COMPSCI 485: Application of Natural Language Processing

Semester Fall 2025

Credits: 3

Instructor: Abhidip Bhattacharyya (he/him)
abhidipbhatt@umass.edu
Office hours: TBD

TA: Anh Pham
acpham@umass.edu
Office hours: TBD

Course Description: **Natural Language Processing** (NLP) is the engineering art and science of how to teach computers to understand human language and perform tasks using data in natural language. NLP is a type of artificial intelligence technology, and it's now ubiquitous -- NLP lets us talk to our phones, use the web to answer questions, map out discussions in books and social media, generate chatbot responses, and translate between human languages. Since language is rich, ambiguous, and very difficult for computers to understand, these systems can sometimes seem like magic -- but these are science and engineering problems we can tackle with data, math, and insights from linguistics.

This course introduces NLP methods and applications including probabilistic models, text classification, linguistic representations, and contextual neural language models to process, understand, and generate text. During the course, students will-

1. learn the core methods of NLP;
2. become familiar with key facts about human language that motivate them, and help practitioners know what problems are possible to solve;
3. become equipped to engage critically with arguments about NLP's ethical and social implications; and
4. complete a series of hands-on projects to implement, experiment with, and improve NLP models, gaining practical skills for natural language systems engineering.

Pre-requisite: This course requires some familiarity with computer programming and probability theory at an intermediate undergraduate level. Formally, the course prerequisite is to fulfill either of:

- Both COMPSCI 220 and COMPSCI 240, with grade "C" or higher; OR
- LINGUIST 492B, with grade "C" or higher

Learning Objectives:

- Be comfortable with mathematical modeling applied to language, and how it can be implemented in software.
- Have practice and comfort with technical writing and poster presentation in the context of summarizing research and the student's own experiments.
- Know basic linguistic concepts and how to manually conduct linguistic analysis, including fundamental challenges in operationalization and construct validity.
- Be able to make informed arguments about social and ethical issues in NLP.
- Improve or introduce practical software skills for data analysis, machine learning, and linguistic processing of text.
- Practice collaboration skills through a group project.
- Critical analysis and reasoning about project results.
- Effective technical communication skills to present project results through written reports, a poster, and presentations.

Course Materials:

1. We will use **reading chapters from** two textbooks for the first half of the course. These books are available as pdf versions.
    - We'll be using draft chapters from the 3rd Edition of *Speech and Language Processing* by Dan Jurafsky and James H. Martin. You don't need to buy the current edition, draft pdfs of the new chapters are available from the textbook website. You can also download (and print, if you desire) the entire book from the website). It is also linked from Canvas.
    - We will have some reading from NLP notes of Jacob Einstein. (or book version; we'll have readings from this one occasionally, and it's recommended in general.)
2. For **advanced topics** reading materials will be mostly from research papers and will be provided accordingly.
3. **Course materials** will be hosted via canvas. This includes assignments, weekly quizzes and submission for in-class assignments and project deliverables.
4. **Course Announcements and Discussion Boards:** Official announcements for the course will go out through Canvas. Piazza will also host discussion forums for the class. Students will be added to Piazza by course staff at the start of the semester.

Course Topics:

Course Topics include

- Words, word counting, lexicons
- Morphology and Computational linguistics tools
- Probabilistic language modeling
- Text classification
- Token classification
- Syntax models

- Vector semantics & word embeddings
- Dependency parsing and constituency parsing
- Information extraction
- Generative models
- Model interpretation

** Topics are subject to change during the semester.**

Coursework and Grading Plan:
Grades are composed of:

- 10%: Exercises/quizzes (approximately weekly). Some of them will be in class exercise and will be graded with a simplified check/check-minus system. Quizzes will be graded in Canvas.
- 30%: Homework assignments (problem sets), which contain both written and programming portions. We plan to have four homework assignments.
- 30%: Midterm, planned to be in-class.
- 30%: Final project, including project proposal (10%), progress report (5%), in-class presentation (2.5%), and final report (12.5%). These are done in groups. (All other coursework is independent.)
    - The project presentation will likely be an in-class slide-based presentation. Final plan will be subject to logistical issues and announced later in the semester.

