Syllabus

CS 678

Advanced Natural Language Processing

Instructors

Antonios Anastasopoulos (antonis [at] gmu [dot] edu)
Office Hours: TBA.

TA

TBA (tba [at] gmu [dot] edu)
Office Hours: TBA

Meets

Tuesday, 4:30 pm - 7:00 pm, 2016 Horizon Hall.

Course Web Page

https://nlp.cs.gmu.edu/course/cs678-fall24/.
We will use **Canvas** and **Gradescope** for course materials/assignments/grading, and **Piazza** for Q&A (sign up link: https://piazza.com/gmu/fall2024/cs678).

Course Description

Massive amounts of information in our daily life are expressed in natural language. In this class, we will study building computing systems that can process, understand, and communicate in natural language. The class will start with an introduction to the foundations of natural language processing (NLP) and relevant theory, and then focus on cutting-edge research problems in NLP. Each section will introduce a particular problem or phenomenon in natural language, describe why it is difficult to model, and demonstrate recent models that were designed to tackle this problem. In the process of doing so, the class will cover different techniques that are useful in creating and applying neural network models. The class will include assignments, short quizzes, and a final project.

Learning Outcomes

- Be familiar fundamental NLP problems and the methods powering the current state-of-the-art in language technologies, such as large pre-trained neural language models;
- Understand the limitations of current technologies and be able to make informed decisions about addressing them using advanced machine learning techniques;
- Design, implement, and evaluate a computing-based solution to address a language-related problem using state-of-the-art tools.

Prerequisites

CS 580 (Intro to AI) or CS 584 (Data Mining). You should be proficient in (a) Algorithms and Data Structures and (b) Probability and Statistics (STAT 344) or equivalent. Students should be experienced with writing substantial programs in Python. Please contact the instructor if you have questions about the necessary background.

Class Format

The class will be in-person. We will follow a **flipped classroom format**. As the class aims to provide skills necessary to familiarize the students with, and to do cutting-edge NLP research, the classes and assignments will be at least partially implementation-focused. In general, each class will take the following format:

- *Reading:* Before most lectures, you will be pointed to some reading materials *(see "Reading Materials" in course schedule)* that you should read, and to pre-recorded videos that you should watch before coming to class that day.
- *Summary/Elaboration/Questions:* The instructor will summarize the important points of the reading material, elaborate on details that were not included in the reading while fielding any questions. Finally, new material on cutting-edge methods, or a deep look into one salient method will be covered.
- *Demo/Code Walk:* In some classes we will walk through some demonstration code that implements a simple version of the main concepts presented in the reading material.

Grading

There will be no midterm or final exam. Your final grade will be dependent on:
Offline Quizzes (10%): (Some of) the lecture videos on Canvas will be accompanied by relevant quizzes, that will need to be taken after you finish watching the videos.

Three homework assignments (35%): In the first weeks of the semester, we will have three programming assignments (each worth 10% of your grade, to be completed independently) to ensure that everybody gets a minimal hands-on experience with building a state-of-the-art NLP model or system. These assignments will be useful (if not necessary) for implementing your projects later in the class.

- HW1: Language Modeling (10%)
- HW2: Implementing a tool-augmented LLM agent (10%)
- HW3: Reproducibility Study (15%)

Bonus questions and points may be available for all homeworks. All homeworks are to be done individually unless denoted otherwise. You should complete the collaboration questions at the end of each homework, to denote whether you received/provided help from/to any *classmates.*

Take-home Quizzes (20%): To make sure everyone learns the core concepts taught in the class, we will have 3 take-home quizzes (worth 2%,9%,9% of your grade respectively). These will be released on Canvas, and you will have 6 days to complete them (individually). They will have a few questions with an answer that will typically not require more than 4-5 sentences.

Project (35%): The bulk of your grade will be based on a group research project related to the topics we will discuss in class. **The groups will be of 2-3 people. If you intend to complete the project independently or with more than 3 people in your group, you MUST ask for permission from the instructor.**
Please check out this webpage for details and requirements for the project. Briefly, the project will consist of the following milestones:

- **Checkpoint 1: Project Pitch (5%):** Make a team, pick a potential project (we highly recommend choosing one of the available projects), and prepare a slide deck for a 5-minute presentation of your idea.
- **Checkpoint 2: Mid-Report Peer-Review (10%):** Your team will (a) prepare a report of your progress (see instructions), and (b) review the mid-project progress report of other teams, providing constructive feedback.
- **Checkpoint 3: Final Report and Presentation (20%):** You will develop your project, write a report (due the last day of the semester) and prepare a presentation (most likely a poster, to be presented during the last scheduled lecture). (**IMPORTANT: see Project Instructions for requirements on the report and in-class presentation**.)

