

By 後端里長伯(hohshen)

目錄

- 1. 除了respect還是respect
- 2. 厲害的鸚鵡
- 3. 今天天氣如何?
- 4. 蛤?
- 5. 讓我看看



敬山姆 阿特曼

世上第一個因AI而丟工作的人 _{我是誰?我在哪裡?}

世上第一個因AI讓董事會丟工作的人

世上第一個因AI改變世界的人

LLM 介紹

最近一直在思考LLM 如何淺顯易懂的介紹LLM 直到找到了這部影片



LLM 介紹

因為類似玩接龍遊戲, 有時可能會一本正經地亂回答 當錯誤一多了就積少.....

LLM 介紹- OpenAl API

讓我們來看看這個例子吧!

OpenAl API (completions)

```
1 POST 'https://api.openai.com/v1/completions'
2 Body {
3    "model": "gpt-3.5-turbo-instruct",
4    "prompt": "XX",
5    "max_tokens": 300,
6    "temperature": 0.8
7 }
```



API設定參數

"model": "text-davinci-003",

"prompt": "XXX",

"max_tokens": 300,

"temperature": 0.8,

"top_p": 0.1

Temperature: 隨機性, 機率的選擇

- (0~2, default: 1)
- 控制每次回答的內容比較一致

Top_p: 創作性

■ (0和1,通常設定在0.1到0.9之間)

Prompt

使用三個引號"""或者三個井字號###區分指令與內容 調控 Temperature 及 Top_p parameters 參數,建議一個一個調

- 1 請寫一篇facebook貼文,主題是
- 2 """
- 3 什麼人適合學習LLM
- 4 """
- 5 貼文內容請滿足以下條件
- 6 1.500個字左右
- 7 2.不需要名詞解釋
- 8 3.請舉例說明
- 9 4.可以加入emoji符號

Prompt: In-Context Learning

Zero shot

- 靠的是模型預訓練的能力而做出的生成回應
- e.g. 莎士比亞是...

One shot

- 造樣照句
- e.g. 台北->台灣 東京-> response 東京->日本

Few shot

- 舉一反三
- e.g. 臨摹法

LLM 介紹- OpenAl API

OpenAl API (images/generations)

```
1 POST 'https://api.openai.com/v1/images/generations'
2 Body {
3 {
        "model": "dall-e-3",
        "prompt": "麥當勞美味套餐",
        "n": 1,
        "size": "1024x1024"
8 }
```

n: 一次產生多少圖

quality: standard(default),hd

LLM 介紹- OpenAl API

Model endpoint compatibility

ENDPOINT	LATEST MODELS
/v1/assistants	All models except gpt-3.5-turbo-0301 supported.retrieval tool requires gpt-4-1106-preview or gpt-3.5-turbo-1106.
/v1/audio/transcriptions	whisper-1
/v1/audio/translations	whisper-1
/v1/audio/speech	tts-1,tts-1-hd
/v1/chat/completions	gpt-4 and dated model releases, gpt-4-1106-preview, gpt-4-vision-preview, gpt-4-32k and dated model releases, gpt-3.5-turbo and dated model releases, gpt-3.5-turbo-16k and dated model releases, fine-tuned versions of gpt-3.5-turbo
/v1/completions (Legacy)	gpt-3.5-turbo-instruct, babbage-002, davinci-002
/v1/embeddings	text-embedding-ada-002
/v1/fine_tuning/jobs	gpt-3.5-turbo, babbage-002, davinci-002
/v1/moderations	text-moderation-stable,text-moderation-latest
/v1/images/generations	dall-e-2, dall-e-3

This list excludes all of our deprecated models.

更多API endpoint

Hello World



開發整合性更高的LLM應用

LangChain六大模組

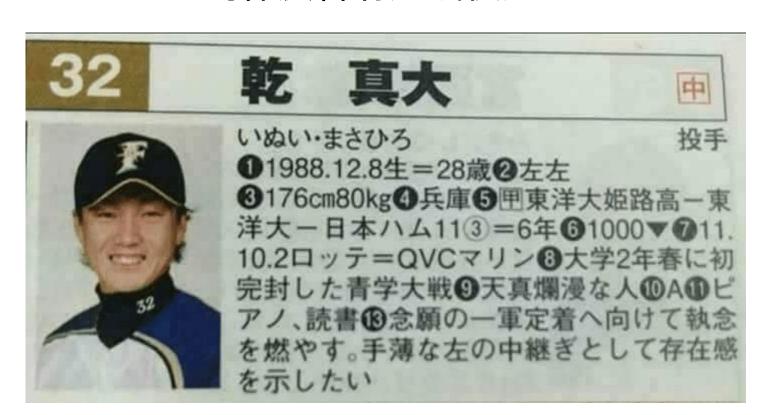
- Model I/O
 - Prompts+Template
 - Model
 - Parser
- Agent
- Chain
- Memory
- Retrieval(Index)

