Syllabus

CS 795

NLP Theory 🔗

Instructors

[Antonis Anastasopoulos](#) (antonis [at] gmu [dot] edu)
Office Hours: Email for appointments.

TA

TBD (TBD [at] gmu [dot] edu)
Office Hours: TBD

Meets

Wednesdays, 4:30 pm - 7:10 pm, Enterprise Hall 173.

Course Web Page

https://nlp.cs.gmu.edu/course/cs795-fall25/.
We will use **Canvas** and **Gradescope** for course materials/assignments/grading, and **Piazza** for Q&A (sign up link:
https://piazza.com/gmu/fall2025/cs795).

Course Description

This PhD-level seminar course will cover seminal topics around the theory of language, learning, and the emerging theory around large language models. The class will require semester-long research projects based on the materials discussed in class.

Prerequisites

No official prerequisites, but highly recommended are CS 678 (Advanced NLP) and CS 747 (Deep Learning). Please contact the instructor if you have questions about the necessary background.

Class Format

The class will be in-person. In general, each class will take the following format:
- *Reading:* Before some lectures, you will be pointed to some reading materials *(see "Reading Materials" in course schedule)* that you should read before coming to class that day.
- *Presentations:* A group of students will be responsible for leading the discussion on each paper.

Grading

There will be no midterm or final exam. Your final grade will be dependent on:
Class participation (50%): You are expected to provide construct feedback to the assigned groups.

Project (50%): Ths will be a group research project related to LLMs and the topics we will discuss in class. **The groups will be of 2-3 people.** More details about the project will be shared in class.

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

Contact the instructor in advance if you know that you will very likely not be able to attend a class due to illness or other incidents.

Tentative Schedule

Our discussions will be framed around the following modules:

- 1. LLMs and Language Understanding
- 2. Language Learning in the Limit
- 3. Learning Under Privileged Information
- 4. Expressivity of Transformers

Date	Module	Details	Reading Materials
08/27	1	Introductions and Class Outline; meaning and language understanding	**Required** • [1] Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. Bender and Koller, 2020 • [2] Eight Things to Know about Large Language Models, Bowman, 2024. **Optional:** • Is it possible for language models to achieve language understanding?, Potts, 2020.
09/03	1	The Nature of LLMs	**Required** • [1] Dissociating language and thought in large language models. Mahowald et al., 2024 **Optional:** • [1] Do Prompt-Based Models Really Understand the Meaning of Their Prompts? Webson and Pavlick, 2022 • Do As I Can, Not As I Say: Grounding Language in Robotic Affordances, Ichter et al., 2023.
09/10	2	Language Learning in the Limit	**Required** • [1] Language Identification in the Limit • [2] Inductive inference of formal languages from positive data, Angluin, 1980. **Optional:** • [1] The Logical Problem of Language Acquisition, Cowie, 1997. • [2] Some decidability results on grammatical inference and complexity, Feldman, 1972.
09/17	2	Language Generation in the Limit	**Guest Lecture by Evgenios Kornaropoulos** **Strongly Recommended:** • [1] Language Generation in the Limit, Kleinberg and Mullainathan, 2024. **Optional:** • [1] Characterizations of Language Generation With Breadth, Kalavasis et al., 2025. • [2] Representative Language Generation, Peale et al., 2025. • [3] Exploring Facets of Language Generation in the Limit, Charikar and Pabbaraju, 2025. • [4] Generation from Noisy Examples, Raman and Raman, 2025.
09/24	1,2	An aside: Hallucinations and Biases	**Required** • [1] Calibrated Language Models Must Hallucinate, Kalai end Vempala, 2023. • [2] Large Language Models Are Biased Because They Are Large Language Models, Resnik, 2025. **Optional:** • [1] (Im)possibility of Automated Hallucination Detection in Large Language Models, Karbasi et al., 2025. • [2] LLMS Know More than they Show: on the Intrinsic Representation of LLM Hallucinations, Orgad et al., 2025. • [3] Foundational Challenges in Assuring Alignment and Safety of Large Language Models, Anwar et al., 2024. (sections 2 and 3) • [4] HALOGEN : Fantastic LLM Hallucinations and Where to Find Them, Ravichander et al., 2025.
10/01	1	The Nature of LLMs (Part II)	**Required** • [1] Embers of Autoregression: Understanding Large Language Models Through the Problem They are Trained to Solve, McCoy et al., 2023. **Optional:** • [1] Reasoning or Reciting? Exploring the Capabilities and Limitations of Language Models Through Counterfactual Tasks, Wu et al., 2024. • [2] Large Language Models Fail on Trivial Alterations to Theory-of-Mind Tasks. Ullman, 2023. • [3] When a Sentence Does Not Introduce a Discourse Entity, Transformer-Based Models Still Sometimes Refer to It. Schuster and Linzen, 2022.
10/08	3	Learning Under Privileged Information Part 1	**Required** • [1] A new learning paradigm: Learning using privileged information, Vapnik and Vashisht, 2009. **Optional:** • [1] On the Theory of Learning with Privileged Information, Pechyony and Vanpik, 2010.
10/15	3	Learning Under Privileged Information Part 2	**Required** • [1] On the Power of Context-Enhanced Learning in LLMs, Zhu et al., 2025. • [2] Transformers as Algorithms: Generalization and Stability in In-context Learning, Li et al., 2023. **Optional:** • [1] Deep Learning under Privileged Information Using Heteroscedastic Dropout, Lambert et al., 2019. • [2] Metadata Conditioning Accelerates Language Model Pre-training, Gao et al., 2025.
10/22	4	Representation Power of Transformers Part 1	Antonis Presenting on Preliminaries (formal languages, logic, circuits) **Optional (in preparation of the whole module):** • [1] What Formal Languages Can Transformers Express? A Survey, Strobl et al., 2024. • [2] Overcoming a Theoretical Limitation of Self-Attention, Chiang and Cholak, 2022.

