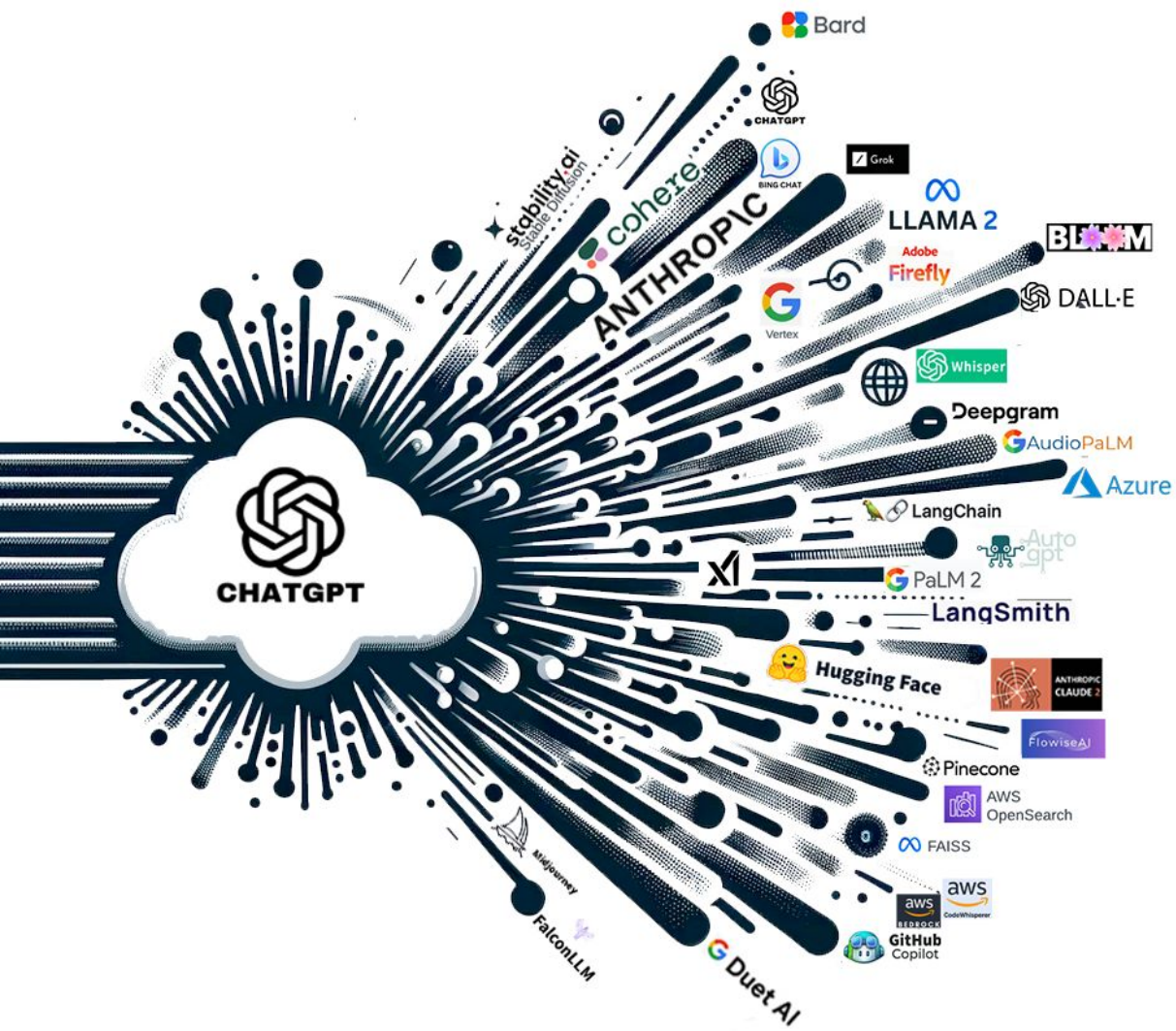


Beyond ChatGPT

Beyond ChatGPT

Cian Clarke

linkedin.com/in/cianclarke

cianclarke.com

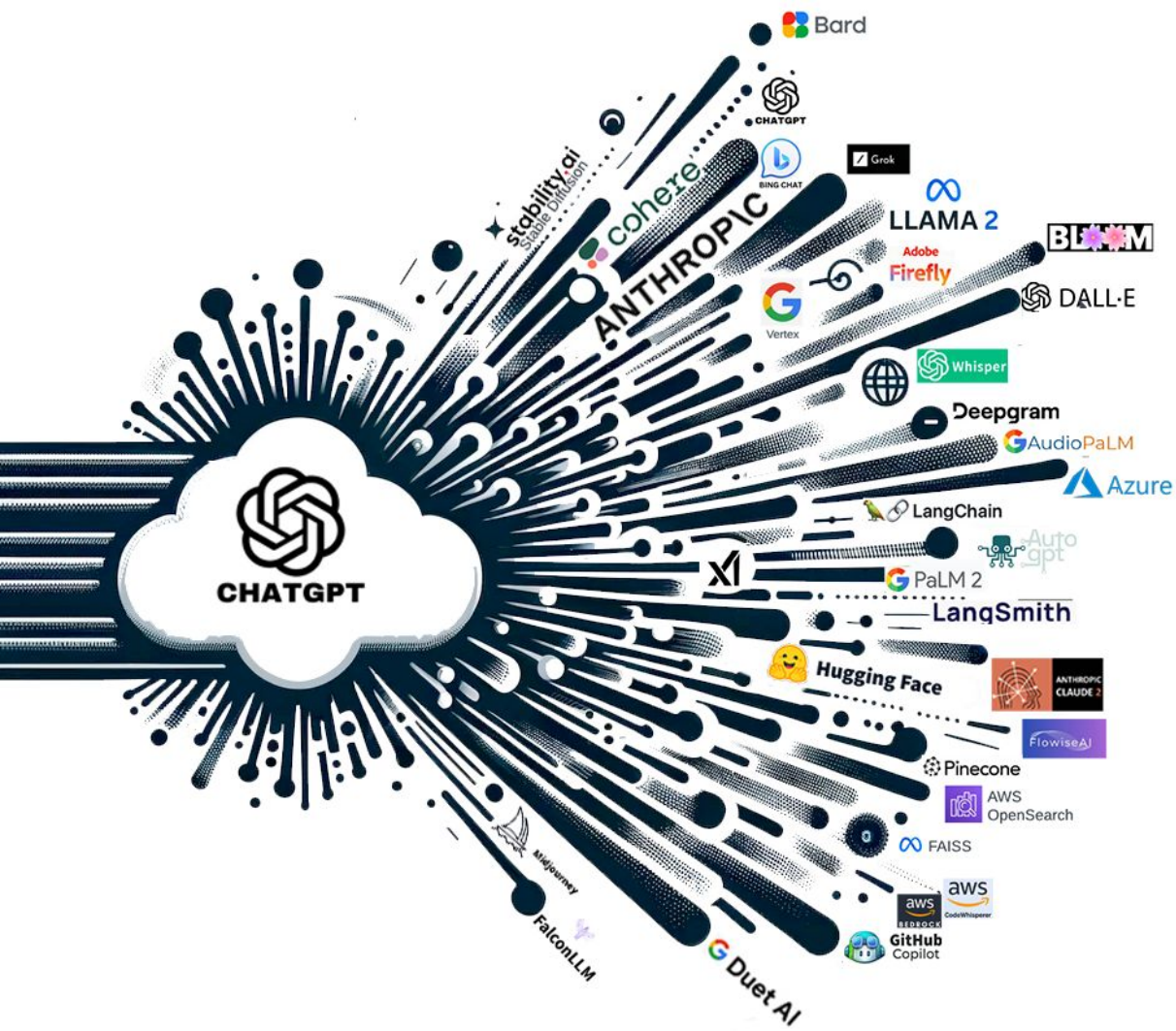
SERVISBOT

What I’ll cover

1. New @ OpenAI
2. LLM Landscape
3. Prompt Engineering Techniques
4. Demos

new @

- Developer Day last Monday
- GPT-4 with 128k token context length
- “GPTs”
- Assistants
- Update to “Training cut off”
- GPT-4 Vision & Fine Tune APIs


