From Prompts to Agents: Teaching AI to Think and Act

Understand LLMs, Prompting, Agents, and Tool Calling in Simple Terms

- Rashika Karki

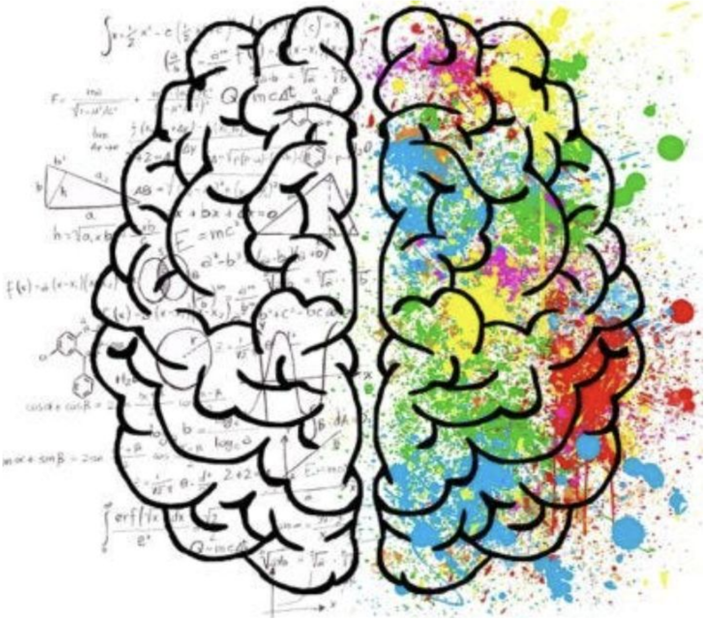
Let’s Start Simple — What Is an LLM?

LLM = Large Language Model

It’s like an AI that has read the entire internet.

It can:

- Answer questions
- Write code or poems
- Explain things like a teacher



Analogy:

Imagine you gave a really smart parrot billions of books — it can now "guess" what comes next in any sentence.

How Does an LLM Work?

- Predicts the next word in a sentence
- Learns language patterns
- Doesn’t “know” facts, it’s just really good at pattern-matching

It's not thinking — it’s generating.

Talking to LLMs = Prompting

You control the AI by how you talk to it.

This is called **Prompt Engineering.**

Basic Prompt:

💬 *“Tell me about Paris.”*

🧠 → Gives you a general overview.

Better Prompt:

💬 *“List 3 fun things to do in Paris with kids under 10, in a fun tone.”*

🎯 → More targeted and helpful.

Few-shot Prompting:

💬 *“Translate: ‘Hello’ → ‘Bonjour’, ‘Goodbye’ → ‘Au revoir’”*
Now translate: ‘Thank you’ →”

👉 The model learns from your examples.

Challenge

You’re given short, casual notes. Your task is to create a prompt that teaches the AI how to rewrite these notes as polite, professional messages.

Input:

Note: can't attend, busy

Note: send it fast

Note: need it now

Output:

Polite Message: I'm sorry, I won't be able to attend as I have a prior commitment.

Polite Message: Could you please send it as soon as possible?

Polite Message: I would appreciate it if you could provide it right away.

Prompts Alone Aren't Enough

LLMs are good at Language but not at real world action.

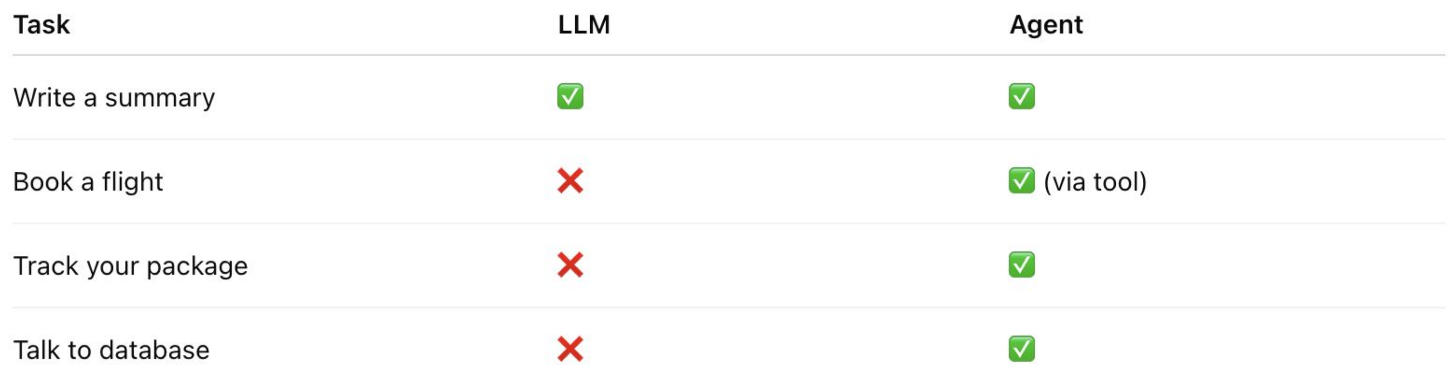
What if we need to:

- Search Google?
- Call an API?
- Book a meeting?
- Analyze a PDF?