Final grades will be computed using the following grading scale:
A: 93-100%

A-: 90-92%

B+: 87-89%

B: 84-86%

B-: 80-83%

C+: 77-79%

C: 74-76%

C-: 70-73%

D+: 65-69%

D: 60-64%

F: below 59.5%

Minor revisions may be made to the grading rubric at the discretion of the instructor.

Course Policies:
Students should make sure they are familiar with all course policies and the relevant University policies linked to below. By staying enrolled in this course, students agree to be bound by all applicable policies.

- **Course Community Code of Conduct:** The instructor and the course staff are committed to providing a friendly, safe and welcoming environment for all, regardless of gender identity and expression, sexual orientation, disability, personal appearance, body size, race, ethnicity, age, religion, nationality, or other similar characteristic. Please be courteous, respectful, and professional in all of your interactions with other students, TAs, and graders in all mediums of communication including but not limited to in person, email, video meetings, chat, discussion forums, and re-grade submissions.

  Demeaning, insulting or harassing any member of the course community over any medium of communication is not acceptable behavior, including in person, through official course platforms and through personal/private platforms (social media, email, DM, text, etc.). Students who engage in such behavior will be warned at most once before the behavior is reported to the Dean of Students office. If you feel you have been or are being harassed or made uncomfortable by a member of this course community, please contact a member of the course staff immediately (or if you do not feel safe doing so, contact the Chair of the Faculty of CICS, Erik Learned-Miller (chair@cs.umass.edu), or the Dean of Students office). We care about making this course a safe and welcoming place for all.

- **Class Attendance Policy:** Students are expected to attend all class meetings unless they are feeling ill. Class participation will be measured with in-class activities and contribute to the final grade.

  For in-class assignments, students must be present to submit and receive credit. No grade will be given for in-class activities if the student is absent.

  Each student is allowed up to **three excused absences** for any reason, without needing the instructor's permission. However, **these absences may not be used on days when project deliverables are due**, such as the proposal or final presentations.

  Any absences beyond the three allowed will result in a **grade penalty**.

- **Exam Absence Policy:** A makeup exam time will be provided to students who are unable to attend regular scheduled exams according to University policy (e.g., in the case of illness, religious observances, official University travel, and other extenuating circumstances). Note that a makeup exam will be provided in the case of official University travel (e.g., to present at a conference), but conflicting research deadlines (e.g., a conference paper submission deadline) are not grounds for requesting a makeup exam. When students are aware of an exam conflict ahead of time, they should contact the course staff in writing as soon as possible and no later than one week before the exam date to arrange a time for a makeup exam. In the case of illness or other unforeseen

extenuating events, students should contact course staff in writing when they are able. Providing documentation for unforeseen exam absences is greatly appreciated.

- **Late Homework Policy:**
  - For homework assignments:
    - Everyone has **three late days** (in total over the semester) to use for homework assignments.
    - After all late days have been exhausted, no more non-excused late submissions will be accepted.
  - We follow University policy in accepting work late due to health, personal, and religious holiday circumstances, which are extensions that do not use late days.
  - Arrangements in non-emergency circumstances must be made in advance (at least 24 hours) with course staff.
  - Job interviews and other schoolwork are not acceptable excuses for late homework.
  - Project assignments will not be accepted late.
  - In case of unforeseen and emergency circumstances (e.g. illness, internet issues, personal emergency), please contact the instructor when you're able to.

- **Re-grading Policy:** Errors in grading can occur despite the best efforts of the course staff. If you believe you've found a grading error, please submit a re-grade request. Re-grade requests must be submitted no later than one week after the graded material is returned. Note that re-grading may result in your original grade increasing, decreasing or remaining unchanged as appropriate.

- **Academic Honesty and Homework Collaboration Policy**:
  - For all non-group (non-project) assignments, all of the content you submit, both code and text, needs to be produced independently. Your work must be in your own words and based on your understanding of the solution. ("Independently" applies to homework; for group projects, work must be produced only by members of the group.)
  - The following are disallowed and extremely likely to result in an academic honesty violations:
    - Sharing completed or in-progress reports or code with another student
    - Copying all or part of completed or in-progress reports or code requested from another student
    - Using another student's code or code output to help debug your code
    - Working with another student while you both write-up or program a solution to a problem § Posting completed solutions on public code repositories (e.g., GitHub, etc.) during or after the course
    - Copying any solution materials (derivations, code, method descriptions) in whole or in part from external sources. Disallowed external sources include books, web pages, homework "help" services such as Chegg, automated generation such as ChatGPT, etc.
  - On your homework, list the names of everyone you collaborated or had discussions with.