Bard
CHATGPT
Grok
BING CHAT
stability.ai
Stable Diffusion
cohere
ANTHROPIC
LLAMA 2
BLOOM
Adobe
Firefly
Vertex
DALL·E
Whisper
Deepgram
AudioPaLM
Azure
LangChain
Auto
gpt
CHATGPT
PaLM 2
LangSmith
Hugging Face
ANTHROPIC
CLAUDE 2
FlowiseAI
Pinecone
AWS
OpenSearch
FAISS
aws
CodeWhisperer
aws
BEDROCK
GitHub
Copilot
Midjourney
FalconLLM
Duet AI

LLM Landscape

Commercial LLMs

OpenAI
ANTHROP\C
cohere
stability.ai
xAI
PaLM 2

OSS LLMs

LLAMA 2
FalconLLM
BLOOM

Chat

CHATGPT
Bard
BING CHAT
ANTHROPIC CLAUDE 2
Grok

Model Hosting & White-label LLMs

Hugging Face
Azure
$10bn
OpenAI
aws BEDROCK
$4bn
ANTHROP\C
Vertex
$2bn

Voice Models

Whisper
Deepgram
AudioPaLM
stability.ai Audio

Image Models

stability.ai Stable Diffusion
Midjourney
DALL·E
Adobe Firefly

Code Assist

aws CodeWhisperer
GitHub Copilot
Duet AI

OSS Frameworks

LangChain
LangSmith
Auto gpt
FlowiseAI

Vector DBs

Pinecone
AWS OpenSearch
FAISS
Weaviate

Embedding Algs

OpenAI
text-embedding-ada
cohere
Vertex BERT
Word2Vec

52n.ai

Prompt Engineering 101

What is Prompt Engineering?

The art of crafting input to an LLM in order to direct the model towards a specific output enables LLMs to achieve more complex tasks.

Principals of Prompt Engineering:

- Instructions should be **Clear** & **Specific**

```
Write something about dogs
=>
Describe the benefits of having a dog as a pet for a family living in an urban environment
```

- Describe a role to play:

```
You are a nutritionist working for a leading research institute. Evaluate this meal plan for.....
```

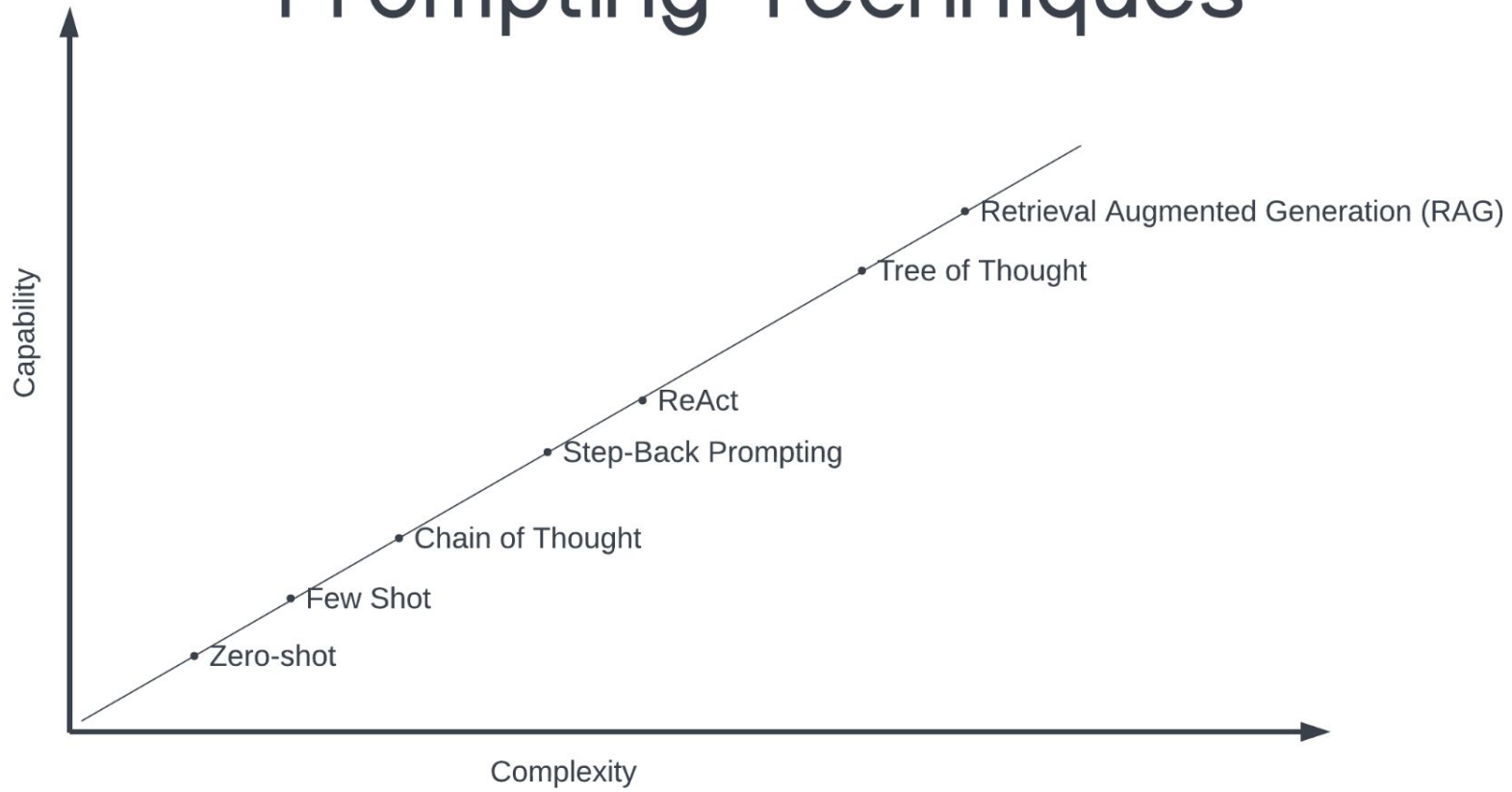
- Iterate & refine prompts

```
Write about Napoleon Bonaparte
=>
Write a detailed account of Napoleon's impact on their society
=>
Explain the economic reforms enacted by Napoleon and there long-term effects on French society
```

