Natural Language Processing

COSC 426 Section A

Instructor Info

Forrest Davis (he/him)

T: 3:30-5:00PM, W: 9:00-10:00AM, R: 12:30-1:30PM

322 Bernstein

fdavis@colgate.edu

Course Info

T,R

9:55-11:10AM

214 Bernstein

Overview

Language is an amazing and intricate human capacity and building computational systems that process human language (Natural Language Processing; NLP) has widespread applications: both practical (e.g., searching the internet, virtual assistants, chatbots, autocorrect, translation systems, audio captioning, and hate speech detectors) and scientific (e.g., as models of the human mind). Coursework has three interrelated goals. First, to develop the language and tools necessary for understanding, building, and evaluating NLP systems. Second, to engage in the scientific process by reading papers, replicating existing results, developing research questions and running experiments to answer these questions. Third, to critically examine the broader impact of NLP systems on society.

Material

Required Text: There is no required textbook. Readings will be drawn from a variety of sources, including contemporary research. A wonderful resource (that we will also draw on) is Jurafsky and Martin's Speech and Language Processing.

Moodle Site: The course Moodle can be found here. We will also use Gradescope, which can be found here.

Course Site: The course site can be found here. It includes links to course materials, slides, and homeworks.

Schedule of Topics: A dynamic schedule of topics for this course can be found here.

Coursework

Class

Each class meeting will feature a mix of lectures, demonstrations, open question and answer, and active problem solving (on your own and in small groups). You are expected to attend class, respond to questions, solve problems, and respect and support your classmates. If there is someone you do not want to work with, please complete this form anytime during the semester (no reason needs to be provided). The course is supplemented by readings from a variety of sources. Before each class meeting you should engage with the listed materials on the course schedule

Homework Assignments

Homework assignments apply the concepts discussed/practice in class (and lab) to more challenging problems. There will be around 3 larger homework assignments throughout the semester. Homeworks will be due Mondays at 11PM.

Midterm Replication, Society Reflection, and Final Projects

The course is centered around a midterm replication project, a society reflection, and a final capstone project (completed in small groups). More details will be provided in the course. Key dates are included below.

- Midterm Replication Project:
  - Project Teams Decided: Sep 13
  - Project Due: Oct 11
- Society Reflection:
  - Reflection Due: Nov 21
- Final Project:
  - Project Ideas Due: Oct 18
  - Proposed Groups Due: Oct 22
  - Groups Finalized: Oct 25
  - Project Proposal Due: Nov 5
  - In Person Project Feedback: Nov 11
  - Project Results Presentation: Nov 19
  - Poster Presentation: Dec 4, 5, or 6
  - Final Paper Due: Dec 19 at 2:00PM

Note that the final project feedback and final project poster presentation may occur outside of our regular class time. The times will be determined well in advance of the dates.

Grading Scheme

Your course grade will be determined as follows (weightings of different project components will be detailed later):

5%	Attendance and Class Engagement
20%	Homework
25%	Midterm Exam
10%	Society Reflection
40%	Final Project (10% per component)

Grading is on an absolute scale (i.e., no curve). Letter grades will be assigned as shown below. A grade of A+ is awarded when the student demonstrates truly exceptional performance and is not simply determined by having a high final grade. I reserve the right to make adjustments.

F	D-	D	D+	C-	C	C+	B-	B	B+	A-	A	A+
< 60	60 - 62	63 - 66	67 - 69	70 - 72	73 - 76	77 - 79	80 -82	83 - 86	87 - 89	90 - 92	$\geq$ 93	*

Policies

Attendance and Class Engagement

Attendance to class is expected and contributes in a small way to your final grade. However, I do not expect you to attend if you are feeling unwell. Additionally, if you have athletics events, scheduled commitments, or other issues please let me know. If you miss a class, you can make up the credit by performing a check-in email. For a check-in, consult with a classmate and/or review the materials, and within 48 hours of the class send me an email containing:

1. a brief summary (2-4 sentences) of what was covered
2. any thoughts/questions you have

Check-ins are not a substitute for attending class regularly. If you accrue > 2 unapproved and/or unannounced absences, you will not receive attendance credit.

While in the class, you are expected to engage with the material and the other students in the course. You should aim to be a good participant: raising your hand, respecting others, actively listening, and making sure to leave space for others to speak. There are no bad questions, and I would always rather you contribute than avoid doing so. Mid-semester and end-of-semester evaluations will be administered with the intention to help me improve your experience in the course.

Time Management and Deadlines

Plan to give yourself time to step away from your work and return to it later. Programming in this way works better than sitting down and trying to complete programs in one shot right before the deadline. I will be reasonably flexible on deadlines. If you need some extra time due to illness, your workload in other classes, and/or personal matters, please let me know. As long as you have made a good faith effort to complete learning activities by the original deadline, I am willing to offer a reasonable extension. I will be less willing to grant an extension if you wait to start a homework assignment until the day before it is due, repeatedly ask for extensions, etc.