Letter Grade	Points (out of 100)
A	94-100
A-	90-93
B+	86-89
B	83-85
B-	80-82
C+	76-79
C	73-75
C-	70-72
D	60-69
F	0-59

Late Day Policy:
In case there are unforeseen circumstances that don't let you turn in your deliverables on time, 5 late days *total* over the course deliverables will be allowed without penalty. These apply to all deliverables (assignments and project checkpoints).

- For Assignments: If you are late beyond the allowed late days, your assignment will be graded down one half-grade (3 points) per day late (e.g., Assume your assignment receives a score of 95/100 before penalty. If you have used out your allowed late days and are now 1 more day's late, then your final score becomes 92/100.)
- For Video Quizzes: These can be completed any time before the semester ends. But you will be better off following along through the semester.
- For Take-Home Quizzes: No late days will be allowed. If you are late, you will receive zero grade.
- For Project Checkpoints: Similar to the assignments, any late days beyond the allowed ones will cause penalty (downgraded one half-grade per day late). However, for Checkpoint 3, late days will NOT be feasible given the school deadline for instructors to complete grading. Note that
- **Contact the instructor IMMEDIATELY and BEFORE DUE DAYS if you know that you will very likely to be late for illness, travel, etc. If you have any questions about the late day policy, similarly reach out to the instructor IMMEDIATELY.**

