://cs187.info>

## Enrollment limitation

We will limit enrollment depending on availability of teaching staff. Enrollment preference will be given to junior and senior CS concentrators. Please fill out [the precourse survey](#) if you are interested in enrolling in the course.

## Texts

In addition to home-grown materials, the following two textbooks will be extensively used. Both are available in open versions online.

Primary textbook

[Speech and Language Processing](#) (3rd ed. draft), Dan Jurafsky and James H. Martin, forthcoming. (Earlier editions, while available in print, are substantially different from this one.)

Supplemental textbook

[Introduction to Natural Language Processing](#), Jacob Eisenstein, [2019](#).

## Coursework

### Course content

The course will cover a range of methods, including structural/formal, statistical, and neural methods, for a range of natural-language interpretation tasks – where the interpretations are structured as classes, sequences, tree structures, or meaning representations.

Emphasis will be on gaining a fuller hands-on understanding of fewer methods rather than a cursory understanding of more.

We will also address the timely issue of ethical concerns with natural-language processing applications through a module in cooperation with the [Embedded EthiCS](#) program.

### Workload

[Last year's Q guide](#) reported a median of 6 hours per week outside of class time.

### Pedagogy

The coursework for CS187 is focused on in-class collaborative labs and a sequence of problem sets.

The course is divided into four segments of about three weeks (about six class sessions) each, covering the four major kinds of outputs of NLP systems. Each segment is introduced by a lecture/discussion constituting the first class session, followed by a series of labs. For each lab,

1. Students prepare through readings in the textbooks and other course materials, and fill out a brief reading survey.
2. During the class meeting time, students work in pairs on a sequence of exercises involving both programming and pencil-and-paper work, to gain understanding of and fluency with the ideas in the readings. Staff engage with students to help out and provide guidance. Students can submit the lab to a grading server that will provide some immediate feedback.
3. After the class meeting time, pairs continue to meet together and finalize their solution to the lab. Each student in the pair submits the solution to the grading server, which will be automatically and manually graded.

Each segment ends in a problem set that is completed individually. Taken together, the problem sets culminate in the building, evaluation, and comparison of two versions of a natural-language question-answering system.

All programming is done in [Python 3](#) using [Jupyter notebooks](#) locally or on [Google Colab](#).

## Class meeting schedule

The spreadsheet below provides a *provisional* calendar showing all class sessions, both lectures and labs with their required pre-session readings, as well as problem set due dates. Note that some class sessions fall on Fridays, especially when no class is held on Monday due to start of term or holidays. *This calendar is provisional and subject to change.*

## Labs

Labs start the second week of classes. The lab sessions involve a short introduction, followed by [pair-programming](#) exercises to be completed in lab and thereafter. Attendance at and participation in labs is *strictly required*, and attendance is taken. See [Absences](#) for further information.

For each lab, you will be assigned a partner to work with. You work on the exercises in the lab *together*. Staff are available to provide help and make sure you are making progress and understanding the concepts. Each pair submits their lab solution by 5 pm EST two days after the lab session (for instance, 5 pm Wednesday for a Monday lab).

## Lectures

A small set of lecture/discussions provide background on the course. *Students are expected to attend and participate in all of these lecture/discussions.*

## Readings

Chapters from the course textbook are assigned before most lectures and labs. In addition, we provide a PDF version of the lab itself (distributed at <https://go.cs187.info/labinfo>) ahead of time for you to look over so that you can spend your time efficiently in the lab session. You should complete these readings and a short survey on the readings by the day before the corresponding lab or lecture. *Every lab session will assume that students have looked over the lab and done the other reading for the session, and your lab partners will be relying on this as well.*

## Problem sets

A series of four problem sets provides the opportunity to use the techniques practiced in labs on a larger scale. Each problem set involves the implementation of a natural-language-processing application built around the task of answering questions in the “air travel information system” ([ATIS](#)) domain.

## Exams

There are no exams in the course.

## Office hours

Course staff will hold regular office hours on a schedule to be determined, typically including the Friday course time when not being used for regular course meetings.

## Discussion forum

Students are encouraged to ask and help answer questions [on the course discussion forum](#), where staff will also be available to help.

For individual administrative questions, please email the course's head TFs at [staff@cs187.info](mailto:staff@cs187.info).

## Grading

Your grade in the course will be based on your performance on the various course components – reading surveys, labs, problem sets – as well as your contribution to the learning of others as evidenced by your collaborations with your partners in labs, including post-lab surveys, your contributions to questions and answers [on Piazza](#), among other factors.

Approximate weightings for the course components are:

<b>Component</b>	<b>Percentage</b>
Problem sets	50%
Individual lab submissions	40%
Other aspects	10%
<i>Total</i>	<i>100%</i>

The other aspects include your contribution to the learning of others and your work on the ethics assignment.

## Absences

Because of the collaborative nature of so much of the instruction in CS187, especially in labs, we require active engagement in those course components. If you plan to participate in academic or extra-academic activities that would regularly preclude you from attending lab, enrollment in CS187 will not be possible.

