



工業技術研究院
Industrial Technology
Research Institute

委託案第一期報告

能源資產績效管理知識 GPT 模型 程式委託開發

年度計畫：自 113 年 8 月 1 日至 114 年 2 月 28 日

委辦機關：財團法人工業技術研究院

執行單位：國立臺灣海洋大學

中 華 民 國 113 年 9 月

目錄

1. 研究背景	2
2. 研究方法	3
3. 建置成果報告	5
(1) POXAGPT Front-End UI 之設計與實作	5
(2) POXAGPT Data Manager 之設計與實作	5
(3) POXAGPT Controller 之設計與實作	7
(4) POXA 自動摘要功能	9
4. 目標達成情形	11

1. 研究背景

目前電力市場和能源政策包含大量複雜的規範和資料，資料分散於 POXA Info 網站的多個 Blog 文章和多份電力政策相關資訊之 PDF 文件中，這使得相關人員在尋找和解析資訊時面臨挑戰，難以快速了解市場的機會與風險。其中 POXA Info 網站¹每週都會分析電力交易市場的相關資料與數據，並會發布一份報告，總結市場活動，報告會包含電價、不同公司參與度和新規則等相關資訊。本委託案將規劃開發能源資產績效管理知識 GPT 模型，簡稱 POXAGPT：一個專門針對電力交易和能源相關資訊的聊天機器人服務，POXAGPT 將透過整合多個資料來源（POXA Info 及台電能源局規範 PDF）並提供問答功能，以簡化資訊檢索和解析的過程，提升使用便利性(Usability)及擴大服務對象。

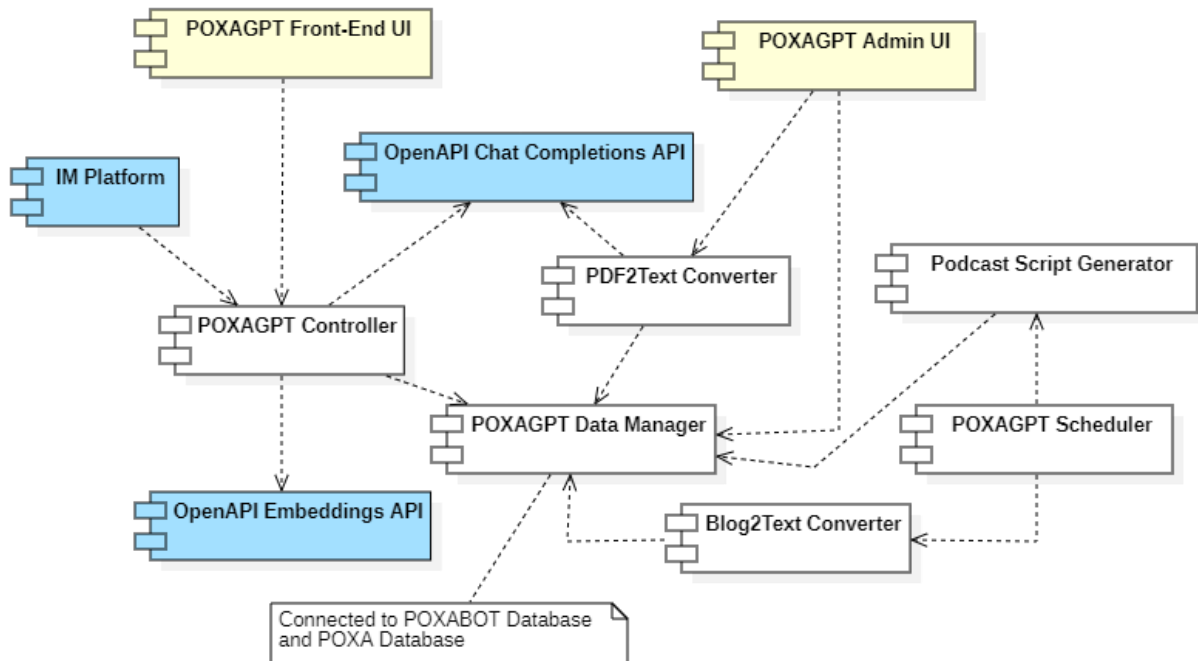
POXAGPT 服務之主要目標包含：

1. 提供一個能夠自動回答使用者問題的聊天機器人。
2. 可快速檢索和解析電力交易和能源政策相關的規範和資訊。
3. 能自動生成 Podcast 的腳本和講稿，以利電力新知資訊之傳播與推廣。

