Syllabus of LLMs

Advanced Topics – Large Language Models

ECE 16:332:539:02 Spring 2025
(Electrical and Computer Engineering, Rutgers University at New Brunswick, NJ)

Instructor: Professor Zhou
Name: Xing Zhou
Email: Xing.Zhou@rutgers.edu
Office: EE – 113

Class Meets:
Time: Thursday @ 2:00 pm ~5:00 pm
Place: COR – 538

I. Course Information

Semester: Spring 2025 (January 21 – May 8, 2025)
Credit Hours: 3
Final Exam: Final Exam + Main Project Presentations
Office Hours: Thursday 5:10 pm~ 6:10 pm

II. Course Overview

Large Language Models (LLMs) have revolutionized Natural Language Processing (NLP) in recent years, establishing the foundation for state-of-the-art systems. This course will explore cutting-edge research on pre-trained language models like GPT, BERT, and their derivatives, which are built on transformer-based architectures. These models, capable of processing and generating human-like text, are crucial for tasks such as translation, question-answering, and summarization. Students will dive into **technical foundations** (such as BERT, GPT, T5, mixture-of-expert models, and retrieval-based models.), **emerging capabilities** (knowledge representation, reasoning, few-shot learning, and in-context learning), **advanced techniques** (like attention mechanisms, fine-tuning and adaptation paradigms and the ethics of LLMs). Hands-on experience with tools like PyTorch and TensorFlow will allow students to build, fine-tune, and apply LLMs, creating applications like sentiment analysis and chatbots. Students are also expected to read and present research, and complete one final project.

Prerequisites: Machine Learning or its equivalent(s), Practical Python or Python for ML

III. Learning goals

1.This course is intended to prepare you for performing cutting-edge research in natural language processing, especially topics related to pre-trained language models. We will discuss the state-of-the-art, their capabilities and limitations.
2. Practice your research skills, including reading research papers, conducting literature survey, oral presentations, as well as providing constructive feedback.
3. Gain hands-on experience through the final project, from brainstorming ideas to implementation and empirical evaluation and writing the final paper.

IV. Recommended Learning Materials

1. Transformers for Natural Language Processing and Computer Vision: Explore Generative AI and Large Language Models with Hugging Face, ChatGPT, GPT-4V, and DALL-E 3 3rd Edition by Denis rothman. Packt Publishing. ISBN -10: 1805128728 / ISBN-13: 978 – 1805128724
2. Natural Language Processing with Transformers by Lewis Tunstall, Leandro von Werra, and Thomas Wolf. O'REILLY. ISBN -10: 1098136799 / ISBN-13: 978 – 1837633784
3. Mastering Transformers: The Journey from BERT to Large Language Models and Stable Diffusion by Savas Yildinm and Meysam Asgari-Chenaghlu. ISBN -10: 1837633789/ ISBN-13: 978 – 1098136796
4. Hands-On Large Language Models: Language Understanding and Generation 1st Edition. By Jay Alammar and Maarten Grootendorst,O'REILLY. ISBN -10: 1098150961/ ISBN-13: 978 – 1098150969
5. Build a Large Language Model (From Scratch) by Sebastian Raschka,Manning. ISBN -10: 1633437167/ ISBN-13: 978–1633437166
6. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent System, 3rd Edition, by Aurélien Géron. O'REILLY. ISBN-10:1098125975/ISBN-13: 978-1098125974
7. Lecture notes will be posted in Canvas

V. Evaluation or Grading (Tentative)

1. Attendance: 10%
2. In-Class Participation: 10%
3. Two Individual Presentations (Topic: Paper Reading or Projects) : 25%
4. Two Individual Projects : 25%
5. Final Exam: 30%

The grading scale is as follows:

A = [90, 100]	B$^+$ = [85, 89]	C$^+$ = [75, 79]	D$^+$ = [65, 69]	F = [0, 59]
	B = [80, 84]	C = [70, 74]	D = [60, 64]	

VI. Course Delivery

1. This course consists of three parts: teaching, discussions and student presentations. Students are expected to attend the class and participate in the discussion.
2. Since this is an advanced course, we will learn this course with questions, and I will post several questions before each lecture.
3. We will be teaching and discussing the foundational technologies and methodologies as well as state-of- the-art some papers and topics in class. I will post all recommended papers and materials before each lecture.
4. All the course materials and communications would be posted on Canvas.

VII. Tips for Success

1. Regular attendance and active participation in class are crucial for success in this course.
2. Ask questions, answer questions, and don't hesitate to make educated guesses.
3. If you don't understand a particular concept, feel free to ask me in class or visit during office hours.
4. LLMs involve many emerging technologies; reading extensively from references can significantly broaden your knowledge.
5. Consistent hands-on practice is essential to digest, consolidate, and master theoretical knowledge and effectively solve problems.

VIII. Course Outline (Tentative)

Part 1 Theory

1.An Introduction to Large Language Models

2. Tokens and Embeddings

3. Looking Inside Large Language Models

Part 2 Applications

4.Text Classification

5.Text Clustering and Topic Modeling

6.Prompt Engineering

7. Advanced Text Generation Techniques and Tools

8.Semantic Search and Retrieval-Augmented Generation

9.Multimodal Large Language Models

Part 3 Training and Fine-Tuning

10.Creating Text Embedding Models

11.Fine -Tuning Representation Models for Classification

12.Fine-Tuning Generation Models
13.Error Generic Data Poisoning Defense
14.Reverse-Engineering Attacks (REAs) on Classifiers

IX. Topics Covered Week by Week (Tentative)

W#	Date	Lecture Topics	Remark
Week 1	01/23	#01: Course Organization +An Introduction to Large Language Models	3 H for lecture
Week 2	01/30	#02: Tokens and Embeddings +Class Discussions on related topics	
Week 3	02/06	#03: Looking Inside Large Language Models + Class Discussions on related topics	
Week 4	02/13	#04: Text Classification +Text Clustering and Topic Modeling +Class Discussions on related topics	
Week5	02/20	#05: Prompt Engineering +Student Presentation ()	2 H for lecture 1 H for InPre
Week 6	02/27	#06: Advanced Text Generation Techniques and Tools +Student Presentation ()	2 H for lecture 1 H for InPre
Week 7	03/06	#07: Semantic Search and Retrieval-Augmented Generation +Student Presentation ()	2 H for lecture 1 H for InPre
Week 8	03/13	#08: Multimodal Large Language Models +Student Presentation ()	2 H for lecture 1 H for InPre
Week 9	03/20	Spring Recess(no class)	3/15 -3/23
Week 10	03/27	#09: Creating Text Embedding Models +Student Presentation ()	2 H for lecture 1 H for InPre
Week 11	04/03	#10: Fine-Tuning Representation Models for Classification +Student Presentation ()	2 H for lecture 1 H for InPre
Week 12	04/10	#11: Fine-Tuning Generation Models +Student Presentation ()	2 H for lecture 1 H for InPre
Week 13	04/17	#12:Error Generic Data Poisoning Defense +Student Project Presentation ()	2 H for lecture 1 H for InPre
Week 14	04/24	#13: Reverse-Engineering Attacks (REAs) on Classifiers +Class Discussions on related topics	
Week 15	05/01	#14: Student Group Project Presentation	
Week 16	05/08	#15: Final Exam (5/08 -5/14) Student Group Project Presentation	1.5 H for Exam 1.5 H for Team-Pre
Week 17	05/15	Summer Vacation	