# 

**Mohannad Elhamod – IS883** 



## Recap

We have used...



- To query OpenAl's <u>pre-trained</u> language model using their API
- No model is stored locally.



#### **Hugging Face**

- To create and <u>fine-tune a</u> language model of our choice <u>on your own data</u>.
- The model is stored locally.



- To create an online service that hosts our solution.
- Makes our solution accessible to others.



## **Prompt Engineering vs Fine-Tuning**

Let's say you have a pre-trained model:

- If the model has not already been trained on the knowledge/patterns you wish to query, you probably need to fine-tune it.
- If it has been, but it is not giving the right answer, then you probably need to better engineer your prompt to extract the knowledge.



## **Prompt Engineering vs Fine-Tuning**

#### Prompt engineering

- More suitable for opensource data
- No need to update the model.
- May cost more in terms of queries

#### Fine-tuning

- More suitable when your data is sensitive/not open source.
- Requires more expertise
- May have significant training computational cost upfront.



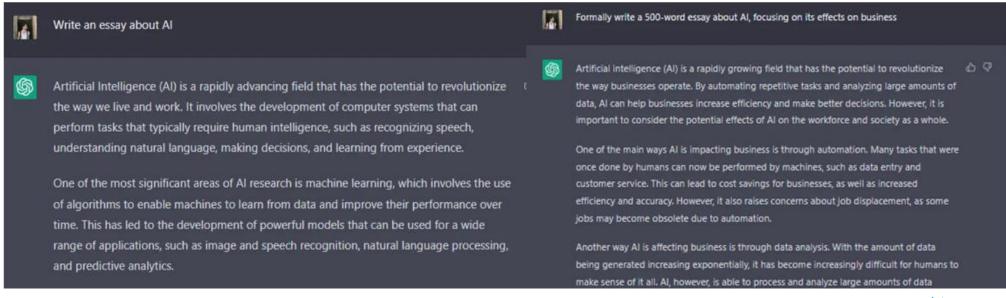
## **General Examples of Prompt Engineering**

- Writing in certain style.
- Prompting for a specific output type.
- Providing the model some examples.
- Asking the model to use certain tools.

• ...



## The prompt makes a big difference!



dataconomy



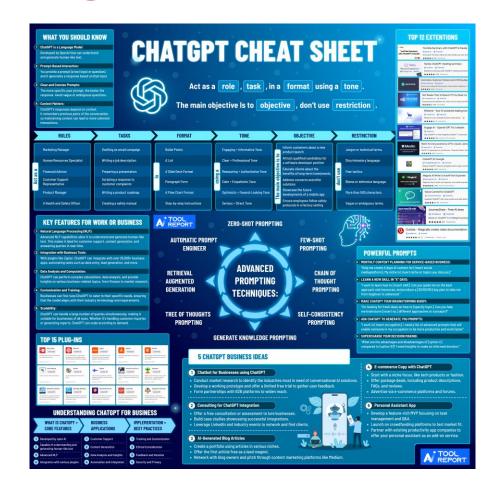
## Things to keep in mind when prompting...

- The more specific the prompt is, the better.
- Ask for the desired data formatting.
- Ask for a specific response length.
- Always be aware of previous context.



## ChatGPT prompt engineering

Link





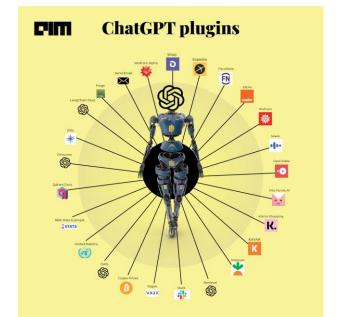
# Using Plugins



## What is a ChatGPT Plug-in?

It is a new feature that allows developers to augment ChatGPT's functionality such that:

- It can provide access to information external to the model.
- It can take actions based on the model's reply.
- Integrates with 3<sup>rd</sup> party functionality.



analyticsindiamag.com

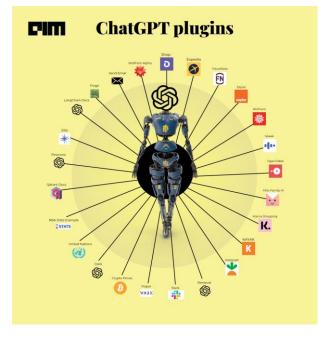


## Why are they useful?

- 1. **Enhanced Capabilities:** ChatGPT Plugins will enhance the AI's capabilities, allowing it to analyze data at a faster rate than ever before.
- 2. Improved Accuracy: With access to a vast array of knowledge and information, ChatGPT Plugins will be able to provide more accurate responses to users.
- Increased Efficiency: ChatGPT Plugins will allow businesses and individuals to automate tasks and processes, saving time and increasing efficiency.
- 4. **Customizability:** Eventually ChatGPT Plugins will be customizable to meet the specific needs of businesses and individuals, making them versatile tools.