¹ <https://info.poxa.io>

2. 研究方法

POXAGPT 服務之預期系統架構如下圖所示，前端模組標示為黃色，後端模組標示為白色，外部模組則標示為藍色。



1. POXAGPT Front-End UI：是 POXAGPT 的主要聊天/問答介面，可讓使用者提問並顯示回覆資訊。
2. POXAGPT Admin UI：是系統管理者介面，允許管理員新增新的電力法規或其他技術資訊之 PDF 文件到系統中。
3. POXAGPT Controller：核心控制器，負責處理使用者查詢和調用其他模組，如直接呼叫 PoxaBot Data Manager 進行資料查詢，亦會外部調用 OpenAI Embeddings API 以實現 Retrieval Augmented Generation (RAG)；透過結合 POXAGPT Data Manager 與 OpenAI 之 Function Calling 機制，以能進行資料查詢(如查詢指定時間之 dReg 商品價格)；以及調用 OpenAI Chat Completions API，以生成一般性的對話回覆資料。
4. POXAGPT Data Manager：資料管理模組，負責從對 Blog 和 PDF 轉換出的 JSON 資料進行讀寫操作，亦即負責與資料庫溝通，以能儲存和檢索轉換後的文本資料。此外，除了 Blog 與 PDF 轉換之文本資料庫外，POXAGPT Data Manager 亦可銜接 POXA Info 系統本身的資料庫，以利 POXAGPT Controller 實現 Function Calling。
5. Blog2Text Converter：Blog 內容轉換器，負責將 Blog 網頁內容轉換為文本，或取得原始的 JSON 資料，再轉換為適當的 JSON 格式 (例如過濾掉不相干的資訊)，最後再透過 POXAGPT Data Manager 將轉換後的資料儲存到資料庫。
6. PDF2Text Converter：PDF 文件轉換器，負責將包含法規或技術內容之 PDF 文件轉換為文本，然後轉換為適當的 JSON 格式，最後透過 PoxaBot Data Manager 將轉換後的資料儲存到資料庫。目前規劃運用 PyPDF2 套件²來擷取 PDF 內文，並透過 OpenAI Chat Completions API 以進行文本校正。
7. Podcast Script Generator 用於生成 Podcast 腳本，根據 Blog 資料生成相關的 Podcast 腳本。

² <https://pypi.org/project/PyPDF2/>

8. POXAGPT Scheduler：負責管理定期任務，安排和執行周期性工作，如每週使用 Blog2Text Converter 進行 Blog 內容的轉換與儲存，及定期調用 Podcast Script Generator 生成 Podcast 腳本。

POXAGPT 服務系統預計以 Python 為主要開發語言，後端模組 API 規劃以 FastAPI 框架建置，前端規劃運用 jQuery/Bootstrap 套件或 react 框架實現使用者介面。

本年度第一期規劃完成之功能雛型包含：

- 設計與實作 POXAGPT Front-End UI。
- 設計與實作 POXAGPT Data Manager，並先預載範例資料，以能成功提供基本問答效果。
- 設計與實作 POXAGPT Controller，以能串接前端、Data Manager 與 Open AI 之 API。Controller 會運用 OpenAPI 之 Chat Completions API，以生成合適的回覆資料。

3. 建置成果報告

(1) POXAGPT Front-End UI 之設計與實作

前端介面的核心概念為「讓使用者能輕鬆上手使用」，主要設計考量包含：

- 參考了目前市面上的聊天機器人介面，如 Facebook messenger。
- 提供功能按鈕，讓使用者可以清楚地知道 POXAGPT 提供了哪些功能，以避免使用者對 POXAGPT 有錯誤的預期。