This requirement is reflected in the grading policy as well. We will reduce the final grade based on unexcused absences from labs as per the following schedule:

<b>Absences from</b>	<b>to</b>	<b>Half-grades deducted</b>	<b>Reduction from A</b>
0	1	0	A
2	3	1	A–
4	5	2	B+
6	7	3	B
8	9	4	B–
10	11	5	C+
12	13	6	C
14	15	7	C–
16	17	8	D+
18	19	9	D
20	21	10	D–
22	23	10	D–
24	and above	11	E

Of course, absences precipitated by any health issues will be excused. Absences based on *required, pre-planned* participation in other academic and extracurricular activities will also be excused upon a timely request several days in advance. Requests for any of these issues should be submitted at <https://go.cs187.info/absence>, with corroborating information provided as soon as possible thereafter.

## Grading of labs

Participation in labs affects your grade in several ways.

- Most importantly, it is through the preparation for and engagement in labs that the concepts and

skills are taught in the course.

- Second, your participation in all aspects of the lab “ preparation, attendance, collaboration with members of your group ” will be reflected in self- and peer evaluations that are used in assessing your contributions to the learning of others.
- Lab problems involving programming are graded for *correctness* automatically based on a set of unit tests upon submission to the grading server. We provide a small set of *open* unit tests visible in the distributed lab notebooks, which typically implement “sanity checks” on your solutions, for instance, does the answer deliver values of the right sort. For grading purposes, we make use of a larger set of *closed* unit tests, not visible in the distributed notebooks, which provide much more detailed checking. Thus, *perfect performance on the open tests is no guarantee of a high correctness score*, so it is recommended that you fully test your code yourself.
- Lab problems requesting an open-ended response are graded for *understanding* according to the following coarse rubric:

<b>Score</b>	<b>Meaning</b>
X	Unsatisfactory: Nothing was turned in beyond the distribution code, or something was turned in that indicates major conceptual problems. (0)
“	Satisfactory: Needs improvement, showing some holes in understanding or substantive problems. (3)
”	Excellent: A good job, showing full understanding of the concepts, with no major problems and few minor ones. (4)
+	Exemplary; the instructors would have been proud to turn this in. (5)

+s are likely to be very rare; graders typically award them to only a few percent of submissions. A student doing very well in the course should be getting mostly ”s, with some -s and perhaps an occasional +. Grades in the ” range are thus quite respectable. Students should not expect to get many +s, though striving for them is a good idea.

These grades will be used in calculating your overall grade. For the purpose of grading, the quality score is converted to a number as per the parenthesized numbers in the table above. The quality scores are normalized separately and averaged to form the problem set quality component of the final grade.

## Grading of problem sets

Problem sets are graded manually using the same rubric as above.

## Submitting coursework

All student solutions to programming exercises, including problem sets and labs, should be submitted to the course grading system. Submissions by email or any other method are not accepted.

### Submitting labs

Lab submissions are due by 5 pm ET two days after the lab (5 pm Wednesday for a Monday lab for instance).

The grading system automatically checks every lab submission to make sure that it compiles cleanly against a series of unit tests. Submissions that do not compile are rejected. *Any attempt at submission that fails to compile cleanly receives a 0. Rejection is equivalent in both process and grading to not having submitted the problem set at all.*

Lab submissions that are accepted (having compiled cleanly against the unit tests) will be graded automatically against the unit tests as well as manually for the open-ended questions.

### Submitting problem sets

Problem set submissions are due as per the calendar below. All problem sets are due on the indicated due date by 5 pm ET unless otherwise indicated.

Occasionally, extraordinary circumstances may make it impossible for you to submit your problem set solutions on time. For this reason, we allow for six “late days” for your problem sets. For each day a problem set is turned in late, *up to two per problem set and allocated “greedily”*, you will be charged one of your allotted late days. For example, if a problem set is due on Wednesday at 5 pm it can be turned

in by 5 pm on Thursday charging one late day, or by 5 pm Friday charging two, but will receive no credit thereafter. Late days are measured only in full days (not in hours). *After two days, we will not accept late homework* and will stop charging late days; that is, only two late days will be charged for unsubmitted work or work submitted exceptionally late. If you turn in a problem set late without sufficient late days remaining, the problem set will not earn any credit. However, we recommend that you turn in such problem sets anyway as we may consider them in allocating final grades in borderline cases.

Late days may not be used for labs, so plan accordingly.

## Course policies

### Academic integrity and collaboration policy

The course, like all courses at Harvard College, operates under the salutary spirit of the [Harvard College Honor Code](#). That spirit is especially important in considering collaboration on course work. Students are encouraged to discuss all aspects of the course work “ readings, labs, problem sets ” with each other; talking together can be a useful method for working out difficulties in solving the problems and in improving your understanding of the concepts. Indeed, we will provide opportunities for this kind of interaction in the weekly sections and office hours.