Any homework that is turned in late without prior approval will receive a 0%.

Academic Honesty & Collaboration

You are expected to abide by Colgate's academic honor code. You are not allowed to discuss the specifics of the programming assignments or share your code (or pseudo-code) with other students in any of the sections of this class (past or present). Assignment code should be discussed with the course TA or the instructors.

Collaborating with peers in the class

You may discuss course concepts, generic aspects of python, and work through the logic of something you don't understand with your peers. However, you should not share code (including psuedo code). Your submitted work should be your own. Here are some concrete rules that exemplify this (but are not intended to be comprehensive):

Do NOT:

- Ask a peer in either section of the course to debug your code.
- Ask a peer for pseudo code for an algorithm needed for a homework.

You CAN:

- Ask clarification questions about the fundamentals of programming (e.g., "How do I create a class in Python?")
- Ask for conceptual clarifications (e.g., "What is the difference between finite state automata and context-free grammars?")
- Try to work through the logic of something you don't understand (e.g., "How is data formatted for a neural network?")

Using generative AI tools

Generative AI systems (e.g., ChatGPT), if used correctly, can serve as powerful tools for learning and idea refinement. In this course, you can use generative AI systems to learn about concepts iteratively through a interaction. However, you cannot ask these systems to directly give you answers or write code for you. That is, you should submit your own work. Here are some concrete rules that exemplify this (but are not intended to be comprehensive):

Do NOT:

- Give the model a problem description and ask it to sketch an algorithm for you or write you pseudo code.
- Give the model the homework description and ask it to organize the code for you (e.g., generate the necessary function headers, write the main functions etc).
- Give the model a function description and ask it to generate code for you.
- Interact with the model and have your assignment open at the same time. Use your interaction with the AI as a learning experience, then close the interaction down, open your assignment, and let your assignment reflect your revised knowledge

Using the AI system in ways as described above will count as cheating even if you cite the AI system as a source.

You CAN:

- Ask clarification questions about the fundamentals of programming (e.g., "How do I loop through a file in Python?")
- Ask for conceptual clarifications (e.g., "What is the difference between recall, precision, and F1?")
- Try to work through the logic of something you don't understand (e.g., "How does the gradient tell the model which direction to move in during optimization?")
- Given a problem description and your proposed algorithm and 'talk' through the potential fallacies.

Remember: Policies around the use of Generative AI tools, like any other course policies, vary across different courses both within and outside the department.

Anonymous Feedback

Your feedback on this course is important for helping me improve the learning environment. You can provide anonymous feedback at any point in the semester via this here.

Getting Help

A key to your success at Colgate is figuring out what resources are available and using them to help you achieve your goals. There are several options for getting help with this course:

1. Drop in during my office hours (noted at the top of the syllabus) or if no office hours times work, arrange an appointment with me – just send me an email with a few times you are available, and we will find a time that works well for both of us.
2. Form a study group with other students in the class and work together on a regular basis (note the Academic Honesty & Collaboration policy above).
3. Attend TA office hours (to be announced).

I also encourage you to reach out to many great resources at Colgate that can assist you with academic, personal, or other needs, including:

- Administrative Deans (https://www.colgate.edu/about/offices-centers-institutes/dean-college/administrative-advising) help you understand policies and procedures, navigate personal challenges, work with faculty, and engage with parents.
- Counseling Center (https://colgate.edu/counseling) staff are trained to help students manage a wide array of emotions. The counseling center meets with over half the student body for clinical services at some point during their four years at Colgate. You can arrange an appointment online or by phone (315-228-7385). For emergencies, a counselor is available 24/7 by calling campus safety at 315-228-7333 and asking for the counselor on call.
- Haven (https://colgate.edu/haven) is a sexual violence response center that provides confidential care, support, advocacy, and trauma-informed clinical services for survivors of sexual assault, intimate partner violence, child/family abuse, stalking, and/or harassment. You can call (315-228-7385) or visit during business hours. You can also contact the Help Restore Hope Center (855-966-9723).
- Student Health Services (https://colgate.edu/offices-and-services/studenthealthservice) provides accessible, convenient, cost-effective, non-judgmental, and confidential care for all students.
- Information Technology Services (https://colgate.edu/its) help desk consultants assist all students with problems concerning email, Portal, Moodle, or your personal laptop. Contact me if problems with your personal computer are affecting your ability to get your work done.
- Chaplains (https://colgate.edu/campus-life/religious-life/officeofthechaplains) provide the community with a dynamic and friendly support.