- Use delimiters

```
The following is a transcript between an agent and a customer:
"""
14:00 Agent: Hi, how can I help you?
14:01 Customer: I need a refund
14:05 Agent: No, gway
"""
Evaluate the interaction for friendliness, professionalism and responsiveness.
```

Prompting Techniques



Prompting Techniques

Zero Shot

Have the LLM attempt a task "zero shot" without any training examples. Most prompts users enter to ChatGPT are "zero shot"

Classify the text into positive, neutral or negative:
Text: Pure deadly
Classification:
Neutral

Few Shot

Supply some previous turns in context to help the LLM generate a correct response

Classify the text into positive, neutral or negative:
Text: That's deadly
Classification: Positive
Text: Fair play
Classification: Positive
Text: Zero craic
Classification: Negative
Text: Pure deadly
Classification:
Positive

Chain of Thought

Provide the LLM with reasoning steps in order to arrive at the correct solution

Few shot: Provide example reasoning of a similar problem statement to assist an LLM in arriving at a correct solution

One shot: Tell the LLM "Let's think step by step"

Tree of Thoughts

Have an LLM explore several different trees of thoughts, with a variety of strategies potentially used to then decide on the most promising result (BFS, DFS, custom)

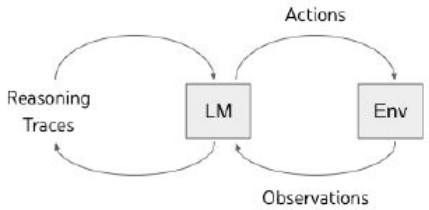
```
Imagine three different experts are answering this question.
All experts will write down 1 step of their thinking, then share it with the group.
Then all experts will go on to the next step, etc.
If any expert realises they're wrong at any point then they leave.
The question is {question}
```

52n.ai

Prompting Techniques

ReAct

Combines "Acting" and "Reasoning" to allow the LLM to achieve complex reasoning tasks



ReAct (Reason + Act)

Example (from ReAct paper):

Question: How many rooms are in the hotel that is home to the Cirque du Soleil show Mystere?

