CMSC470 Syllabus - file last edited 1/26/2025

UNIVERSITY OF MARYLAND

Course Information

Course Title: Introduction to Natural Language Processing
Course Number: CMSC470
Term: Spring 2025
Credits: 3
Course Dates: 1/28/25 – 5/13/25
Course Times: TuTh 12:30pm - 1:45pm
Location: CSI 3117
Final Exam: May 20, 2025 4pm-6pm

Professor: Rachel Rudinger (she/her/hers)
Email: rudinger@umd.edu
TA: Fuxiao Liu
Email: fl3es@umd.edu
TA: Ishani Mondal
Email: imondal@umd.edu
TA: Shramay Palta
Email: spalta@umd.edu
Office Hours: TBA

Course Description

Welcome to CMSC470!

This is an upper-level undergraduate elective course on natural language processing (NLP). This course will introduce fundamental concepts and techniques for automatically processing and generating natural language with computers. We will study the machine learning techniques, models, and algorithms that enable computers to deal with the ambiguity and implicit structure of natural language. We will apply these techniques in a series of homework assignments.

Prerequisites

Minimum grade of C- in CMSC320, CMSC330, and CMSC351; and 1 course with a minimum grade of C- from (MATH240, MATH461); and permission of CMNS-Computer Science department.

We will work extensively with probability, statistics, mathematical functions such as logarithms and differentiation, and linear algebra concepts such as vectors and matrices. You should be comfortable manipulating these concepts.

We will make extensive use of the Python programming language. It is assumed that you know or will quickly learn how to program in Python. The programming assignments will be oriented toward Unix-like operating systems (Linux, OS X). It may be possible to complete the course using other operating systems, but you will be responsible for troubleshooting any issues you encounter.

We will introduce linguistics concepts as required throughout the course, so a background in linguistics is not required.

Here are some reference materials that you might find useful if you'd like to refresh your memory:

- Probability primer
- Review of calculus and linear algebra for machine learning

- Basic Python/Unix tutorial
- Core Unix tools to work with text data (aka Unix for Poets)

These resources are also available on ELMS in the "Helpful Primers and Tutorials" sub-folder.

Class Format

CMSC470 will meet in person in CSI3117. Lectures will not be recorded. To get the most out of this course, I strongly encourage you to attend all in-person lectures. If you are unable to attend a lecture due to illness, there are several steps I recommend you take to catch up once you have recovered. These include: reviewing the lecture slides (which will be posted to ELMS), doing any accompanying readings or practice problems, and attending office hours to resolve any unresolved questions you may have about the content. **New for Spring 2025**, we will have a small (low-stakes) in-class quiz at the start of each Tuesday class based on feedback from prior semesters; read the section on In-Class Quizzes for more details. The course will also entail homework a mix of coding and non-coding homework assignments, two in-class midterms, and a final exam.

Online Resources

- Course Website: elms.umd.edu
- Piazza – Course discussion and questions page will be available on Piazza. You can access the course Piazza through ELMS.
- Gradescope – Programming assignments will be done through gradescope. Instructions and links will be available through ELMS.
- Zoom – If any sessions are to be held over Zoom, the associated link will be posted under the "Zoom" tab on ELMS.
- All other course resources including this syllabus, links to readings, homework assignments, practice problems, and slides will be available through the ELMS course page.

Piazza

Announcements, discussions, and questions will be through Piazza. Please sign up for the course on Piazza with the signup link. If you have a question about an assignment, you are encouraged to set the post as "public" to the class so others with the same question may also view the response, though of course private posting is also acceptable for questions of a personal nature. Posts may be made anonymous to other students but not to instructional staff.

Readings

All readings will be selected from resources that are available online freely for UMD students.
PDF copies of these books are located in the Readings sub-folder on ELMS.
Primary resources:

- Speech and Language Processing (3rd edition draft) by Dan Jurafsky and James H. Martin
- A Course in Machine Learning by Hal Daumé III

Additional resources:

- Natural Language Processing by Jacob Eisenstein
- Natural Language Processing with PyTorch: Building Intelligent Language Applications Using Deep Learning , by Brian McMahan and Delip Rao
- Linguistics Fundamentals for NLP by Emily Bender

Assignments

The assignments for this course will include a number of non-coding and coding homework assignments (roughly 4-5 total). The homework assignments are to be completed individually (unless otherwise indicated). For homework assignments, discussion is permitted; **however, all solutions must be written up independently.**

In-Class Quizzes

Each Tuesday class will start with a short quiz (~5 minutes) consisting of 1-2 comprehension questions pertaining to content from the previous two lectures, or possibly prior lectures. The purpose of these quizzes is to encourage you to stay on top of the material, to help you prepare for exams, and to get more frequent feedback on your comprehension. These quizzes will constitute no more than 5% of your final grade. Because each quiz is a very tiny percentage of your final grade, and answers will be shared immediately afterwards, make-up quizzes will not be offered for any reason. Instead, to accommodate for inevitable conflicts/missed quizzes, I will drop the lowest ~25% of your quiz scores.

