chatGPT and Beyond: LLMs in the classroom

ccl

May 2, 2023

Plan for this tutorial

- ▶ What large language models (LLMs) are
- ▶ How LLMs work, at a high-level
- ▶ Whether LLMs can be used to automate homework
- ▶ How LLMs can be used by instructors and students
- ▶ How to grade in a post-chatGPT world
- ▶ Natural language meta-programming AKA prompt engineering
- ▶ The future of LLMs: what a year from now might look like?

What is an LLM?

We can unpack the meaning in reverse

Model A model is the final product of building a program with machine learning

Language The model takes in human language as input, give back human language as output

Large The model was built from literally billions and trillions of examples

How are they made?

LLMs are trained, in part, like a game of fill-in-the-blank.

Fill in the blank
You get a point for each of these you get right, the goal is to memorize enough to maximize your points

- ▶ "The sky is _____"
- ▶ "To _____ go"
- ▶ "a war-_____ country"

Imagine doing this a trillion times: what could you learn?

What can they do?

On one hand
All an LLM can do is predict the likely text to follow from the given text

On the other. . .
Text prediction at this scale means a system that has memorized *patterns* in language

What the model learns: patterns and stories

- some patterns are simple
- some patterns are subtle and nuanced
- some patterns are counter-intuitive

Examples of patterns

Simple patterns

- ► Cliches
- ► Turns of phrase
- ► Common metaphors

Example of patterns

Complicated patterns

- Genre: cozy mystery, YA dystopia, sword & sworcery fantasy
- Audience: undergrads, pop-sci enthusiasts, fifth graders, philosophy phd students
- Form: cover letter, five paragraph essay, resume, sonnet, article in Nature

Demo time: chatGPT

- chatGPT from OpenAI is the most powerful and most accessible LLM *at the moment*
- We'll talk later about the ethical issues around the company and why I'm looking forward to non-OpenAI products
- https://chat.openai.com

Stories are patterns

Philosophic aside

- We interact with stories of the world, shorthands for objects.
- A chair is hard to define, but you know it when you see it.
- LLMs memorize these stories

LLMs can't do knowledge retrieval

- Putting all this together, do not think of an LLM as doing knowledge retrieval.
- It's reflecting back commonly held beliefs

When is common wisdom good enough?

- When you're intentionally *wanting* common stories
- When you're a subject-matter expert who can discern when something is a good explanation or not
- When you need a starting place for a new subject
- When you're unpacking commonly used jargon

Bias and wrongness

What are the failure modes?

- ▶ Common wisdom can be wrong
- ▶ What the LLM produces might not even be common wisdom
- ▶ Don't trust citations and references
- ▶ Don't trust "facts"
- ▶ Why I don't call it "bias"

Rule of thumb

LLM text should never be "final product"

Your turn

We're going into breakout rooms where folks will have 15 minutes
to try out ideas

Every group should pick a representative to give a summary of
what was talked about afterward

To get you started:

- ▶ Pick a topic from a class you teach and ask for an explanation
- ▶ Ask for a definition of a term
- ▶ Enter an assignment as a prompt
- ▶ Make up a backstory and ask for a cover letter or resume

How'd it go?

Breakout-rooms feel free to report!

My thoughts on LLM as answer generators

- "Write me an essay about..." responses tend to be pretty bad
- Rote questions lead to rote answers
- Packing and unpacking definitions reflects common wisdom
- Answers are useful in very limited domains
- Creativity and novelty require human intervention

Can LLMS automate assignments?

Note: We're going to ignore the question of grading writing *skill* per se for the moment

Can students use this to cheat?

- *sorta*
- Paid essay writing is down approximately 40%
- These were the students already not doing the assignments
- The more creative the work the more immune it is to chatGPT

More interesting: text transformers rather than answer generators

- ▶ The more context you give an LLM the more interesting things it can do
- ▶ You can expand outlines into full drafts
- ▶ You can edit drafts with natural language
- ▶ If students are providing the logic, does it matter if some of the text is generated?

Demo time, again!

Back to the browser!

Your turn!

Another 15 minutes of trying out things we've seen so far

- ▶ Drop in a page from a book and ask questions
- ▶ Analyze an abstract
- ▶ Generate a lesson plan from an outline
- ▶ Edit your own writing interactively
- ▶ . . .
- ▶ DO NOT ENTER IN STUDENT ASSIGNMENTS, FOR PRIVACY REASONS

Grading writing itself

- *This is harder*
- Ask students to practice writing
- Judge on the content: competent but uninspired

Why to never use LLM detectors

An important concept from statistics

- Imagine a tool for detecting cheating with a false positivity rate of 5%.
- $\frac{1}{100}$ students are cheating
- The detector flags a student paper: what are the odds they're innocent?

Prompt engineering: a preview

You can do really *weird* things with LLMs

- ▶ Prompts that generate prompts
- ▶ Getting wildly different results by asking for "chains of reasoning"
- ▶ Giving examples of how to perform a task

Prompt engineering demo

Once more unto the LLM

Prompt engineering references

- https://www.promptingguide.ai/
- https://learnprompting.org/

The problems with OpenAI and chatGPT

- ▶ Massive energy and environmental cost
- ▶ Model was built off of low-wage labor
- ▶ Closed sources: we kinda know how they built it but not really
- ▶ Privacy: what you paste into chatGPT, openAI keeps

The future of LLMs

- ► More efficient models
- ► Open source models
- ► Specialized models
- ► Integration of models with other tools
- ► Using models to query your personal writing, papers, documents

The future of LLMs

We haven't even touched on LLMs as coding-assistants. . .

The future of LLMs

The technology of LLMs is here to stay and it's going to become far more common, accessible, and cheaper to access
We can treat them like they're "cheating machines" or we can look at them as things to let ourselves and students process & analyze information faster

Questions?

Hope this was a helpful introduction and feel free to message me at clarissa.littler@pcc.edu