COGS 8X.XX: Models of Language and Conversation

Meeting times	TBD
X-hour	TBD
Classroom	TBD
Instructor	Dr. Jeremy R. Manning
Email	jeremy@dartmouth.edu
Office location	349 Moore Hall
Office hours	By appointment

Course Description

Recent advancements in Artificial Intelligence (AI) have led to the development of powerful large language models (LLMs) capable of generating human-like text. These models have revolutionized the way we think about language, cognition, and intelligence. This course introduces students to the various approaches used in building, studying, and using conversational agents. We will use readings, discussions, and hands-on demonstrations to explore historical approaches, current approaches, and hints about the "next generation" of these models.

We will investigate what these models can and cannot do from various cognitive, philosophical, computational, and ethical perspectives. A special focus will be on which aspects of human cognition, experiences, minds, and language these models can vs. cannot capture.

Course Goals

This course is intended to train students to: - Explore a variety of approaches to constructing conversational agents, including string manipulation models, rules-based (decision tree) models, and transformer-based models. Students will implement and train these models, largely from scratch, to gain a deep understanding of how they work. - Understand different classic and modern approaches to text embeddings (LSA, LDA, word2vec, USE, BERT, etc.) and dimensionality reduction (PCA, UMAP). Students will use these approaches to carry out analyses of text corpora. - Gain a general understanding of computational linguistic text analyses including part of speech tagging, sentiment analysis, etc. Again, students will implement and apply these tools to text data. - Explore the psychological and cognitive applications of LLMs, such as modeling thought processes, language comprehension, and social cognition - Critically assess the abilities and limitations of conversational agents in replicating human-like reasoning and understanding - Discuss the philosophical and ethical implications of AI, including questions around consciousness, agency, and the future of AI - Develop and implement experiments using LLMs to explore research questions in psychology and cognitive science

Pre-Requisites

Students must have prior experience with Python programming in order to enroll in this course. Prior coursework on statistics or probability (e.g., PSYC 10, AP Stats, or similar) is also highly recommended. An online statistics course is fine as a stand-in, but I will expect you to know about hypothesis testing, probability distributions, and have some intuitions about how to design and interpret statistical analysis before you start this course. Additional prior coursework and/or experience with linguistics, linear algebra, statistics, machine learning, artificial intelligence, data science, philosophy of mind, and/or creative writing will all be useful, but are not required.

Course Materials

We will use a variety of freely available online materials and research papers, which will be provided throughout the course. You will also need an internet-enabled computer or tablet capable of displaying and outputting

graphics and running a standard web browser (e.g., Chrome or Firefox).

Format and Overview

This course follows an experiential learning model. Students will engage with lecture materials through hands-on programming exercises, experiments with models and tools, and group discussions. Problem sets will deepen your understanding of the course material, and small projects will allow you to apply concepts to real-world research problems in the relevant domains.

Classes will include: - Lectures: Foundational topics, discussions on implementing and leveraging conversational agents, text embeddings, and text analysis. - Hands-on Labs: Experimenting with conversational agents, designing simple conversational agents, and conducting research using these models. - Problem Sets and Projects: Bi-weekly problem sets to apply what you've learned. These will typically take the form of small-scale "research" projects.

Platforms and Tools

- Google Colaboratory: For developing and running Python code.
- Hugging Face Models: We will use pre-trained models available through Hugging Face's model hub.
- GitHub: Used for managing and sharing code, data, and project work.
- Discord: To facilitate discussions outside of class, share ideas, and collaborate on projects.

Grading

Grades will be based on the following components:

- Problem Sets (75%): A total of 5 problem sets designed to reinforce key concepts (each is worth 15% of the final course grade).
- Final Project (25%): You will carry out a larger scale (relative to the problem sets) "research" project on a topic of your choosing. This will include:
  - Python code, organized as a Colaboratory notebook
  - A "presentation" to the class (also submitted as a YouTube video), along with an in-class discussion of your project
  - A brief (2–5 page) writeup of the main approach and key findings or takeaways

Students may work together on all of the assignments, unless otherwise noted in class or in the assignment instructions. However, each student must submit their own problem set and indicate who they worked with. Final projects will (typically) be completed in groups of 2–3 students, with the entire group turning in the same project (and receiving the same grade for it).

Grading Scale: A (93–100), A- (90–92), B+ (87–89), B (83–86), B- (80–82), C+ (77–79), C (73–76), C- (70–72), D (60–69), E (0–59). All grades will be rounded to the nearest integer (e.g., a 92.5 avearge will result in a final grade of "A", whereas a 92.4999 average will result in a final grade of "A-"). Out of fairness to all students in the course, there will be no "negotiations" about grading– e.g., your grade will be determined solely by the numerical average of your assignment scores.

Late Policy

Problem sets will receive a 10% deduction for each week late, rounded up to the nearest whole week (e.g., from a grading standpoint submitting an assignment 1 minute late is the same as submitting it 1 day late, is the same as submitting it 6 days late).

Your final project must be submitted on time in order to receive credit for it.

I strongly encourage you to submit your assignments before the last possible moment to avoid grading penalties, unexpected circumstances (e.g., illness, emergencies, etc.).

Academic Honor Principle

Students are expected to adhere to Dartmouth's Academic Honor Principle. You are encouraged to collaborate and discuss ideas with classmates, but all submitted work must be your own (aside from the final projects, which will be completed in small groups). If you're unsure about what constitutes a violation, please ask for clarification.

Use of Generative AI

Given that this is a course about the technologies used to make modern generative AIs like ChatGPT, Claude, Llama, Gemeni, etc., you will be using these models extensively in class and in your assignments. There are just two "rules" about using GenAI: - Most importantly, you are responsible for the content of your assignments, whether written by you "from scratch" or with the help of a GenAI agent. - Second, you are bound by the Academic Honor Principle to acknowledge any use of GenAI in your work. This can be done by either using a brief comment in your code, by explicitly citing the tools you use, or by adding a note to the relevant section(s) of your assignment. Each situation is unique, but you need to make it clear exactly what work is your own vs. produced by GenAI. You must also include a chat history (including any prompts you used) as an addendum to your assignment(s).

Scheduling Conflicts

Attendance is expected for all classes unless previously arranged. A critical part of the course is the in-class discussions and demos, and those will only work if you are physically present in class. If you anticipate any conflicts due to religious observances or other commitments, please inform the instructor by Week 2 to make appropriate arrangements.

Student Needs

We strive to create an inclusive learning environment where all students feel supported and engaged. If you require any accommodations, please contact the Student Accessibility Services office, or discuss your needs with the instructor privately.