Why quizzing? Frequent low-stakes testing has been scientifically shown to improve learning outcomes. This is because (1) testing your knowledge gives you a more accurate sense of how well you have learned material (while merely reviewing content can create a false sense of mastery); (2) spaced repetition of concepts aids long-term retention; and (3) the act of *retrieving* knowledge from memory (rather than just reviewing it) reinforces learning. This is also in direct response to feedback from prior classes requesting more small-stakes assessments. The primary purpose of these quizzes is **not** testing as an end in itself; rather, it is testing as a means to better learning outcomes. If you are interested in learning more about the cognitive science behind testing and effective learning, I highly recommend the book Make It Stick. This is a new/experimental course feature, and I welcome your feedback.

Grading

In-Class Quizzes	5%
Homework	30%

In-Class Midterm 1	20%
In-Class Midterm 2	20%
In-Person Final	25%

Late Policy

Over the course of the semester, you have eight late days **total** to use across any homework assignments, with a **maximum of two late days per assignment**. Late days may NOT be used on any exams. Late days are rounded UP, so an assignment submitted 24 hours and 1 minute past the deadline will count as two late days. Late days may be used at your discretion, but the purpose of the late days is to give you flexibility when illness or other unforeseen emergencies arise, so you are **very strongly encouraged** to save your late days for such situations.

Midterms and Final Exam

There will be two in-class midterms and an in-person final exam. In accordance with the University exam schedule, the class final will be held on May 20, 4pm-6pm.

Course Policies and Academic Accommodations Office Hours

Each member of the instructional staff (Professor and TAs) will typically host one hour of office hours per week. The time and location of these sessions will be announced after the start of the semester.

Course Policies and Academic Accommodations

We will follow the standard University of Maryland Undergraduate Course Policies . Please familiarize yourself with them if you are not already.

Any student eligible for and requesting reasonable academic accommodations due to a disability is asked to provide, to the instructor by email, a letter of accommodation from the Accessibility and Disability Service (ADS) office **within the first two weeks of the semester**. More details about these resources are in the section below on “Resources & Accommodations.”

If you plan to observe any holidays during the semester that are not listed on the university calendar, please provide a list of these dates **by the end of the first two weeks of the semester**.

Academic Integrity

The University's **Code of Academic Integrity** is designed to ensure that the principles of academic honesty and integrity are upheld. In accordance with this code, CMNS does not tolerate academic dishonesty. Please ensure that you fully understand this code and its implications because all acts of academic dishonesty will be dealt with in accordance with the provisions of this code. All students are expected to adhere to this Code.

Course Policy on the Use of Artificial Intelligence

Unless otherwise stated, you are permitted to use AI language models such ChatGPT and Llama to assist with submitted course assignments. However, you must carefully read and adhere to the following rules regarding the use of AI in this course:

(1) If you use AI to assist with written responses, you must disclose this fact. The disclosure must include the following information:
   a. What model(s) you used.
   b. Date(s) when the model was accessed.
   c. Briefly describe how you used the model to help you answer the question. (E.g., "I used the model's output verbatim," "I lightly edited the model's output," "I made substantial changes to the model's output," "I used the model's output for ideas but wrote the response myself," etc.

(2) If you use AI to assist with coding assignments, you must also disclose this fact BOTH with in-line or block comments indicating which line(s) of code are partially or fully AI-generated, and a separate written disclosure with the same information listed above.

(3) You may NOT use outputs that somebody else got from a model, and you may NOT use another student's prompts or inputs to a model. Similarly, you may not share your answers or model outputs with other students, in accordance with class policy. Discussion of homework questions is, however, permitted.

(4) Failure to follow guidelines (1), (2), and (3) will be considered a violation of academic honesty.

(5) For your own benefit, you are **strongly encouraged** to screen-capture or otherwise save/document the original model outputs that you used to assist with the assignment. You do not need to submit these files, but you should retain them until final grades are posted at the end of the semester. In the event that a question of academic honesty or copying work arises, this will help you document your work process.

(6) For graded assignments in which the use of AI is permitted, you will not be penalized nor will you receive bonus points for the use of AI. All submitted work will be graded on its substance and quality. Be aware that AI-generated can sound convincing while being entirely incorrect. It is your responsibility to ensure the correctness of your submitted work.

(7) You take full responsibility for all written answers and code you submit in this course, **regardless** of whether you used an AI to assist in writing it. Be aware that code or text generated by language models may be unsafe (e.g., generated code may contain malware, and generated text could contain language constituting hate speech). It is your responsibility to review the contents of your submitted work for possible violations of course, department, or university policy, or any other laws or regulations you are subject to.

