Model building using an LLM
Prompt Engineering and homework 1

Lars A. Lochstoer

UCLA Anderson School of Management

Why prompt engineering?

LLMs like ChatGPT and Gemini has billions of parameters

- Very general
- It has read tons of (almost all?) textbooks, academic papers, literature, webpages, newspapers, etc.

LLMs use something called "context" and "embeddings"

- It generates tokens iteratively based on your prompt
- This prediction model will predict words based on the context you give it
- In-context learning: It's like your are focusing the model's attention on certain sub-parts of the model

Thus, prompt engineering critical for success in use of LLMs

Examples

Q: "*In one sentence, explain an ols regression to someone who does not understand math well, but instead is more of an emotional learner.*"

A: "*Think of OLS regression as finding a line that best goes through a cloud of points so you can see a trend.*"

Q: "*In one sentence, as a PhD level econometrician, explain OLS regression*"

A: "*OLS regression estimates β by solving $\hat{\beta} = (X'X)^{-1} X'y$, the coefficient vector that minimizes squared residuals and projects y onto the column space of X, yielding the BLUE under standard assumptions.*"

Examples: define the use case

"*Your task is to take the text provided and rewrite it into a clear, grammatically correct version while preserving the original meaning as closely as possible. Correct any spelling mistakes, punctuation errors, verb tense issues, word choice problems, and other grammatical mistakes.*"

Now, you have focused the LLM by providing context that it will actively use when answering subsequent prompts.

The LLM will effectively rely on what it has learned already about grammar, as your context will in fact make the model effectively weigh that part of its parameter space

- Note, you are not changing any of the billions of parameters, of course.
- We are just focusing more on some than others
- More on this later

Few-shot prompts

Focus the model by first providing specific examples of Q's and A's:

> *The following are multiple choice questions about high school computer science.*
>
> *Let x = 1. What is x << 3 in Python 3? (A) 1 (B) 3 (C) 8 (D) 16*
> *Answer: C*
>
> *Which is the largest asymptotically? (A) O(1) (B) O(n) (C) O(n2) (D) O(log(n))*
> *Answer: C*
>
> *What is the output of the statement "a" + "ab" in Python 3? (A) Error (B) aab (C) ab (D) a ab*
> *Answer:*

You can ask it to not answer anything outside of the bounds you are giving it:

- *You are an AI assistant that helps people find information. You can only talk about pets and nothing else. If you don't know the answer, say, \"Sorry bud, I don't know that.\" And if you cannot answer it, say \"Sorry mate, can't answer that - I am not allowed to\".*

Prompt skeleton

LLMs estimate $p(x_t \mid x_{<t}, \text{instructions})$.

- Finite context. Place constraints early; put the ask last.

Prompt outline:

- **ROLE**: who the model is.

- **TASK**: single objective.

- **CONSTRAINTS**: time, tokens, style, safety.

- **FORMAT**: JSON/Markdown schema.

- **EXAMPLES**: k few-shot pairs.

- **RESOURCES**: tools or context ids.

- **ASK**: one specific output request.

From vague to testable

Bad: "Summarize this paper."

Better:

- Task: summarize for a technical reader.

- Constraints: ≤ 150 words; 3 key claims; 1 limitation.

- Format: Markdown with headings {Claims, Limitation}.

- Ask: end with 1 replication question.

HW1 and model building

In homework 1, we will use an LLM to help build a state space model for U.S. inflation data.

I suggest prompts that iteratively gets us to write out the math, do the coding, and estimate the model.

This is an example of how you increase your value!

- It would take much longer if you would have to do this without the help of the LLM
- At the same time, it's not like the LLM would just get to the best answer without your expert guidance.
  - There are lots of modeling choices to be made
  - Taste, intuition, and experience helps guide the model
  - Us, humans, have a good sense of what is "in the air" currently that people want to focus on