CS-397: SEMINAR IN LARGE LANGUAGE MODELS

Spring 2025

Instructor: David Demeter	**Email:** David.Jr1@Northwestern.edu
Time: Wednesdays 3:00PM to 5:50PM	**Place:** Technological Institute, Room L160

Course Description: Large Language Models (LLMs) assign probabilities to sequences of words, and are used as a general-purpose mechanism to perform text classification, text generation, question answering, machine translation, and many other tasks. In recent years, deep neural networks have provided radically improved LLMs. This course will cover both the fundamental technologies that comprise large language models, along with applications of the techniques to important tasks in artificial intelligence. Students will be required to read and present research papers, and to write a four-page semi-novel research paper.

Office Hours: By appointment via Zoom (https://northwestern.zoom.us/j/93429505268).

Course Materials: No textbook. The course material will be comprised of research papers in the field.

Course Goals: The goal of this course is to familiarize graduate students and advanced undergraduates with the current state-of-the-art in large language models and natural language process. Students will read recently published papers in the field.

Prerequisites: CS-349 or permission of the instructor.

Grading Policy: Grades are assigned using the standard scale (given in the "introduction" lecture notes), so 93-100 points is an A, 90-93 points is an A-, etc. Score will not be rounded. Points will be allocated as follows:

	Points
Class Participation	35.0 pts
Paper Presentations (2 for 20 points each)	40.0 pts
Research Paper	25.0 pts

All students are required to join a group of two, to form exactly eleven groups for the class, no later than the end-of-day on Wednesday, April 2nd. After that, I will make random assignments for any students not in a group. Groups will remain fixed for the remainder of the quarter. Each group is responsible for leading two Paper Presentations, ideally separated by a few weeks. I will ask two separate groups to coordinate to presented different parts of "A Survey of Large Language Models". Students will have the opportunity to *bid* on individual papers, with final paper assignments to the group being made by the instructor. Bids are due no later than the end-of-day on Friday, April 4th. Papers will be assigned on Saturday, April 5th, with bids being taken into consideration. As a seminar class, student class participation is vital. Unless other arrangements are requested and approved, students are expected to attend class in-person. Class participation in a hybrid in-person/remote setting does not work.

Class participation will count for 5.0 points per week starting with Week #2. **You must submit a short two-sentence Self-Assessment of your participation before the end of each class while your recall is still fresh.** Delivering a paper presentation does not count towards class participation. I cannot accept late submissions. The two weeks with the lowest class participation scores will be dropped for each student, with the exception of Weeks #9 and #10 when groups present their own papers.

Paper Presentations:

- Task (1.0 pts): Describe the task being performed in the paper, its motivation (e.g., what problem is being solved, along with an example) and a description of existing methods.
- Model (2.5 pts): Define the model or methodology used in the paper (e.g., equations and/or diagrams), define important terms/concepts and clearly state model inputs and outputs.
- Dataset (1.0 pts): Describe the dataset, its source and/or construction and provide relevant statistics.
- Experiments (2.5 pts): Describe the formulation of the experimental setup, and evaluation metric(s) used to measure performance. If this is a well-established dataset, briefly mention current state-of-the-art performance.
- Results (0.5 pts): Provide a headline of the important result(s) -- discussing every data point provided should be avoided. Describe the method used to construct the baseline, and the extent to which the model presented in the paper "beat" the baseline.
- Analysis (2.5 pts): Discuss the limitations of the approach, the potential driver of results and any other important insights.
- Discussion (10.0 pts): Guide a discussion of the paper by asking questions, by asking the class to respond to insights presented, and by answering questions from the class (and instructor).

Not every paper will have all of the elements, but most paper will have a majority of these elements. In some instances the group will need to explore (and present) related work. A PowerPoint (or similar) presentation must be prepared for each Paper Presentation with the content scoped for a 25-minute presentation, followed by a 25-minute discussion. For the avoidance of doubt, Paper Presentations are group assignments and Self-Assessments are graded on an individual basis.

Groups will also be required to write a four-page research paper as a final project. groups will present their papers to the class during Weeks #9 and #10. The topic must be approved by the instructor -- more details to follow.

Course Objectives:

- Have a general understanding of the current state-of-the-art Large Language Models and associated Natural Language Processing tasks,
- Understand how to used at least one Large Language Model (via the course project), and
- Be able to understand, and think critically about, recent research papers in the field.

Calendar:

Week #1: Introduction

- Introductory remarks
- Review of selected background material

Week #2: Large Language Models

1. L Ouyang et al. "Training language models to follow instructions with human feedback." (2022) (abs/2203.02155)
2. WX Zhao et al. "A Survey of Large Language Models." (2023) (abs/2303.18223)

Week #3: Prompting

3. J Wei et al. "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." (2022) (abs/2201.11903)
4. P Sahoo et al. "A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications." (2024) (abs/2402.07927)
5. D Zhou et al. "Least-to-Most Prompting Enables Complex Reasoning in Large Language Models." (2022) (abs/2205.10625)

Week #4: Dense Representations

6. T Zhang et al. "BERTScore: Evaluating Text Generation with BERT." (2019) (abs/1904.09675)
7. N Reimers et al. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." (2019) (abs/1908.10084)
8. P Lewis et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." (2020) (abs/2005.11401)

Week #5: Efficiencies

9. N Houlsby et al. "Parameter-Efficient Transfer Learning for NLP." (2019) (abs/1902.00751)
10. EJ Hu et al. "LoRA: Low-Rank Adaptation of Large Language Models." (2022) (abs/2106.09685)
11. V Sanh et al. "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter." (2019) (abs/1910.01108)

Week #6: Efficiencies (*Cont.*)

12. T Dettmers et al. "QLoRA: Efficient Finetuning of Quantized LLMs." (2023) (abs/2305.14314)
13. O Press et al. "Train Short, Test Long: Attention with Linear Biases Enables Input Length Extrapolation." (2021) (abs/2108.12409)
14. NF Liu et al. "Lost in the Middle: How Language Models Use Long Contexts." (2023) (abs/2307.03172)

Week #7: External Tools and Knowledge

15. T Schick et al. "Toolformer: Language Models Can Teach Themselves to Use Tools." (2023) (abs/2302.04761)
16. A Bosselut et al. "COMET: Commonsense Transformers for Automatic Knowledge Graph Construction." (2019) (abs/1906.05317)
17. T Mihaylov et al. "Can a Suit of Armor Conduct Electricity? A New Dataset for Open Book Question Answering." (2018) (abs/1809.02789)

Calendar (*Cont.*):

Week #8: Logic and Reasoning
18. X Lu et al. "NeuroLogic Decoding: (Un)supervised Neural Text Generation with Predicate Logic Constraints." (2020) (abs/2010.12884) 19. A Lewkowycz et al. "Solving Quantitative Reasoning Problems with Language Models." (2022) (abs/2206.14858) 20. C Bhagavatula et al. "Abductive Commonsense Reasoning." (2019) (abs/1908.05739) 21. WL Chiang et al. "Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference." (2024) (abs/2403.04132)
Week #9: Student Presentations
Six student groups.
Week #10: Student Presentations
Five student groups.

Northwestern University Syllabus Standards

This course follows the Northwestern University Syllabus Standards (https://www.registrar.northwestern.edu/registration-graduation/northwestern-university-syllabus-standards.html). Students are responsible for familiarizing themselves with this information.