(8) You do not need to disclose use of AI on non-graded written communication (e.g. private email, Piazza posts, etc.), but all messages posted or sent by you will be construed as your own speech and subject to the same code of conduct expectations as outlined in this syllabus and in university policy.

(9) **The use of AI will NOT be permitted on any in-class exam.**

This is a new, experimental policy on AI and is subject to change during the semester. If these rules change, I will clearly communicate the change before releasing any assignments to which

the new policy applies. If you have any confusion or concerns regarding this policy, please do not hesitate to ask for clarifications.

Course Guidelines

Names/Pronouns and Self-Identifications
The University of Maryland recognizes the importance of a diverse student body, and we are committed to fostering inclusive and equitable classroom environments. I invite you, if you wish, to tell us how you want to be referred to both in terms of your name and your pronouns (he/him, she/her, they/them, etc.). The pronouns someone indicates are not necessarily indicative of their gender identity. Visit trans.umd.edu to learn more.

Additionally, how you identify in terms of your gender, race, class, sexuality, religion, and dis/ability, among all aspects of your identity, is your choice whether to disclose (e.g., should it come up in classroom conversation about our experiences and perspectives) and should be self-identified, not presumed or imposed. I will do my best to address and refer to all students accordingly, and I ask you to do the same for all of your fellow Terps.

Communication with Instructor:

Email: If you need to reach out and communicate with me, please email me at rudinger@umd.edu.

ELMS: I will send IMPORTANT announcements via ELMS messaging. You must make sure that your email & announcement notifications (including changes in assignments and/or due dates) are enabled in ELMS so you do not miss any messages. You are responsible for checking your email and Canvas/ELMS inbox with regular frequency.

Communication with Peers:

With a diversity of perspectives and experience, we may find ourselves in disagreement and/or debate with one another. As such, it is important that we agree to conduct ourselves in a professional manner and that we work together to foster and preserve a classroom environment in which we can respectfully discuss and deliberate controversial questions.

I encourage you to confidently exercise your right to free speech—bearing in mind, of course, that you will be expected to craft and defend arguments that support your position. Keep in mind, that free speech has its limit and this course is NOT the space for hate speech, harassment, and derogatory language. I will make every reasonable attempt to create an atmosphere in which each student feels comfortable voicing their argument without fear of being personally attacked, mocked, demeaned, or devalued.

Any behavior (including harassment, sexual harassment, and racially and/or culturally derogatory language) that threatens this atmosphere will not be tolerated. Please alert me immediately if you feel threatened, dismissed, or silenced at any point during our semester together and/or if your engagement in discussion has been in some way hindered by the learning environment.

Resources & Accommodations

Accessibility and Disability Services
The University of Maryland is committed to creating and maintaining a welcoming and inclusive educational, working, and living environment for people of all abilities. The University of Maryland is also committed to the principle that no qualified individual with a disability shall, on the basis of disability, be excluded from participation in or be denied the benefits of the services, programs, or activities of the University, or be subjected to discrimination. The **Accessibility & Disability Service (ADS)** provides reasonable accommodations to qualified individuals to provide equal access to services, programs and activities. ADS cannot assist retroactively, so it is generally best to request accommodations several weeks before the semester begins or as soon as a disability becomes known.

For assistance in obtaining an accommodation, contact Accessibility and Disability Service at 301-314-7682, or email them at **adsfrontdesk@umd.edu**. Information about **sharing your accommodations with instructors**, **note taking assistance** and more is available from the **Counseling Center**.

Student Resources and Services
Taking personal responsibility for your own learning means acknowledging when your performance does not match your goals and doing something about it. I hope you will come talk to me so that I can help you find the right approach to success in this course, and I encourage you to visit **UMD's Student Academic Support Services website** to learn more about the wide range of campus resources available to you.

In particular, everyone can use some help sharpening their communication skills (and improving their grade) by visiting **UMD's Writing Center** and schedule an appointment with the campus Writing Center.

You should also know there are a wide range of resources to support you with whatever you might need (**UMD's Student Resources and Services website** may help). If you feel it would be helpful to have someone to talk to, visit **UMD's Counseling Center** or **one of the many other mental health resources on campus**.

Basic Needs Security
If you have difficulty affording groceries or accessing sufficient food to eat every day, or lack a safe and stable place to live, please visit **UMD's Division of Student Affairs website** for information about resources the campus offers you and let me know if I can help in any way.

Course Evaluation
Please submit a course evaluation through CourseEvalUM in order to help faculty and administrators improve teaching and learning at Maryland. All information submitted to CourseEvalUM is confidential. Campus will notify you when CourseEvalUM is open for you to complete your evaluations for fall semester courses. Please go directly to the **Course Eval UM website** to complete your evaluations. By completing all of your evaluations each semester, you will have the privilege of accessing through Testudo, the evaluation reports for the thousands of courses for which 70% or more students submitted their evaluations.