demo 00_hello 01_prompt 02_template

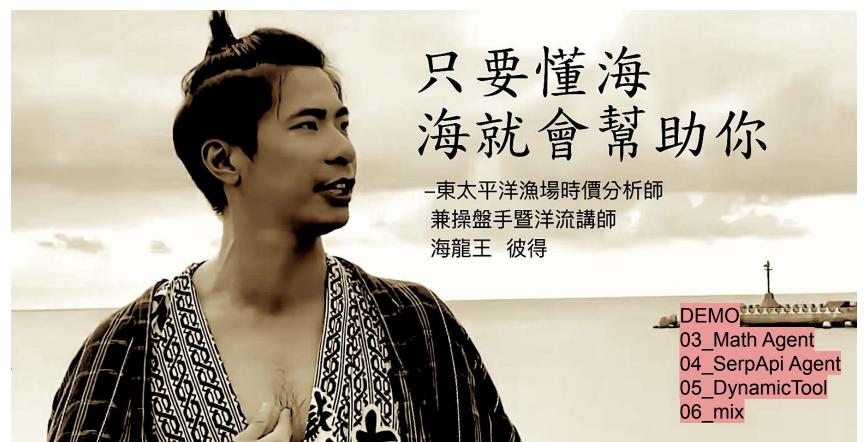
今天天氣如何?

同學請回答

為什麼會有天氣很大?



專業的事就是交給專業的!





陳先生,聽說你昨天開車衝進淡水河



哇!和運怎麼都知道



總機就是要專業分工

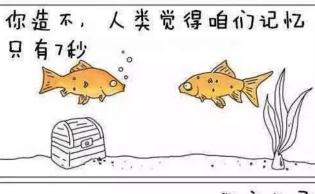
DEMO 07_simple_seq 08_seq 09_router 10_parser

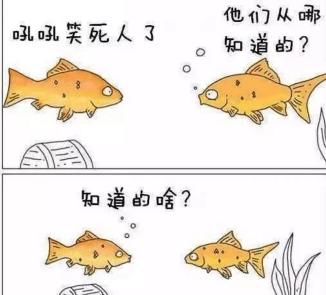
一顆睿智的蛤蜊曾說過



『蛤?』

蛤?





北部の加雪田田県

大家知道金魚缺少了什麼嗎?

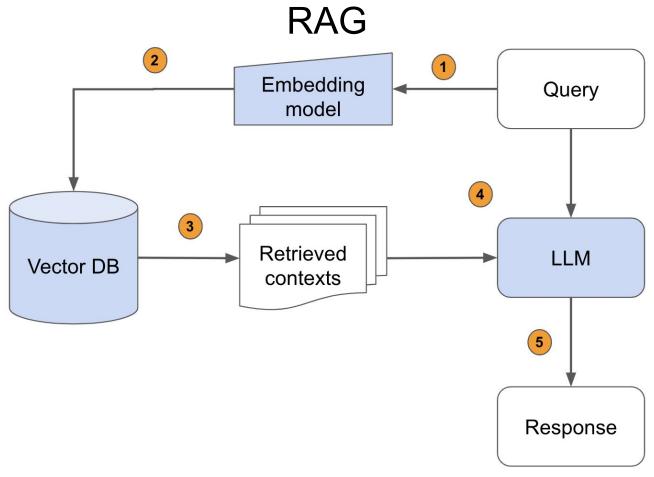
demo 11_memory 12_history

Embedding similarity search

There are many ways measuring the similarity of vectors

Metric	Formula	Interpretation
Euclidean distance	$d = \sqrt{\sum_{i}^{n} (x_i - y_i)^2}$	x d y
Manhattan distance	$d = \sum_{i=1}^{n} x_i - y_i $	$d_1 = \frac{y}{d_2}$ $d_2 = d_1 + d_2$
Chebyshev distance	d = max(x: - y:)	$d_1 = \frac{y}{d_2}$ $d_2 = \max(d_1, d_2)$
Cosine distance	$\Theta = \frac{x \cdot y}{ x \cdot y }$	\overrightarrow{x}

demo 13 embending



Demo 14_rag

檢索增強生成(Retrieval-Augmented Generation, RAG)

RAG

RAG v.s. Fine Tuning

	微調模型	RAG
比喻	就像考試前認真讀書,考試 closed	就像考試 open book,帶筆記去考試,若 筆記上有寫可以回答的很好
缺點	訓練模型需要花時間和計算成本,不可能隨時訓練更新資料	仍有 Token 長度限制 要用工具抓資料因此處理時間較長
優點	品質可能更好,這需要機器學習專業知識	不用擔心新資料更新

讓我看看



demo 15_image

生成圖片



demo 16_image_gen

謝謝大家