This Is Where Agents Come In

An **AI agent** is like an intelligent assistant that can **perceive**, **reason**, and **act** to achieve a goal. It uses an LLM (like GPT-4) as the "brain" but goes beyond just answering questions.

Agent = LLM + Memory + Decision-making + Tools



Task	LLM	Agent
Write a summary	✅	✅
Book a flight	❌	✅ (via tool)
Track your package	❌	✅
Talk to database	❌	✅

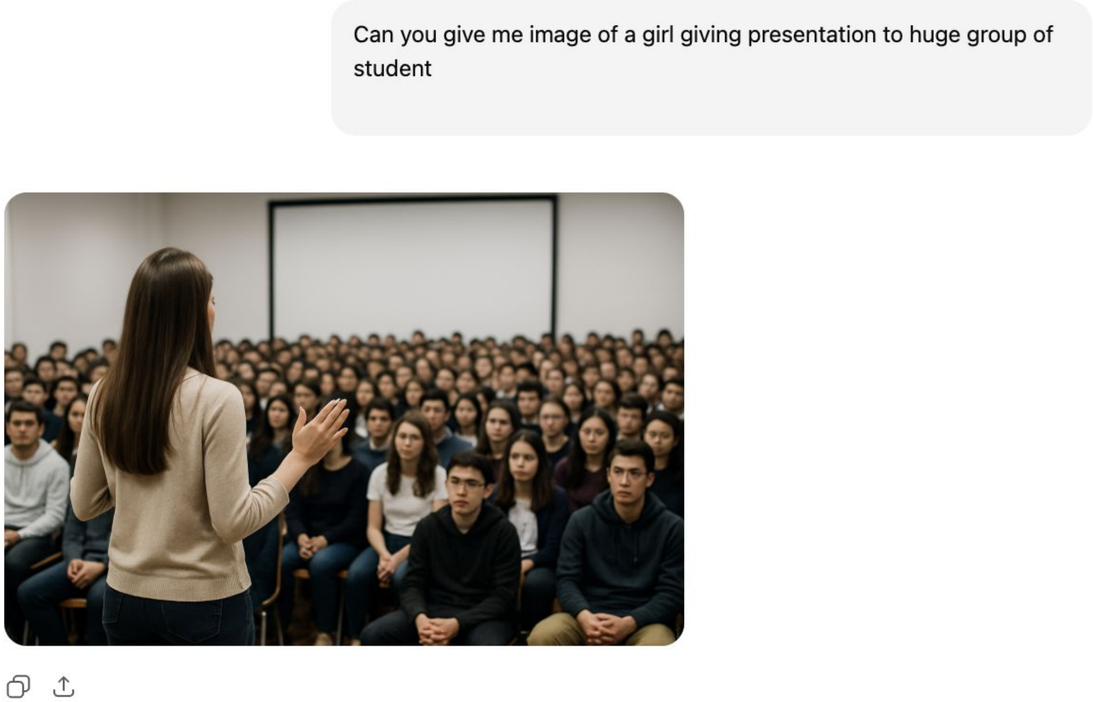
What is Tool Calling?

When the agent **uses a tool**, it’s called **Tool Calling**.

🛠️ A tool can be:

- An API
- A search engine
- A database
- A calculator
- A Python function

How do you think ChatGPT Image generation work?



Can you give me image of a girl giving presentation to huge group of student

Step 1: Prompt Interpretation

- You input a natural language description (e.g., *"girl giving presentation to students"*).
- ChatGPT processes the prompt to understand the **semantic intent** — people, actions, settings, emotions, etc.

Step 2: Tool Calling Under the Hood

- ChatGPT internally **calls a tool** (like `dalle_text2img`) to generate the image.

This is similar to calling an **API function** with structured parameters:

```
{
  "prompt": "girl giving presentation to a huge group of students",
  "size": "1024x1024"
}
```

Step 3: Image Generation by DALL·E

- The request is passed to **DALL·E model**, which:
  - Converts text to a **latent image representation**.
  - Decodes this representation into a high-res visual.

Step 4: Response

- ChatGPT receives the image output and returns it in the chat.
- You see the **image** in a conversational format.

Let's create your own Agent.

https://github.com/RashikaKarki/Workshop---Agent-and-Tool-Calling

Recap — The Big Picture

- **LLMs**: Speak and understand text
- **Prompting**: How we guide them
- **Agents**: Let LLMs *think and take action*
- **Tool Calling**: Let agents *act*