Medium.com



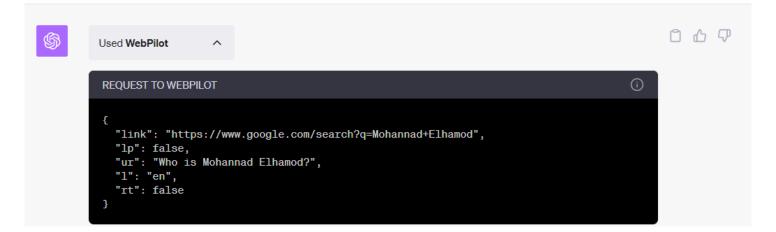




## Plug-ins behind the scene

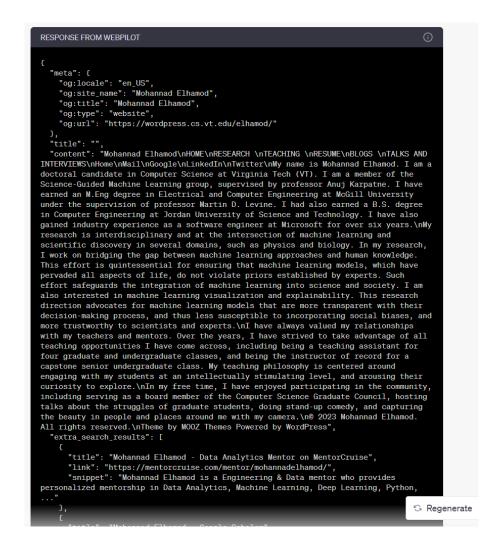


Who is Mohannad Elhamod





## Plug-ins behind the scene





## How do these plug-ins work?

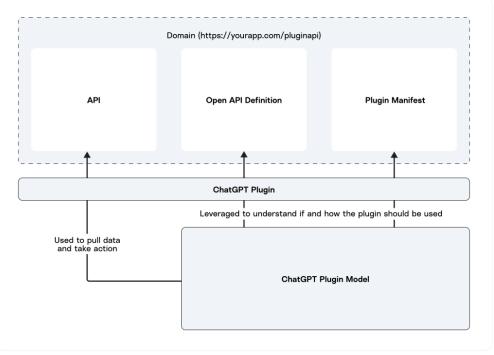
Each Plug-in has a description.

When faced with a prompt, Chat GPT determines which available plug-

in is most suitable.

ChatGPT invokes the plug-in (e.g., through an API).

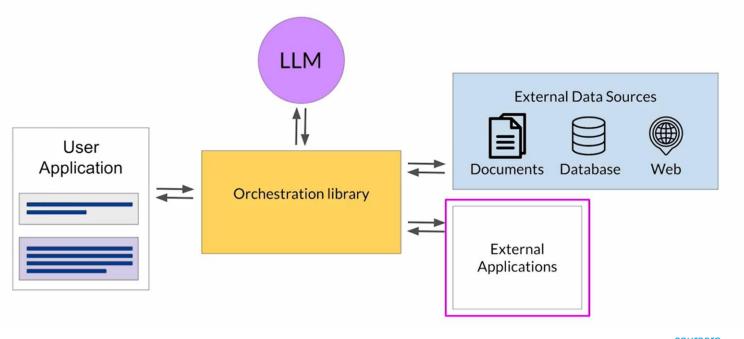
- ChatGPT processes the returned response.
- The cycle is repeated, as necessary.





nuvalence.io

## How do these plug-ins work?

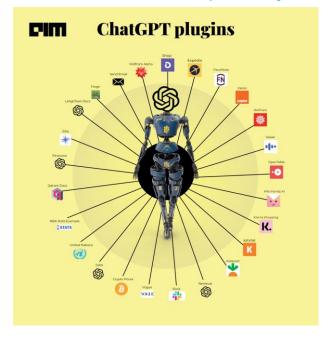




## Let's look at it live

Demo

#### analyticsindiamag.com



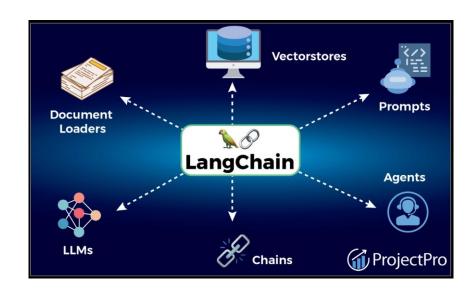


# Langchain



## What is LangChain?

A framework/package that simplifies the use of LLMs (e.g., OpenAl API) by providing functionality for document reading, chatbots, summarization, etc.





