CMPU 366

Natural Language Processing

Fall 2025

Monday 1:30–2:45 p.m.
Wednesday 1:30–2:45 p.m.

New England 105

Professor Gordon

cs.vassar.edu/~cs366

Overview

This course examines the intersection of human language and computation, focusing on how computers understand, analyze, and generate natural language. Students explore both traditional techniques (e.g., n-grams and rule-based systems) and modern methods (e.g., neural networks and deep learning) for language modeling and manipulation. Through hands-on projects, they apply these approaches to tasks such as information extraction, text classification, machine translation, and language generation. The course also addresses the ethical implications and societal impacts of language technologies, encouraging critical reflection on their development and use.

Course goals

1 *Understand the foundations of NLP*
Grasp core methods in NLP grounded in statistics, algorithms, and linguistics, along with key facts about human language that motivate computational approaches.

2 *Gain practical skills with NLP tools*
Develop fluency with Python and widely used NLP libraries, implementing and experimenting with models across tasks such as classification, generation, and language understanding.

3 *Critically evaluate NLP systems*
Learn how to measure model performance, recognize strengths and weaknesses of different approaches, and thoughtfully assess trade-offs in system design.

4 *Engage with research and open-ended inquiry*
Read and interpret current NLP research papers, and experience the process of an open-ended project that involves building, testing, or analyzing a language technology.

5 *Explore applications and societal impact*
Investigate diverse NLP applications, including those with positive societal impact, while learning to identify limitations and pitfalls of data collection and annotation practices.

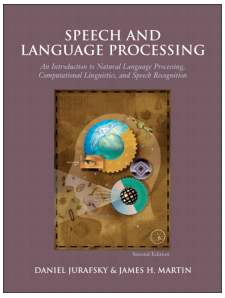
6 *Reason about ethics, bias, and equity in NLP*
Reflect on how NLP systems can perpetuate stereotypes, biases, and inequities, and develop frameworks for anticipating and mitigating these harms.

Prerequisites

- Two 200-level Computer Science courses.
- CMPU 250 or MATH 244 or permission of the instructor.

Readings

Our primary textbook is available for free online:



> *Speech and Language Processing*, third edition (draft)
> Dan Jurafsky and James Martin

Additional readings will be posted on the course website.

Course conduct

Natural language processing as a field is prone to touch on topics related to culture and identity. The unique experiences you bring to the class from your own experiences not only strengthen our community but also actively contribute to the learning of everyone in the

classroom. On the flip side, false cultural assumptions and negative comments about others both causes harm and makes our collective work as conscientious NLP scholars harder.

As your instructor, I am committed to creating a classroom environment that welcomes all students, regardless of race, gender, social class, religious beliefs, etc. We all have implicit biases, and I will try to continually examine my judgments, words and actions to keep my biases in check and treat everyone fairly. I expect that you will do the same with respect to me and the other members of the class, and that you will let me know if there is anything I can do to make sure everyone is encouraged to succeed in this class.

Coursework and grades

The elements of the course will be weighted approximately as follows:

Assignments	40%
Midterm exam	10%
Special topics presentation	10%
Final project	40%

Remember that success in the course is more than just good grades. It means that you are being challenged to grow as a learner, that you are engaging actively with tasks that feed your growth, and that you are creating excellent work by completing challenging tasks with an appropriate level of support. It also means that you are building your lifelong learning skills so that once the course is over, you are better and stronger as a learner and can continue to learn new things independently.

Late work

For flexibility, you have four "late days" that you can use to extend the deadline of any homework assignment (not final project) by 24 hours. You can use them all for one assignment or spread them out.

If you run out of late days, late assignments will incur a penalty of 10% for each day past the deadline. No late days can be used for the final project.

If you know you won't be able to submit an assignment, talk to me. If you're experiencing a major problem – medical, psychological, family, etc. – that is interfering with your ability to complete your class work, you should talk with the Dean of Studies, your class advisor, or Health Services, who will recommend appropriate accommodations to all of your professors, who genuinely want you to succeed.

Getting help

I expect all students to participate in class discussions. In return, I will make myself available to answer questions, listen to concerns, and talk to any student about topics related to the class. I welcome your feedback throughout the semester about how the course is going.

Our course uses Ed for announcements and online discussions. You can use it to ask questions about course concepts, assignments, and logistics. If you go 30 minutes working on an assignment without making progress, please ask for help! You can post questions on Ed, come to my office hours, or schedule a meeting.

Posts on Ed can be either public or private. Public posts are visible to everyone, while private posts are visible only to course staff. Any questions having to do with your particular solution to an assignment should be private; all other posts should be public. (If you have a question about something, it's very likely that other students do, too!)

Because Ed is an extension of our classroom discussion, we expect everyone to behave accordingly: No disrespect, rudeness, or abuse will be tolerated.

Ed is a discussion forum, so please feel free to respond to questions and comments – it's great when students can learn from each other!

Instructor availability

Office hours will be listed on the department website.

I typically only check email and other messages between 9 a.m. and 5 p.m. on weekdays. If you send a message that needs a response during those times, you can expect to get a response the same day. Otherwise, you can expect one when I'm back online.

