INFO 7374 Syllabus

Course Information

Course Title: Advanced Techniques with Large Language Models

Course Number: INFO 7374

No class during finals week

Grading

There will be no quizzes and no exams. Grades will be based on assignments. There are assignments.

Late Days

- 6 late days (max 3 late days per assignment)

Academic Honesty Policy

- ChatGPT as a learning tool is ok, but using it to do your homework is not

Course Objectives

- Deep learning fundamentals
- Understand how LLMs work theoretically and how they are built
- How to work with open source models and tooling
- Adapt LLMs to your custom use cases
- How to use LLM tools for information retrieval and fact grounded generation
- Advanced prompting techniques
- How to create agents to do useful tasks
- How to create high quality datasets
- Understan what problems LLMs are well suited for and where there limitations are

Schedule

Week	Topic	Recommended Reading
Week 1	Course introduction, deep learning fundamentals, backprop, MLPs, gradient descent	Yes you should understand backprop Introduction to Deep Learning Neural Networks and Deep Learning
Week 2	Language modeling, tokenizers, and transformers (GPTs)	GPT-2 Attention is All You Need The Illustrated Transformer
Week 3	Pretraining, Data Cleaning, Data mixes	LLaMA: Open and Efficient Foundation Language Models DoReMi: Optimizing Data Mixtures Speeds Up Language Model Pretraining Multi Query Attention Grouped Query Attention Open Pre-trained Transformer Language Models Chinchilla Scaling Laws RefinedWeb Dataset
Week 4	Instruction Tuning, Instruction Data Collection, Supervised Finetuning	Training language models to follow instructions with human feedback Alpaca: A Strong, Replicable Instruction-Following Model

Week 5	Parameter Efficient Finetuning	Understanding Parameter-Efficient Finetuning of Large Language Models: From Prefix Tuning to LLaMA-Adapters
Week 6	Prompting, Hallucinations, Prompt injections, and LlamaGuard	Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4 LlamaGuard
Week 7	RLHF, RLAIF, DPO	Illustrating Reinforcement Learning from Human Feedback (RLHF) RLHF: Reinforcement Learning from Human Feedback Spinning Up - PPO https://huyenchip.com/2024/02/28/predictive-human-preference.html
Week 8	Embeddings and Basic Retrieval Augmented Generation	Retrieval-Augmented Generation for Large Language Models: A Survey
Week 9	Advanced Retrieval Augmented Generation	ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT
Week 10	Verifiers + LLM Programs	
Week 11	Agents	LLM Powered Autonomous Agents
Week 12	Multimodal Models	CLIP Flamingo Llava
Week 13	Time Series Forecasting,	Are Transformers Effective for Time Series Forecasting?

	Tabular Data, Recommendation Systems	