HotpotQA label: 2,664 **Outdated** ✗

(1d) ReAct (Reason + Act)

Thought 1: I need to search Cirque du Soleil show Mystere, find the hotel it is in, then find the number of rooms in the hotel.
Act 1: Search[Cirque du Soleil show Mystere]
Obs 1: Could not find Cirque du Soleil show Mystere. Similar: ['Cirque du Soleil' …

Thought 2: To find the hotel, I can search Mystere (Cirque du Soleil).
Act 2: Search[Mystere (Cirque du Soleil)]
Obs 2: Mystère (my-steer) is … Held in a custom theatre at the Treasure Island Hotel and Casino …

Thought 3: Mystere is held in a custom theatre at the Treasure Island Hotel and Casino. So I need to search Treasure Island Hotel and Casino next and find the number of rooms in the hotel.
Act 3: Search[Treasure Island Hotel and Casino]
Obs 3: Treasure Island Hotel and Casino … with 2,884 rooms and 220 suites …

Thought 4: Treasure Island Hotel and Casino has 2,884 rooms and 220 suites. So the answer is 3,104.
Act 4: Finish[3,104]

Up-to-date ✓

Prompting Techniques

Retrieval Augmented Generation (RAG)

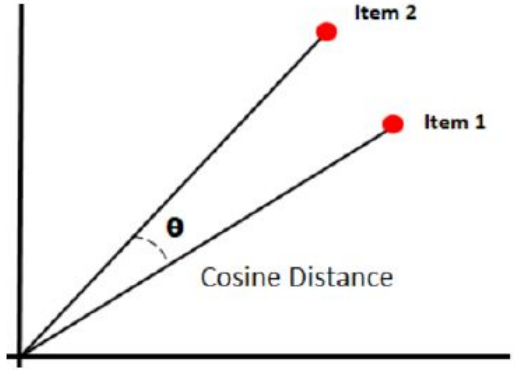
When asking a complex question which relies on obscure, or often propietary knowledge, an LLM will often

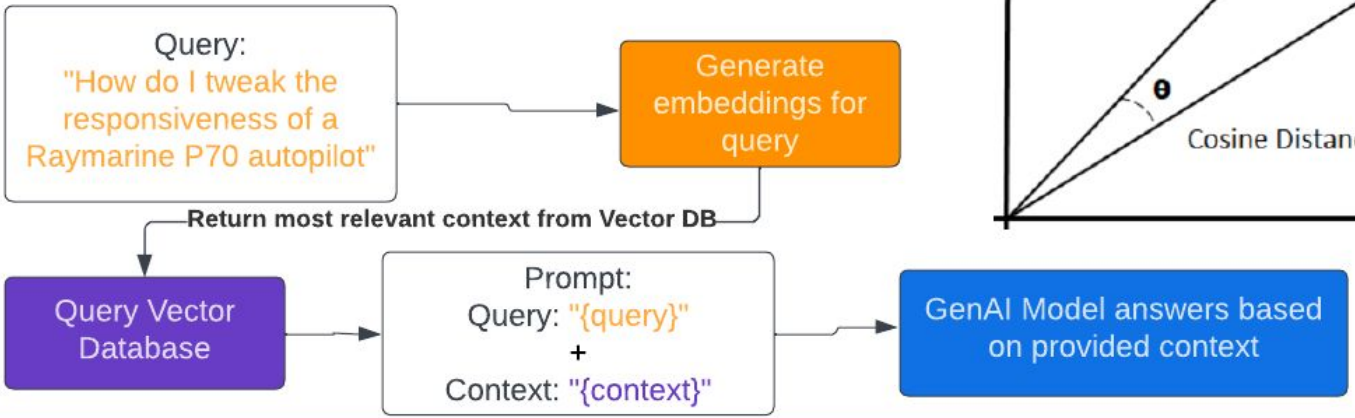
- refuse to answer (best outcome) - OR
- hallucinate a false answer (most likely, most harmful)

RAG reduces the liklihood of this scenario by incorporating a knowledge retrieval capability from a store of documents, in turn providing this context to an LLM to phrase more accurate answers.

Enables:

* Ingesting of knowledge beyond "training cutoff"
* Query custom knowledge about a company/customer
* Search of own documents





Who to follow & Other Resources?

- Discord
  - Langchain, OpenAI, AutoGPT, Midjourney
- https://cobusgreyling.medium.com
- https://www.linkedin.com/in/ruben-hassid/
- https://www.promptingguide.ai/techniques
- https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

?

linkedin.com/in/cianclarke

cianclarke.com

SERVISBOT