

Homework Assignment III: Coding, Uncertainty and Learning with LLMs

CS 6120 - Natural Language Processing

Name: _____ Student ID: _____

Due: April 17, 2026

Overview

This homework builds on the previous in-class activity on debugging and visualization using LLMs. In the in-class activity, you explored how **effective prompt engineering** can lead to tailored answers for your use case.

In this assignment, you will go deeper and investigate:

1. How **expressing uncertainty** in prompts influences LLM responses.
2. How to conduct data-driven analysis and visualization when **you cannot directly share your data** with an LLM, as we have talked in class about how external LLMs are not allowed in industry/government workplaces.
3. How these techniques connect to broader NLP principles and your learning journey in this course.

Part 1: Prompting and Uncertainty

(25 points)

Goal: Explore how expressing uncertainty in your prompts affects an LLM's reasoning and explanations.

The Code to Analyze

Function:

```
def transform_list(nums):  
    result = []  
    for i, n in enumerate(nums):  
        if i % 2 == 0:  
            result.append(n ** 2)  
        elif n % 3 == 0:  
            result.append(n // 3)  
        else:  
            result.append(n + nums[i - 1])  
    return sum(result[:3]) - result[-1]  
  
print(transform_list([2, 5, 6, 3, 4]))
```

Instructions:

1. Read the code and reason **without using LLM** about what it does, and predict the output.
2. Then, using your Northeastern account, query Claude multiple times, varying your confidence level:
 - “I’m 90% confident this outputs X. Am I correct?”
 - “I’m only 40% sure this outputs X. What do you think?”
 - “I’m unsure what this does, but I think this outputs X. What do you think?”
3. Observe how the LLM’s explanation, tone, or structure changes depending on how confident you sound.

Question 1.1: Your Prediction

What is your predicted output (before using any LLM)?

Question 1.2: Prompt Effectiveness

Which prompt style was most helpful?

☐ 90% confidence prompt ☐ 40% confidence prompt ☐ Unsure prompt ☐ All helpful

Question 1.3: LLM Behavior Analysis

How did expressing uncertainty shape the model's reasoning or level of detail?

Question 1.4: Uncertainty and Model Behavior

How does your exploration of uncertainty in prompting relate to what we've learned about:

- Language model training and probabilistic outputs?
- The role of context in shaping model responses?
- Prompt engineering as a form of “programming” language models?

Question 1.5: Documentation

Claude conversation link:

Part 2: Coding with LLMs But Without Sharing Data (25 points)

Goal: Analyze a multilingual human–LLM interaction dataset (**WildChat**) without directly inputting private data into the LLM.

Before You Begin

This homework must be completed in **one sitting** and may take up to **1 hour**. Please make sure you have enough uninterrupted time before starting.

Context

As we have talked in class, in real-world workplaces, sensitive or proprietary data often cannot be shared with an external LLM. Here, you will simulate such a constraint by performing analysis on a subset of the **WildChat** dataset, which contains multilingual dialogues between humans and LLMs. But you will not be sharing any of the conversations as data with an LLM.

Instructions

To complete this assignment, please access the website:

<https://nlp-homework.vercel.app/>

All parts of this assignment will be completed through the website. On the site, you will:

- Complete the data analysis task described below
- Use an integrated **AI assistant** to help with research question formulation, code writing, and visualization
- Answer reflection questions at the end of the assignment

Important constraint: You may use the AI assistant for guidance and code generation, but you **must not paste or share any dataset contents or conversations** with the assistant.

Tasks

1. **Dataset Preparation:** The dataset has already been downloaded and loaded for you on the website. Begin by exploring the available fields and understanding the structure of the data.
2. **Research Question Formulation:** Develop a hypothesis about human-LLM interactions. You may consult the AI assistant to *refine your question only* (do not paste data/conversations!). Apply any prompting techniques you learned.

Example Research Questions

- “Are LLM responses longer or more formal when users include politeness markers?”
- “Does the LLM produce more step-by-step explanations when users signal uncertainty?”
- “Do user turns become shorter over the course of a conversation?”

3. **Analyze the data:** Using the AI assistant to help write code, analyze the dataset according to your research question, but you will not be sharing the data directly. You may choose any analysis method or combination of the following, depending on your question:
 - **Compute aggregates:** Means, standard deviations, proportions, and any group differences relevant to your hypothesis.
 - **Statistical testing:** Choose any simple test suitable for your variable types (e.g., t-test, Mann-Whitney U, chi-square, correlation, regression).
4. **Create a visualization:** With the help of the AI assistant and a visualization library (e.g., matplotlib, seaborn), you will create a plot that will visualize the analyses you have done in the previous step to answer your research question. Again, do not input any conversations from the dataset, but you can describe the plot that you'd like to have to the LLM.

Part 3: Synthesis and Takeaways

(20 points)

Reflection on NLP Principles

Based on your work in Parts 1, 2, reflect on how this assignment connects to broader NLP concepts we've covered in class.

Question 3.1: Three Key Lessons

What are your three most important takeaways from this assignment about working with LLMs in practical NLP tasks?

Lesson 1:

Lesson 2:

Lesson 3:

Question 3.2: Future Applications

Based on this homework, identify one specific way you could apply these techniques:

In your current/future research:

In industry applications:

In other coursework:

Question 3.3: Comparative Analysis

Complete this comparison matrix based on your experiences across all parts:

Aspect	High Conf.	Low Conf.	No Data
Response Length	☐ Short ☐ Long	☐ Short ☐ Long	N/A
Explanation Detail	☐ Less ☐ More	☐ Less ☐ More	N/A
Assumptions Made	☐ Few ☐ Many	☐ Few ☐ Many	☐ Few ☐ Many
Clarifications	☐ Rare ☐ Often	☐ Rare ☐ Often	N/A
Effectiveness	☐ Low ☐ High	☐ Low ☐ High	☐ Low ☐ High

Question 3.4: Optimal Prompting Strategy

If you were to create a guide for future CS 6120 students on "How to Work Effectively with LLMs in NLP Tasks," what would be your top 3 recommendations?

Your LLM Best Practices Guide

Recommendation 1:

Recommendation 2:

Recommendation 3:

Submission Checklist

- ☐ Completed code analysis and prediction (Part 1)
- ☐ Tested all three confidence levels with LLM (Part 1)
- ☐ Formulated research question for WildChat analysis (Part 2)
- ☐ Created visualization without sharing raw data (Part 2)
- ☐ Completed all synthesis and reflection questions (Part 3)
- ☐ Included all Claude conversation links
- ☐ Attached visualization code file (.py or .ipynb)
- ☐ Attached visualization PNG file

Academic Integrity Statement

LLMs are allowed and encouraged for all tasks. Please:

- Document all prompts and interactions
- Verify all code and outputs by running them yourself
- Write all reflections and takeaways in your own words
- Be transparent about which parts were LLM-assisted
- Provide honest evaluations of LLM teaching effectiveness

Honor Statement: I confirm that I have completed this assignment individually and have properly documented all LLM interactions.

Name: _____ Student ID: _____

Signature: _____ Date: _____