Week 1, Day 2: Course Overview

Press Space for next slide

Agenda

- More details about our modules
  - Overview of planned topics
  - Google doc activity
- Go over "loose prerequisites" and "crash course"
  - Basic ML and data science in Python
  - Relevant libraries and other tooling discussion
- Finding research papers (and their actual contents) using Google Scholar, Semantic Scholar, and more
  - Managing references, notes, tasks (group discussion!)

Module 1: Intro

Key goals:

- Gain exposure to different notions of what human-centric and/or data-centric AI is.
- Learn about different frameworks that have been put forward aimed at researchers and/or designers of AI products

Key question:

- In what situations would adopting a "human-centered" or "data-centric" approach to AI cause me to change my AI product?

Some of the Frameworks we'll see

- Chancellor's Human-Centered ML
    - when to use ML, problem statements take positions, domain expertise, preparing to fail…
- DataPerf team and data-centric AI
    - focus on changing data instead of changing model
- Schneiderman's HCAI
    - balancing human control and automation; "Reliable, Safe & Trustworthy"
- Zhu's Value Sensitive Algorithm Design
    - incorporate tacit knowledge and feedback when you're designing your algorithm
- FairML
    - how do we ensure ML predictions meet some standard of fairness
- This is just a sample itself. There's more! It's an active conversation.

Important: I want to give you exposure to lots of (fresh) ideas in the space before I insert my own ideas!

Module 2: Data valuation and scaling

Data valuation also has many definitions. We'll read about some of them

- But we'll mostly focus on a "technical ML approach": data influence
- What is the "effect" of observation i on some model A
- Data influence seeks to explain all the different counterfactual worlds – alternate universes in which we had a slightly different dataset

Human behavior can access those counterfactuals.

- Human actions in a data market will control what data ends up in training.
- What happens if people change their behavior?
- What if people protest , or engage in collective action by training data creators (e.g., you!)

Data Valuation Key Reading

We'll lean heavily on

Hammoudeh, Z. and Lowd, D., 2024. Training data influence analysis and estimation: A survey. Machine Learning, 113(5), pp.2351-2403.

Data Scaling

Data scaling is about predicting how much performance (i.e., reduced test loss) we'll get out collecting from training data, typically from the same distribution we've already been collecting from

Both are about data counterfactuals.

- Influence is typically more about "what if we lost this one observation, or this small group" and scaling is typically more about "what if we got access to a whole bunch of new data"
- We're starting with this because it will be a core concept for other discussions!
- "Realizing" data counterfactuals is a core way for humans to "act on" AI systems

Data Scaling Key Readings

We'll lean heavily on

Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., Patwary, M.M.A., Yang, Y. and Zhou, Y., 2017. Deep learning scaling is predictable, empirically. arXiv preprint arXiv:1712.00409.

for a foundational work on data scaling

and

Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L. and Hobbhahn, M., 2024. Will we run out of data? Limits of LLM scaling based on human-generated data. arXiv preprint arXiv:2211.04325, pp.13-29.

for a look at these trends in very recent context.

Human-centered AI without data influence

i.e., why aren't all human-centered AI classes about data?

- To be clear – there's plenty of human centered AI work that doesn't focus on data
- Convincing model developers to change their practices
- Legislation or licensing that focuses on allowable model behaviors / outputs
- But we'll focus more on the data side.
- A bonus – we're pretty much all data creators so the "data lever" is available to everyone

Some other syllabi if you're curious

- https://philpeople.org/public_cache/file?
  content_type=application%2Fpdf&key=z1d1wj8yk9ztczwt4d9wspqbl6qb

Module 2's Learning goals and Testable material (Broad strokes)

After this module, you can expect to be able to solve some problems such as:

- Explaining how one might calculate training data influence
    - high-level understanding of what different techniques are actually doing
    - high-level understanding of the role of gradient descent – can learn in class, OK if no deep learning background
- Explaining how training data influence might be used
    - applications of training data influence (overflows into next module)
- Describing observed data scaling patterns at a high level
- Intrepret new data scaling results
    - If I give you a plot from a hypothetical paper, interpret it

Module 3: Online platforms and Content Ecosystems

Theme: The impact of Wikipedia and similar platforms on AI ("historical" and present-day)

- Wikipedia is really central to "AI"
  - Wikipedia Content is Used Directly by an Intelligent Technology ("Wikipedia Articles as a Meal to be Served")
  - Wikipedia as Training Data ("Wikipedia as a Recipe")
  - from https://github.com/nickmvincent/UGCValueRoundup/blob/main/wikipedia.md

Reasoning quantitatively about content ecosystems

- Data napkin math

Many other examples

- Anyone heard of Flickr?

- Anyone heard of "Twitter"?

- Then of course private social media

    - Instagram

    - TikTok

    - your new startup?

Module 3' Learning goals and Testable material (Broad strokes)

After this module, you can expect to be able to solve some problems such as:

- Identify particularly prominent data sources in modern AI training
- Answer some historical questions about 21st century AI milestones and their relationship to user-generated content
- Identify some ways to collect and analyze data from UGC platforms

Module 4: Frontiers in Data Governance: Voting, Markets, and More

- Like human-centered AI, "data governance" also means a lot of things and has been used in a lot of ways
- Lots of overlap between talking about data using the lens of governing (let's write laws or vote on things), economics (let's set up a market!) and ecosystem (let's observe how data flows and replenishes). All are useful (and wrong in some way – the "real" model is a complex combination)
- We'll mainly think about protocols and procedures that manage the creation and maintenance of datasets
  - Data governance often includes a component of opt in and opt out… this brings us back to influence / data counterfactuals!
  - Data governance involves governing online platforms… this brings us back to Wikipedia etc.

Voting and Markets

We'll talk about experimental ways to vote with your data, or vote on how data should be used

- opt in, opt out

Data has some unique economic properties

- But it can still be exchanged in a market
- Different market conditions and market participants will lead to… you guessed it… different data counterfactuals

"and more…"

Based on your interests, we can talk about some of the philosophical arguments for different governance approaches "democratizing AI" pluralism and values pluralism etc.

Putting it all together: A multitude of data counterfactuals

Some (such as myself) might say: the data is upstream of everything else we do in AI/ML. Without the data, all of the math and coding can't really produce much of anything.

This is half-true when it comes to research (you can still do lots of math) but very true when it comes to products.

Google Doc exercise

Nick will create a Google doc

- Tools I've used
- Tools I'd like to use
- "Famous" datasets I've worked with (e.g., MovieLens, CIFAR-10, MNIST, …)
- One thing I want to learn for my job…
- One thing I want to learn because I'm curious…

~5 mins filling in (or "plus one-ing" your classmates' entries) ~rest of time for discussion and crash course / pre-req talk