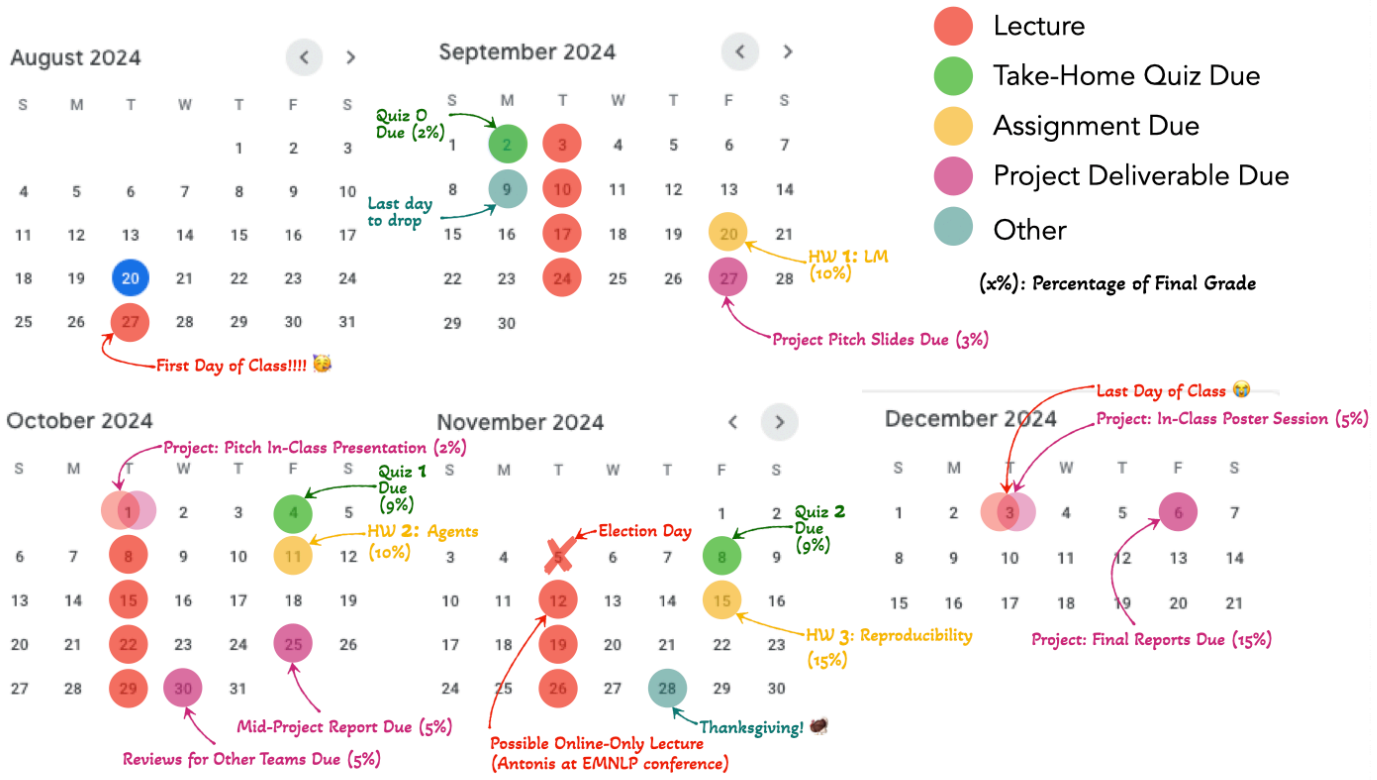
Readings

For each topic/class the instructor will provide a list of papers as suggested readings. Students should be able to understand the course content just by following the lecture and by doing the readings. However, the following textbooks serve as good references.

- Jurafsky and Martin, Speech and Language Processing, 3rd edition [online] **(Referred to as "JM")**;
- Jacob Eisenstein, Natural Language Processing [online] **(Referred to as "Eisenstein")**;
- Yoav Goldberg, Neural Network Methods in Natural Language Processing [publisher] [online primer pdf] **(Referred to as "Goldberg-Publisher/Primer")**; Note that the "publisher" version can be downloaded if you use the school VPN.

Tentative Schedule [TO BE UPDATED]

We will try to cover a lot of ground in the first weeks in order to lay the foundations for the projects, but then we will focus more on specific NLP tasks and Linguistics phenomena. **Pay attention to the Assignments/Quizzes/Project Checkpoints due dates. See the current deadline calendar below**


August 2024
September 2024
October 2024
November 2024
December 2024
Lecture
Take-Home Quiz Due
Assignment Due
Project Deliverable Due
Other
(x%): Percentage of Final Grade
First Day of Class!!!! 🥳
Quiz 0 Due (2%)
Last day to drop
HW 1: LM (10%)
Project Pitch Slides Due (3%)
Project: Pitch In-Class Presentation (2%)
Quiz 1 Due (9%)
HW 2: Agents (10%)
Mid-Project Report Due (5%)
Reviews for Other Teams Due (5%)
Election Day
Quiz 2 Due (9%)
HW 3: Reproducibility (15%)
Possible Online-Only Lecture (Antonis at EMNLP conference)
Thanksgiving! 🦃
Last Day of Class 😭
Project: In-Class Poster Session (5%)
Project: Final Reports Due (15%)

Date	Topic	Video Modules	Deliverables this Week	Suggested Reading Materials
08/27	Introduction and Class Outline; Intro to Neural Networks	0	Take-Home Quiz 0 Released	JM Ch3 (Language Modeling), Ch7 (NNs and FFNN); Eisenstein Ch2; & Goldberg-Primer Ch6.1-6.3 (FFNN); Introduction to Pytorch
09/03	Language Modeling and Basic Measure Theory	1, 2	Quiz 0 due	Cotterell Course Notes SS2.1, 2.2, 2.3, 2.4, 3

09/10	Word Embeddings, Tokenization, and Generation	3		Mikolov et al., 2013a & 2013b;
				Nucleus sampling (Holtzmann et al., ICLR 2020);

09/17	Classical LMs: Finite-state LMs, n-gram LMs	4	Assignment 1 due	An easy-to-read blog post on attention

09/24	Recurrent Neural Networks and Transformers	5, 6	Project Pitch Due, Take-Home Quiz 1 Released	JM Ch9 (RNNs and LSTMs, Encoder-Decoder) JM Ch10.1-10.2 (Transformers), Ch11 (MLMs and pre-training); Goldberg-Publisher Ch10.4;
				Bahdanau et al. 2015 (attention);
				Vaswani et al., 2017 (Transformer);
				Peters et al., 2018 (ELMo);
				Devlin et al., 2019 (BERT)
				An easy-to-read blog post on Transformer language models;
				An easy-to-read blog post on attention

