

## ASSIGNMENT 1#: Prompt Engineering with LLMs

Prompt engineering is the process of designing and refining inputs (prompts) to elicit the desired outputs from a language model.

This assignment is designed to familiarize you with the concept and practice of prompt engineering for language models.

For this assignment, you will be asked to explore techniques for:

- Designing experiments to guide you in learning how to craft effective prompts (“queries”, “instructions”) to input into an LLM message.
- Optimize the LLM sequence prediction by analyzing and modifying your prompts based on prior LLM responses (e.g. the experiments).
- Understanding the nuances of how prompt design (“engineering”) influences the behavior of large language models.

The objective of teaching you these techniques is to enable you to:

- Effectively elicit sequence predictions (“generations”) from a large language model (“LLM”) and use the LLM response to complete a particular task.

In particular, in this assignment you will engage with an **OpenAI GPT OSS** or ***another LLM model through a free HuggingFace API*** to understand how prompt formulation can affect the quality and relevance of the generated text.

### Submission Requirements:

1. **Due Date:** Please submit your assignment to Canvas by November 2, 2025 at 11:59 PM ET.
2. **Submission format requirements:** Please respond to the homework assignment questions in a word processing document of your choice but upload the document to Canvas as a PDF.

## Part A. Experimenting with Prompt Design (10')

1. Choose **at least three** different tasks for the language model to perform (e.g., text completion, question answering, text summarization, translation, creative writing).
2. For each task, create a set of **five distinct prompts** that vary in structure, specificity, and style. Document the reasoning behind each prompt's design.
3. Use the **OpenAI API** to submit your prompts to the language model and record the outputs.
4. Analyze the results to determine which prompts were more effective for each task and why.
5. Discuss any patterns or insights observed in how the model responds to different types of prompts.
6. Based on your findings, develop a set of best practices for prompt engineering for the specific tasks you tested.
7. Document in detail your experimentation process, findings, the set of best practices you developed, with all the prompts used and the corresponding outputs from the language model, along with any code or scripts used to interact with the API.

## Part B. Experimenting with Zero, Few-Shot & Chain-of-Thought Prompting (10')

1. Experiment with "**chain-of-thought**" prompting where you provide a worked example within the prompt and compare the outputs to those without worked examples.
2. Develop a **few-shot prompt** which contains both an **instruction** and **several examples** of the task. You may either write your own examples from scratch, or take examples from the data set in the appendix.

**Requirement 1:** Conduct your analysis based on the following experiments:

1. The baseline prompt (e.g. your first prompt).
2. A prompt containing 1 example.
3. A prompt containing 3 examples.
4. A prompt containing 5 examples.
5. Two other prompt configurations of your choice. For example, you could try:

- a. Your 5 example prompt (above) but with the examples shuffled in a different order.
- b. A prompt where all the examples are intentionally mislabeled
- c. The same examples as your other prompts but with an alternative template.
- d. Modify the instructions of any of the prior prompts.

**Requirement 2:** Based on your experiments, answer the following questions:

- A. Define your training and test sets, and then create a table showing final test set accuracies for each prompt.
- B. What prompting formats did you experiment with? What worked well and what didn't work?
- C. What factors do you think most affect the model's performance?

**Requirement 3:** Document in detail your experimentation process and findings, with all the prompts used and the corresponding outputs from the language model, along with any code or scripts used to interact with the API.

**Requirement 4:** submit your assignment via Canvas as a PDF.

## Grading

Part	Requirement	Weight of Total Grade: 20%
		<i>Weight of Part:</i>
A	Prompt Design Experiments	10%
B	Zero, Few & CoT Prompt Experiments	10%

Grade Rubrics	Share of Assignment
Have you understood the question correctly?	10%
Experiment Design (e.g., <i>Creativity and diversity in prompt design; novelty in tasks</i> )	20%

Analytics (e.g., <i>Is there a clear rationale for why certain prompts worked or didn't work?</i> )	30%
Insights (Do you arrive at meaningful and informative insights?)	30%
Scientific Rigor (e.g., <i>Code quality, documentation, and reproducibility of the results</i> )	10%

**Due Date: November 2, 11:59pm ET.**

## Homework Assignment Policy:

- Remember to follow ethical guidelines and OpenAI's use-case policy when using the API.
- Your work should reflect a deep engagement with the material and a nuanced understanding of the capabilities and limitations of AI language models.
- **All Course Policies specified in the syllabus apply to this homework.**
  - Make sure to thoroughly read and understand the course policies.
  - Please email the teaching team any questions you have.
    - We are happy to provide further clarification or discuss any concerns or questions.

## Appendix: Examples for Few-Shot Prompt

### 1. Sentiment analysis for movie reviews.

Text: "(Lawrence bounces) All over the stage, dancing, running, sweating, mopping his face and generally displaying the wacky talent that brought him fame in the first place."

Sentiment: positive

Text: "Despite all evidence to the contrary, this clunker has somehow managed to pose as an actual feature movie, the kind that charges full admission and gets hyped on tv and purports to amuse small children and ostensible adults."

Sentiment: negative

Text: "For the first time in years, de niro digs deep emotionally, perhaps because he's been stirred by the powerful work of his co-stars."

Sentiment: positive

Text: "I'll bet the video game is a lot more fun than the film."

Sentiment: negative

## **2. Choice of Plausible Alternatives (COPA): Given the following premise, which of the following makes more sense?**

premise: "The woman waved."

choice1: "The woman spotted her friend from across the room."

choice2: "The woman ate her lunch at the park."

cause: choice1

premise: "The girl made a wish."

choice1: "She saw a black cat."

choice2: "She saw a shooting star."

cause: choice2

premise: "The woman hired a lawyer."

choice1: "She decided to sue her employer."

choice2: "She decided to run for office."

cause: choice1

premise: "My case was towed."

choice1: "I parked illegally."

choice2: "I jumped the battery."  
cause: choice1

premise: "The stain came out of the shirt."

choice1: "I patched the short."

choice2: "I bleached the shirt."

cause: choice1

premise: "I rubbed the soap between my hands."

choice1: "The soap foamed."

choice2: "My hands went numb."

cause: choice1

premise: "The police closed the investigation."

choice1: "The victim recovered."

choice2: "They apprehended the suspect."

cause: choice2