CS 395T: Foundations of Modern Generative AI

UT Austin

Prof. Noah Golowich

Course Information

- Monday / Wednesday, 2-3:30 p.m.
- Instructor office hours: Wednesday, 3:30-4:30pm, GDC 4.808 starting September 2.
- Course location: for now, WAG 308 (may change!)
- **Note**: instructor is traveling on Wednesday August 26, so Zoom lecture (see website for zoom link) — sorry!
- **TA**: Joey Zhou. Office hours: Friday, 10-11am GDC 4.416 (or Zoom). First few weeks, Zoom-only.
- **Website**: https://noahgol.github.io/teaching/cs395t-f26/

Course Philosophy

- Last few years: remarkable progress in AI.
- How/Why does AI work? What conceptual frameworks can help us to understand it / make informed predictions?
  - Ideas from CS theory give us some leads…
  - But ultimately don't really give us a satisfying answer in many cases
- **Goal of this course**: teach both theoretical & empirical results, so as to understand (a) success stories from theory helping understand AI, (b) understand where it falls short and how we might hope to proceed.
- **Hope**: you are inspired to work on some of those questions!

Structure of Lectures

- ~13 lectures: given by instructor (mostly on blackboard).
- ~11 lectures: student presentations on more recent papers (typically last 1-3 years) on research topics
- ~4 lectures: project presentations (more on that later)
- Typical lecture:
  - **Empirical component**: what experiments were run, results.
  - **Theoretical component**: what theoretical models have been introduced corresponding to the empirical component? Is it satisfying? Is it useful?

Prerequisites

- No formal prerequisites
- Informal prerequisite: “mathematical maturity”.
- Don’t need experience with modern deep learning frameworks (pytorch/jax/etc). But I encourage you to “get your hands dirty” and tinker with things! Codex/Claude Code make this really easy.

Assignments

- Paper presentation: ~11 lectures on course website are designated as “student presentations”. You sign up for 1 slot (roughly 3 students per lecture) and present one (or a few) papers on a topic for that lecture. Will require some coordination between students in the same lecture. Signup sheet will be distributed in a couple weeks.

- Final project: broadly scoped, generally speaking anything relating to course content is fair game, as long as it can’t be 1-shotted by a coding agent.

  - **Theoretical project example**: come up with some theoretical model helping to explain the results of some empirical paper.

  - **Empirical project example**: does some algorithm with provable guarantees suggest an improvement to some procedure used to train LLMs?

  - Deliverables: project proposal, in-class presentation, final report.

  - Small teams (2, maybe 3 students) allowed.

Brief Overview of Topics

- Lecture 1-2: review / history of deep learning
- Lecture 3-5: transformer architecture, pre-training/scaling laws
- Lecture 6-8: optimization (Adam, Muon/preconditioning)
- Lecture 9-10: LLM inference (chain-of-thought, spec dec, flash attention)
- Lecture 11-15: post-training: DPO, RL, continual learning, long-context
- Lecture 16-20: diffusion models
- Lecture 21-23: AI safety (watermarking, backdoors, multi-agent, monitoring)
- Lecture 24-28: project presentations

Note: above is ambitious, scope may be decreased if we move more slowly…

Week 1

- Today’s lecture (**theory component**): basics of MLPs, perceptron algorithm, back-propagation (on blackboard).
- Wednesday’s lecture (**empirical component**): finishing up above + summary of pre-2017 deep learning (CNNs, RNNs/LSTMs, GANs). **Over zoom (see website for link).**
- **Goal**: just to give a flavor for deep learning before the last decade. Many entire courses could be taught on this!