- If you find, use, or build off of published material, for example on the web or from a textbook, you must cite the source. Always explain the answer in your own words.
- For written assignments, all your writing must be original or properly paraphrased, with citations as appropriate.
- Detected cheating on homework assignments will result in a grade penalty on the assignment up to and including the full value of the assignment. Any misuse of course platforms is grounds for an F in the course. Cheating on exams is grounds for an F in the course.
- We follow the university's Academic Honesty Policy and Procedures. If you are not aware, note that academic honesty is significantly more important than completely finishing assignments or being late. We will report and sanction instances of dishonesty with no leniency. If you have questions about a particular situation, please ask.
- Some examples of the collaboration policy:
  - Acceptable: Alice and Bob discuss alternatives for storing large, sparse vectors of feature counts, as required by a problem set.
  - Unacceptable: Alice and Bob sit down together and write code for storing feature counts.
  - Acceptable: Bob is confused about how to implement the Viterbi algorithm, and asks Alice for a conceptual description of her strategy.
  - Unacceptable: Alice and Bob divide the assignment into parts, and each write the code for their part, and then share their solutions with each other to complete the assignment.
  - Acceptable: Alice asks Bob if he encountered a failure condition at a "sanity check" in a coding assignment, and Bob explains at a conceptual level how he overcame that failure condition.
  - Unacceptable: Alice or Bob obtain a solution to a previous year's assignment or to a related assig

Course Outline:

week	topic	assignment
1	Lecture 1: Introduction, course description	Hw1
2	word, token, normalization, morphplogy Language models	HW2-LM
3	Language Model- classifiers Naive Bayes classifiers	
4	Logistic Regression and SGD	HW3- NB

		Form group
5	Tokens classification and HMM	
6	Parsing 1. Constituency parsing 2. Dependency parsing	HW4- Bert
7	Midterm Word representation and NN	
8	RNNS, ELMO and Attention Transformers, BERT	Proposal and HW5(paper reading)
9	● Proposal ● Pretraining and Generative LLMs	
10	Prompting IR and RAG	
11	Model Interp Multimodal LLMs	
12	Break	
13	Final Presentation	
13	●	

Academic Honesty Policy:

Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructor should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course

instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent (https://www.umass.edu/honesty/).

Accommodations Statement:
The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify me within the first two weeks of the semester so that we may make appropriate arrangements. For further information, please visit Disability Services (https://www.umass.edu/disability/)

Title IX Statement (Mandated Reporter Version): **(Please use the applicable version)**
In accordance with Title IX of the Education Amendments of 1972 that prohibits gender-based discrimination in educational settings that receive federal funds, the University of Massachusetts Amherst is committed to providing a safe learning environment for all students, free from all forms of discrimination, including sexual assault, sexual harassment, domestic violence, dating violence, stalking, and retaliation. This includes interactions in person or online through digital platforms and social media. Title IX also protects against discrimination on the basis of pregnancy, childbirth, false pregnancy, miscarriage, abortion, or related conditions, including recovery. There are resources here on campus to support you. A summary of the available Title IX resources (confidential and non-confidential) can be found at the following link: https://www.umass.edu/titleix/resources. You do not need to make a formal report to access them. If you need immediate support, you are not alone. Free and confidential support is available 24 hours a day / 7 days a week / 365 days a year at the SASA Hotline 413-545-0800.

For purposes of Title IX reporting, I am a considered a "responsible employee" at UMass (https://www.umass.edu/titleix/about). **That means if you tell me about a situation involving sexual assault, sexual harassment, domestic violence, dating violence, stalking, and retaliation, I must share that information with the Title IX Coordinator**. Making a report to the Title IX Coordinator is my legal obligation, meets the University's goal of providing members of our community with supportive resources they might need, and enables the University to obtain a more accurate picture of the extent of sexual violence in our community. **It will be completely up to you to determine if and how you want to work with the Title IX Coordinator's office. You will not be in trouble for reporting to me that you have**

experienced any of these situations, and the law prohibits retaliation against anyone who participates in a Title IX process.