10/29	4	Representation Power of Transformers Part 2	**Required** • [1] Theoretical limitations of self-attention in neural sequence models, Hahn, 2020. • [2] Masked hard-attention transformers and Boolean RASP recognize exactly the star-free languages, Yang et al., 2023. **Optional:** • [1] Formal language recognition by hard attention Transformers: Perspectives from circuit complexity, Hao et al., 2022. • [2] Characterizing the Expressivity of Transformer Language Models, Li and Cotterell, 2025. • Why are Sensitive Functions Hard for Transformers? Hahn and Rofin, 2024.
11/05	4	Representation Power of Transformers Part 3	**Required** • [1] Attention is Turing Complete, Pérez et al., 2021. • [2] Transformers as transducers, Strobl et al., 2024. **Optional:** • [1] Logical languages accepted by transformer encoders with hard attention, Barceló et al., 2024.
11/12		Guest Lecture - TBD	
11/19	4	Transformers and Impossible Languages	**Required** • [1] Mission: Impossible Language Models, Kallini et al., 2024. • [2] Can Language Models Learn Typologically Implausible Languages?, Xu et al., 2025. **Optional:** • [1] Lexinvariant Language Models, Huang et al., 2023. • [2] Priorless Recurrent Networks Learn Curiously, Mitchell and Bowers, 2020.
11/26		NO CLASS (Thanksgiving)	
12/03		Final Project Presentations	

Honor Code

The class enforces the GMU Honor Code, and the more specific honor code policy special to the Department of Computer Science. You will be expected to adhere to this code and policy.

- **IMPORTANT NOTE about the use of ChatGPT and similar technologies:** Unless permission to do so is granted by the instructor, you (or your group, if a group assignment) may not use assistive technology, artificial intelligence, or other tools to complete assignments which can generate, translate, or otherwise create/correct code or answers (many types of assistive technology may be permitted, but you must ask permission).

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.

Last updated on Jan 1, 0001

Published with Wowchemy Website Builder