## Let's see how it works.

**Demo** 



## Roles in OpenAl API

#### System role

The system role also known as the system message is included at the beginning of the array. This message provides the initial instructions to the model. You can provide various information in the system role including:

- A brief description of the assistant
- Personality traits of the assistant
- Instructions or rules you would like the assistant to follow
- Data or information needed for the model, such as relevant questions from an FAQ

You can customize the system role for your use case or just include basic instructions. The system role/message is optional, but it's recommended to at least include a basic one to get the best results.

#### Messages

After the system role, you can include a series of messages between the user and the assistant.

```
{"role": "user", "content": "What is thermodynamics?"}
```

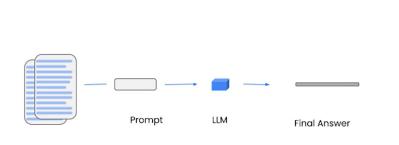
To trigger a response from the model, you should end with a user message indicating that it's the assistant's turn to respond. You can also include a series of example messages between the user and the assistant as a way to do few shot learning.



## Q: What to do if document is too long

It is not going to fit in one context window

#### Stuff method



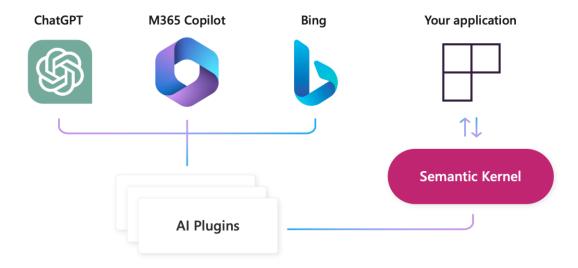


## Microsoft Semantic Kernel



### **Semantic Kernel**

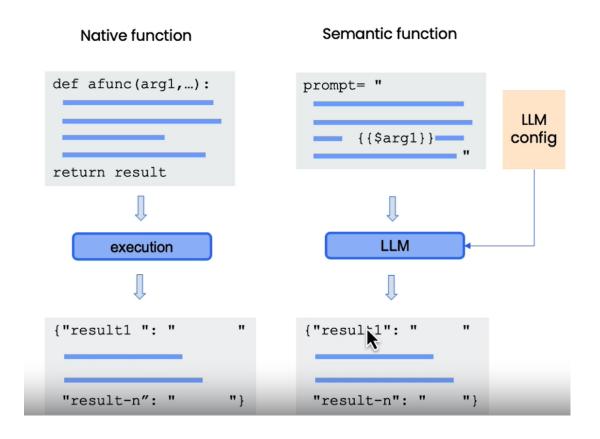
Another framework for developing plugins.





## **Semantic Kernel**

Mostly used for "semantic" functions.





## **Example 1: SWOT analysis**

#### A business strategy tool:

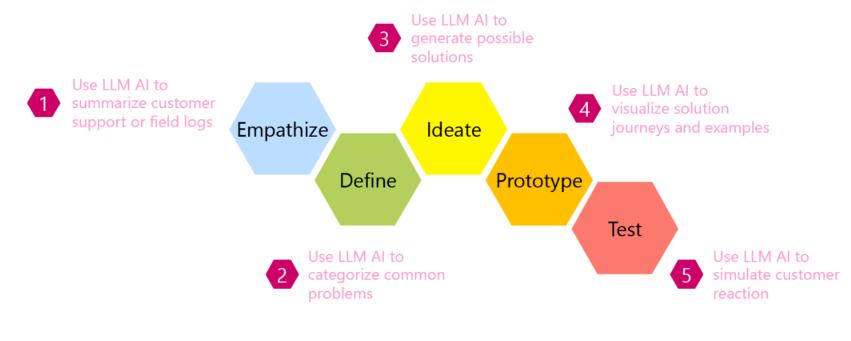
- Strengths
- Weaknesses
- Opportunities
- Threats

#### **SWOT ANALYSIS**





## **Example 2: Design Thinking**



deeplearning.ai





https://github.com/elhamod/IS883/blob/main/Semantic\_Kernel.ipynb

