CS 232: Artificial Intelligence

**Class Meetings:**
T F 9:55-11:10
Sci L039

**Instructor:** Carolyn Anderson
carolyn.anderson@wellesley.edu

**Drop-in hours:**
M Th 4-5:30
Sci W422

**Tutors:**
Heeba Nazeer & Lepei Zhao

**Tutor hours:**
TBA

**Grader:**
Cynthia Wang

**Announcements:**
We will use the course Google Group for course announcements.

**Course Description**
What is artificial intelligence (AI) and should humans fear it as one of "our biggest existential threats"? In this course, we will grapple with these difficult questions and investigate them in different ways. We will discuss the development of the field from the symbolic, knowledge-rich approaches of 20th century AI (e.g., rule-based systems), to statistical approaches that rely on increasingly large amounts of data, including an overview of contemporary deep learning techniques. We will explore how to apply these techniques in several AI application areas, including robotics, computer vision, and natural language processing, and consider ethical issues around AI in society. By the end of the semester, students should be able to answer the starting questions in-depth and with nuance.

**Distributions:** MM  
**Prerequisites(s):** CS 230

**Learning Goals**
- Learn about a range of AI approaches, including rule-based/symbolic systems, statistical approaches, machine learning, and deep learning.
- Learn to recognize and choose appropriate techniques for a range of problems, including search, classification, and generation.
- Implement AI techniques in a variety of domains, including game-playing, robotics, computer vision, and natural language processing.
- Evaluate the performance of AI models with respect to both scientific validity and societal impact.
- Critically consider the ethical consequences of current and future AI technology.

**Flexibility and Feedback**
I welcome feedback throughout the course. You can email me, or, if you prefer to be anonymous, you can use the [Anonymous Question Form](#).

**Course Work**
Expectations

- You are expected to come to class and to participate. However, you should not come to class if you are feeling sick; I’ll work with you to catch up after you are feeling better.
- Whenever you have questions about any course material, whether it is current material or from earlier in the course, please submit questions using the Anonymous Question Form. I strive to answer these questions within a day in the Q&A Document.
- You are expected to seek help when you have trouble with the course material. Sometimes students are embarrassed to seek help, especially if they are far behind. I am happy to help you with any course material, no matter how long ago it was taught.

Assignments

There will be 10 two-week homework assignments, typically composed of:

- short answer questions that ask you to reflect on the weekly readings
- short math questions to reinforce the theory behind the techniques we are studying
- programming exercises that you ask to implement AI algorithms or models
- evaluation questions that ask you to analyze the performance of a model that you have implemented

You should expect roughly 50% of each assignment to be Python programming.

Assignments will be due on Mondays at 10pm.

Final Project

In lieu of a final exam, there will be a final project. You will pick one of two topics: (1) cultural biases in large language models, or (2) the impact of language variation on language model performance. You will research one of these aspects and work to develop a probe task that can be applied to a neural network language model to investigate one aspect of the topic. Once you have developed your individual probe task, you will work with other group members to compile your findings into a research report. This report will be due at the end of finals period. Your group will present its findings in a short presentation on the last day of class.

Assignment Submission

We will use Gradescope for homework submissions. Some homework problems will involve writing text, rather than code. You may intersperse these answers as comments in your Python program, or submit them as a separate PDF.

Guest Lectures

I have invited AI researchers to give research talks as part of this course. Attending these talks is mandatory. Some lectures may need to be held outside of normal class hours.

Readings

The main textbook for this course is You Look Like A Thing and I Love You by Janelle Shane. Let me know if you have difficulty obtaining a copy. Technical readings will come from the
3rd edition of *Artificial Intelligence: A Modern Approach* by Peter Norvig and Stuart Russell. There is a PDF online. Other occasional readings will be linked to on the course schedule.

**Course Policies**

**Late Policy**

- Assignments will typically be due at **10pm Monday each week**.
- **You have 5 late days for the semester.** You can use them all for one assignment, or spread them out across the semester. **I will not accept late work beyond these days.**
- **I will not answer questions on late work during help hours.**
- **If you are falling behind, email me as soon as possible.** We will figure out a plan that supports your learning and your well-being. **You do not need to explain any personal details to justify your request.**

**Collaboration Policy and the Honor Code**

- **Collaboration:**
  - You **may** discuss high-level ideas or strategies with other students, but you **must not** communicate detailed algorithms, implementations, or formulae.
  - **Wait 30 minutes after discussions with other students** before writing your solution. This helps you know if you actually understand the solution.
  - **You may not share code from homework problems with other students at any point in the course (even after the assignment deadline).**
- **Reference:**
  - You may consult external reference resources for general concepts and techniques, but you **must cite them.**
  - You must **not** consult solutions to this or any similar assignment from other NLP courses, books, or online resources.
- **Code reuse and adaptation:**
  - **You must not reuse or adapt any code beyond what is given to you in the assignment and in CS232 lectures.**
- **After the semester:**
  - **Do not post homework solutions publicly** (i.e. on GitHub or your website). You may show your work to potential employers or others who aren’t students.

Some of these rules will be relaxed for the final assignment, which is a group project.

**Grading**

- **Homework** 80%
- **Final project:**
  - Prep work 5%
  - Probe task 5%
  - Presentation 5%
  - Report 5%

This course will comply with the [Wellesley College grading policy](#).
Seeking Help

Lots of help is available in this course!

- If you go 20-30 minutes without making progress, please ask for help!
- You are encouraged to post questions to the Anonymous Question Form. However, you may not post code. For code-specific questions, email me directly.
- Come to tutor hours! You don’t need a specific question to attend.
- Come to my drop-in hours! You don’t need to have a specific question. I’m also happy to talk more generally about NLP if no one is waiting for homework help. If the posted time doesn't work for you, you can book a one-on-one meeting using this calendar.
- You can request a one-on-one tutor from the PLTC.

Disabilities and Accommodations

My job is to help you learn, and part of that is making accommodations for any disabilities you might have. Please feel free to offer suggestions about how I can make the course more accessible to you. I will keep the details of our conversations confidential. You are welcome to share anything that will be helpful, but you should not feel pressure to share details with me.

If you would benefit from reasonable academic adjustments, I recommend contacting Accessibility and Disability Resources (ADR) to get a letter outlining your accommodation needs as soon as possible, as some may take significant time to arrange. If you need immediate accommodations, please meet with me ASAP. If you are unsure but suspect you may have an undocumented need for accommodations, ADR can help with referrals for assessments.

Disclosures of Discrimination and Sexual Misconduct

Wellesley College considers diversity essential to educational excellence, and we are committed to being a community in which each member thrives. The College does not allow discrimination or harassment based on race, color, sex, gender identity or expression, sexual orientation, ethnic or national origin or ancestry, physical or mental disability, pregnancy or any other protected status under applicable local, state or federal law.

If you or someone you know experiences discrimination or harassment, support is available:

- Confidential reporting:
  - Health Services (781.283.2810), the Stone Center Counseling Service (781.283.2839), and Religious and Spiritual Life (781.283.2685) are not required to report allegations of sexual misconduct to the College.
- Non-confidential reporting:
  - You can let me know. As a faculty member, I am a mandatory reporter: I am obligated to report allegations of sex-based discrimination to the Title IX Office.
  - You can report directly to the Nondiscrimination/Title IX Office (781.283.2451) to receive support, and to learn more about your options for reporting.
  - You can report to the Wellesley College Police (Emergency: 781.283.5555, Non-emergency: 781.283.2121) if you believe a crime has been committed.