10/01	Modern LLMs: Alignment and Prompting	7	Quiz 1 Due	Exploring the Limits of Transfer Learning... (Raffel et al., JMLR 2020, "T5");
				Pre-train, Prompt, and Predict... (Liu et al., 2021, survey paper);
				Prefix tuning... Prompts for Generation (Li & Liang, ACL 2021);
				Chain-of-Thought Prompting (Wei et al., 2022);
				Chameleon Plug-and-Play (Lu et al., 2023);
				Scaling Laws for Neural Language Models (Kaplan et al., 2020);
				Training Compute-Optimal Large Language Models (Hoffmann et al., 2022);
				PaLM: Scaling Language Modeling with Pathways (Chowdhery et al., 2022)

10/08	LLM Harms, Biases, and Ethics Discussion	8	Assignment 2 Due	Ethics: Zhao et al., 2017;
				Rudinger et al., 2018;
				Gebru et al., 2018

10/15	Machine Translation and Multilingual NLP	9, 10		Eisenstein 18.1-18.2
				Neural Machine Translation... with Subword Units (Sennrich et al., ACL 2016);
				Beyond English-centric multilingual machine translation (Fan et al., 2020);
				MAD-X: Multi-task cross lingual transfer (Pfeiffer et al., EMNLP 2020);
				Hershcovich et al., 2022;
				Liu et al., 2021;
				Bird, 2020;
				Lent et al., 2021

10/22	Morphology, Syntax, and Parsing	11	Mid-Project Report Due	JM Ch8; JM Ch12.1-12.2, 12.6, 13.1-13.4 (constituency), 14 (dependency);
				Chen&Manning, 2014;
				Dozat&Manning, 2017

10/29	NLP Interpretability and Explainability	12	Report Reviews Due, Take-Home Quiz 2 Released	Explainability: LIME; e-SNLI; Hewitt&Liang'19; CheckList
11/05	NO CLASS (election day)		Take-Home Quiz 2 Due	

11/12	Special Topics: Retrieval-Augmented LMs, Tool Augmentation, Agents	13, 14	Assignment 3 Due	Interactivity: Wang et al., 2016;
				Hancock et al., 2019;
				guidelines for human-AI interaction;
				InstructGPT&RLHF

11/19	Special Topics: Multimodality (speech, vision)	15, 16		TBA

11/26	Special Topics: Security, Prompt Injections, Data Poisoning	17		TBD
12/03	Project Presentations		Final Project Presentation and Report due	

Honor Code

George Mason has established institutional academic stnadards. Three fundamental principles to follow at all times are that: (1) all work submitted be your own, as defined by the assignment; (2) when you use the work, the words, or the ideas of others, including fellow students or online sites, you give full credit through accurate citations; and (3) if you are uncertain about the ground rules on a particular assignment or exam, ask for clarification. No grade is important enough to justify academic misconduct.

```
The class enforces the <a href="https://oai.gmu.edu/full-honor-code-document/">GMU Honor Code</a>, and the <a href="https://cs.gmu.edu/resources/h
```

- **IMPORTANT NOTE about the use of AI Assistants:** Use of Generative-AI tools should follow the fundamental principles of the Honor Code. All of the work you submit **HAS TO BE YOUR OWN.** We will not explicitly monitor the use of ChatGPT or other LLM-based assistant, but we may run your outputs through a LLM-generated content detector. Note that these systems are prone to mistakes and hallucinations. If you do use generative AI software, you will be responsible for any incorrect, biased, or unethical information that is submitted. For example, if you use ChatGPT for a report and it hallucinates a citation, or it produces a verbatim repetition from a paper without a proper citation, **THIS IS A VIOLATION OF ACADEMIC INTEGRITY** and you will be referred to the appropriate office, as per the Honor Code requirement.

Note to Students

Take care of yourself! As a student, you may experience a range of challenges that can interfere with learning, such as strained relationships, increased anxiety, substance use, global pandemics, feeling down, difficulty concentrating and/or lack of motivation. All of us benefit from support during times of struggle. There are many helpful resources available on campus and an important part of having a healthy life is learning how to ask for help. Asking for support sooner rather than later is almost always helpful. GMU services are available, and treatment does work. You can learn more about confidential mental health services available on campus at: https://caps.gmu.edu/. Support is always available (24/7) from Counseling and Psychological Services: 703-527-4077.

Disabilities

If you have a documented learning disability or other condition which may affect academic performance, make sure this documentation is on file with the Office of Disability Services and come talk to me about accommodations. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Services, I encourage you to contact them at ods@gmu.edu.

NEXT
CS 678 Course Project

Last updated on Jan 1, 0001

Published with Wowchemy Website Builder