However, except where explicitly stated otherwise, *all assignments should be completed individually*. By this we mean:

- It is allowed and encouraged to have high-level discussions with others, but working together one-on-one directly on developing your solutions goes beyond such high-level discussions, as does sharing code (no matter how small a snippet). Such modes of “collaboration” are expressly disallowed.
- Making use of others’™ code that you come across is not appropriate, even if cited. Other people’s™ solutions to these problems may exist on the web, or in the files of past students, or on roommates’™ laptops, or in the recycling bin near your house printer. It is not advisable to be reconnoitering for them for help. What counts as “other people’s™ solutions” includes answers written by previous or current students in the course; answers posted on programming forums such as StackOverflow, whether in response to your query or others’™; and answers generated by NLP systems such as ChatGPT or Copilot. It ought to go without saying that all individually submitted code should be the student’s™ own. But we said it anyway.
- For labs completed in pairs, members of the pair will of course work closely together to develop a single coded solution. But the same rules apply to members of the pair looking outside the pair for help. High-level discussion is fine; sharing code is not.
- It is inappropriate, and considered a violation of the collaboration policy, to publicly post or privately distribute your coursework, including your solutions to labs and problem sets.

Please see the section on [Plagiarism and Collaboration in the Handbook for Students](#) for further clarification.

### Auditing policy

Because of the highly interactive nature of the class sessions, auditing the course is not allowed.

### Simultaneous enrollment policy

Students in CS187 are required to attend all lab slots. Ordinarily, you may not enroll in courses that meet at the same time or overlapping times, as described in the [Harvard College Handbook for Students](#).

Because attendance at lab is strictly required, if you have conflicts with the course time, you will not be able to enroll in CS187. See [the College policy on simultaneous enrollment](#) for further information.

### Late enrollment policy

Because the course ramps up quickly through collaborative activities, it is difficult to join the course late. We therefore strongly recommend against joining the course after the first week, and will under no circumstances approve joining the course after the second week. Keep in mind also that labs begin the second week of the course, so that joining the course late would lead to missed labs, with the concomitant



effect on the final grade as described in [the section on absences](#).

## Pass/fail policy

CS187 can be taken pass/fail with the following proviso: Because the course relies so much on collaborative effort in labs and on problem sets, all students, including those enrolled pass/fail, are under a moral obligation to the others in the class to participate fully in labs and all its associated processes, including preparation for lab (especially doing the reading before lab), lab attendance, engagement during lab, submission of labs, and the post-lab surveys. Thus, taking the course pass/fail is unlikely to provide much time relief in the course, though it may reduce grade anxiety. Students interested in taking the course pass/fail should email the course staff ([staff@cs187.info](mailto:staff@cs187.info)) well before the pass/fail deadline (fifth Monday of term).

## Course climate

It is my intention to have a course climate that is open to and inclusive of everyone “ of all races, genders, sexual orientations, ethnicities. The primary preventative for inappropriate interactions is empathy. I urge all participants in the course “ staff and students “ to think empathetically as we all interact with each other. Because I strive to be a moral person, I attempt to be empathetic myself. But I know that even the best of intentions can inadvertently fail. If you find any of my own behaviors to infringe on the open climate I strive for, please let me know, either directly or through a friend, so I can benefit from the event and improve my own behavior. If you would prefer to go through an independent, neutral, and confidential service, the [Harvard University Ombudsman Office](#) can serve as intermediary. For concerns of sexual harassment or other sexual misconduct, the [Harvard Office for Gender Equity](#) can provide support and help with your various options.

## Mental health

If you experience significant stress or worry, changes in mood, or problems eating or sleeping this semester, whether because of CS187 or other courses or factors, please do not hesitate to reach out immediately, at any hour, to any of the course staff to discuss. Everyone can benefit from support during challenging times. Not only are we happy to listen and make accommodations with deadlines as needed, we can also refer you to additional support structures on campus, including, but not limited to, the below.

- [Counseling and Mental Health Services](#), 617-495-2042
- [Let’s Talk](#)
- [Room 13](#), 617-495-4969

## Accommodations for students with special requirements

We will happily accommodate any student with special requirements. Students needing academic adjustments or accommodations because of a documented disability should contact the [Disability Access Office](#) (AEO) to obtain a Faculty Letter, and then contact the course administrator ([staff@cs187.info](mailto:staff@cs187.info)) by the end of the second week of the term. Failure to do so may result in our inability to respond in a timely manner. All discussions will remain confidential.

## Accommodations for students with severe allergies

Because labs are held at typical meal times, students occasionally bring food with them. Unfortunately, labs are held in a large open room, so that students may be present who have severe food allergies to the food brought in. In order to prevent problems, if you have a severe food allergy that precludes your being present in the lab room with the allergen, please let us know ([staff@cs187.info](mailto:staff@cs187.info)) so that we can inform students in the class to avoid bringing the allergen into lab sessions.