Collaboration policy

The goal of the course is to gain an understanding of computer science and data science, not to compete with each other. Indeed, in CS as a field, collaboration is the norm. As such, you're encouraged to discuss class material with other CMPU 366 students when reading, studying, and thinking about the problems.

Assignments: Limited collaboration

For homework assignments, you are permitted to discuss high-level ideas with other students, but you must produce your code and other responses on your own.

In particular, the following activities are *not* allowed when working on homework assignments:

- Sharing code files with another student for any reason,
- Sitting next to another student while writing up solutions, looking at what the other is typing,
- Sending code for a homework question to a classmate "just so they can look at it to figure out how to do the problem",
- Obtaining a solution online or from someone not in the course,
- Leaving your work in unprotected directories or services (including GitHub, GitLab, or CourseHero) where other students can find them, or
- Showing a classmate your code so they can help you find or debug an error.

However, the following scenarios are fine:

- Asking the professor or coaches for help,
- Classmates discussing an assignment at a general level (rather than the code) – discussing what the question is asking, what topics it draws on, and other similar non-code issues,
- Asking a classmate what general causes of a particular error message might be, or for debugging strategies, without showing them your code.

Projects and special topics: Permitted collaboration

For the final course project and the special topic presentations, you are permitted – and strongly encouraged – to work in a group with other students. You are expected to contribute fairly to the group's work, and you will be asked to describe your contribution at the end.

You are allowed to discuss your work with other groups, but – as on assignments – this should be at a high level and not involve sharing code or data.

Exams: No collaboration

Exams must be completed individually. Do not discuss the contents with anyone but the professor until the exams are graded or example solutions are released.

Academic integrity

Please read the CS department's guide to academic integrity. In particular, note that: cs.vassar.edu/integrity

1. You may not copy code written by anyone else, e.g., a classmate, a friend, an online source, a book. (You *are* permitted to use the code provided by the instructor or included in the assigned readings!)
2. **This includes any generative AI tools, such as ChatGPT, Copilot, Claude, or Gemini.** The output of these systems is not considered to be your own work, and therefore assignments that use generative AI are violations of academic integrity unless an assignment specifically allows their use.
3. Using code or other material from sources as "inspiration" and submitting highly derivative solutions is considered copying. (That is, you can't "paraphrase" someone else's work!)
4. You may not post a public question that contains any part of your code.
5. You may consult online resources as part of your coursework, but you may not copy code from an online source. If you get an idea of how to solve a problem from an online source, include a citation near the top of your source code, e.g.,

```
# Consulted online source for list comprehensions:
# https://diveintopython3.net/comprehensions.html
```

You do not need to include a similar notation if you consulted with a classmate; we expect that – just don't share code!

If you haven't already done so, you should read "Going to the Source", available from the Dean of the College website. Note that the guidelines that apply to writing in general apply equally to the writing of computer programs. In particular, *copying anyone else's code without attribution is plagiarism*.

Quoting from Chapter XI, "In suspected cases of plagiarism, the instructor prepares a written statement of complaint to the Academic Panel." Please don't put yourself or your professor in that position. When in doubt, *stop* and ask me first.

Academic accommodations

Academic accommodations are available for students who are registered with the Office for Accessibility and Educational Opportunity (AEO). Students in need of disability (ADA / 504) accommodations should schedule an appointment with me early in the semester to discuss any accommodations for this course that have been approved by the Office for Accessibility and Educational Opportunity, as indicated in your AEO accommodation letter.

Religious observances

Students who wish to request academic accommodations for a religious observance should email me no later than the end of the add period. While requests can be made at any time, accommodations may be more limited if submitted after this period. If you have questions or concerns about your request, please contact the Office of Religious and Spiritual Life and Contemplative Practices or the Office of the Dean of Studies. Please note that accommodations do not exempt you from completing any coursework missed as a result of a religious observance.

Title IX

If you (or someone you know) have experienced sexual harassment or sexual violence, including sexual assault, relationship violence, or stalking, know that you are not alone. Vassar has staff trained to support you in navigating campus life, accessing health and counseling services, providing academic and housing accommodations, and more. Vassar strongly encourages all community members to take action, seek support, and report incidents to the EOAA / Title IX Office at 845-437-7924 or titleix@vassar.edu.

For more information about confidential support on- and off-campus, see offices.vassar.edu/savp.

Please be aware that faculty members must disclose information about suspected or alleged sexual harassment or other potential sexual misconduct to the EOAA / Title IX Office. When the EOAA / Title IX Office receives information, they will reach out to offer information about resources, rights, and procedural options. Response to this outreach and further participation with the EOAA / Title IX Office is voluntary. Meeting with EOAA / Title IX does not mean you must file a formal complaint or report to another entity, including law enforcement. This meeting can be for information and support only or to file a complaint and receive support.

Acknowledgments

This course – and this syllabus – is based in large part on work by other professors, especially Carolyn Anderson at Wellesley College. The policy on asking and answering questions and on communication on Ed is adapted from Laney Strange, Northeastern University. The statement on course conduct is adapted from Xanda Schofield, Harvey Mudd College.