

# HW4

Lukas White

September 2025

## 1 Introduction

### 1.1

Not every task requires the smartest model—a simple retrieval or intent classification task may be handled by a smaller, faster model, while harder tasks like deciding whether to approve a refund may benefit from a more capable model

What I want to do this week is build upon the previous things we learned; set up a simple ai agent, run an agent on a simple LLM running on my machine. And with that, we will try to process a basic command using inputted data, such as weather prediction or something along those line (give 3 more examples).

### 1.2

After some implimentation shown it **2.1** we can try to understand what is not working. The LLM seems to be confused. Some of the times, it doesn't do what is told since it does not really understand what we are asking it. Like with urgent, it does not seem to understand what urgent means. Also, it does not seem to understand how to convert units. This is more of an issue with the locally-run LLM than it is with the ai agent. To really determine if this is the case, I will ask it the same questions as before, however, I will feed it more precise wording and also, for the case of Celsius to Fareinheit, the formulas.

Now with **2.2** we can see that, after given important data like: formulas, current date, and more precise instruction the LLM can understand what to do and follow tasks correctly. I am actually shocked this locally-run LLM can do this. This means we can dive into a lot deeper and complex tasks.

## 2 LLM Testing

### 2.1

Prompt: Pretend you are a simple weather predictor. Based on the season being autumn, predict tomorrow's weather in one short sentence.

AI Output: It will be cloudy and rainy.

Prompt: Here is my to-do list: 1. Finish report (due Friday) 2. Buy groceries 3. Call mom (birthday tomorrow) 4. Book dentist appointment Please reorder these tasks from most urgent to least urgent.

AI Output: 1. Finish report 2. Buy groceries 3. Call mom 4. Book dentist appointment

Prompt: Here is a list of movies: The Matrix - sci-fi Inception - sci-fi The Notebook - romance Shrek - comedy The Godfather - crime Toy Story - animation Superbad - comedy Finding Nemo - animation Recommend me a good comedy movie to watch.

AI Output: Superbad - comedy

Prompt: Convert these temperatures from Celsius to Fahrenheit: 0, 20, 37, 100

AI Output: 0, 20, 37, 100

## 2.2

Prompt: Pretend you are a simple weather predictor. For autumn, give tomorrow's weather in one short sentence including temperature range in Celsius and condition (e.g., rain, sunny, cloudy).

AI Output: Tomorrow's weather: cool (10–15°C), cloudy skies, and light rainfall.

Prompt: Here is my to-do list: 1. Finish report (due Friday) 2. Buy groceries 3. Call mom (birthday tomorrow) 4. Book dentist appointment Reorder strictly from most urgent (soonest due or time-sensitive) to least urgent.

AI Output: 1. Call mom (birthday tomorrow) 2. Finish report (due Friday) 3. Buy groceries 4. Book dentist appointment

Prompt: Here is a list of movies: The Matrix - sci-fi Inception - sci-fi The Notebook - romance Shrek - comedy The Godfather - crime Toy Story - animation Superbad - comedy Finding Nemo - animation

Recommend exactly one comedy movie by name.

AI Output: Shrek - comedy

Prompt: Convert these Celsius temperatures to Fahrenheit using the formula  $F = (C \times 9/5) + 32$ : 0, 20, 37, 100

AI Output:  $9.0^{\circ}\text{C} = 48.2^{\circ}\text{F}$ ;  $5.0^{\circ}\text{C} = 41.0^{\circ}\text{F}$ ;  $32.0^{\circ}\text{C} = 89.6^{\circ}\text{F}$ ;  $0.0^{\circ}\text{C} = 32.0^{\circ}\text{F}$ ;  $20.0^{\circ}\text{C} = 68.0^{\circ}\text{F}$ ;  $37.0^{\circ}\text{C} = 98.6^{\circ}\text{F}$ ;  $100.0^{\circ}\text{C} = 212.0^{\circ}\text{F}$