目前的前端介面是以 React 和 Bootstrap 進行實作，已建置之成果畫面如下：



(2) POXAGPT Data Manager 之設計與實作

目前 Data Manager 是透過 MongoDB 進行資料儲存，我們設計了以下的 Collection：

- WebInformation.article
 - 存放 Blog 文章資訊之 Collection 結構：
 - title
 - content
 - label
 - block
 - blockTitle
 - blockContent
 - section
 - sectionTitle
 - sectionContent
 - 如何運用於 QA 問答功能：
 - title：用於判斷問題的時間點
 - content：用於建立資訊搜尋的索引
 - label：用於文章分類

- block：用於生成回答的文章資料來源(相似度分析&關鍵字搜尋)
- section：用於生成回答的文章資料來源(相似度分析&關鍵字搜尋)

```
"title": "POXA 2024 8/12 每週動態",
"content": "感謝訂閱POXA Info 每週最新動態！即時備轉價格大漲21.66%，調頻/E-dReg的投標量跟得標量近年的變化趨勢為何？與收入跟成本的關係",
"labels": {
  "0": "E-dReg",
  "1": "dReg"
},
"block": {
  "0": {
    "blockTitle": "📢 台電最新公告",
    "blockContent": "8/6 修正「公告事項4-3」，僅修正部分文字敘述。",
  },
  "1": {
    "blockTitle": "📉 本週主題分析",
    "blockContent": "本週調頻備轉減少至777.5MW，主要是又有1家綠悠能源轉換至E-dReg，價格連續199天維持0元，平均得標率小升至64.3%，",
  },
  "2": {
    "blockTitle": "📊 市場最新動態",
    "blockContent": "「調頻服務」平均結清價格維持在0元/MWh；目前累計參與容量減少至777.5MW，較上週減少5MW，來自於1家現有廠商綠悠能",
  },
  "3": {
    "blockTitle": "💡 台電電力供需資料",
    "blockContent": "本週再生能源佔比維持12%，最高滲透率降至27.99%，再生能源佔比與上週相同，雖然用電負載回復至正常水平，但再生能",
  },
  "4": {
    "blockTitle": "POXA ENERGY 專屬分析 🧐",
    "blockContent": "分析電網頻率及供需變化對dReg/sReg儲能充放電排程的影響。",
  },
  "5": {
    "blockTitle": "📰 資訊來源網站",
    "blockContent": "整理各種POXA定期追蹤的原始資料來源網站。",
  },
}
```

```
"section": {
  "0": {
    "sectionTitle": "📢 台電最新公告",
    "sectionContent": "📢 台電最新公告\n8/6 修正「公告事項4-3：日前輔助服務市場需求量估算方式說明文件」，僅修正部分文字敘述。",
  },
  "1": {
    "sectionTitle": "📉 本週主題分析",
    "sectionContent": "📉 本週主題分析\n本週調頻備轉減少至777.5MW，主要是又有1家綠悠能源轉換至E-dReg，價格連續199天維持0元，平",
  },
  "2": {
    "sectionTitle": "📊 市場最新動態",
    "sectionContent": "📊 市場最新動態\nPOXA Energy整理台電電力交易平台公開資料，追蹤每週市場變化的趨勢，完整資料可以參考台電",
  },
  "3": {
    "sectionTitle": "💡 台電電力供需資料",
    "sectionContent": "💡 台電電力供需資料\nPOXA想藉由監測再生能源的供給跟佔比情況，來了解其對輔助服務市場的影響，預期可以分析",
  },
  "4": {
    "sectionTitle": "📰 資訊來源網站",
    "sectionContent": "📰 資訊來源網站\n電力交易市場\n輔助服務市場，https://etp.taipower.com.tw\n輸配電等級儲能專區（可以看見目",
  },
}
```

- WebInformation.synonyms
 - 用於判別同義詞，以提升 ETP (Exchange-Traded Products)和 QA 問答的答題準確率

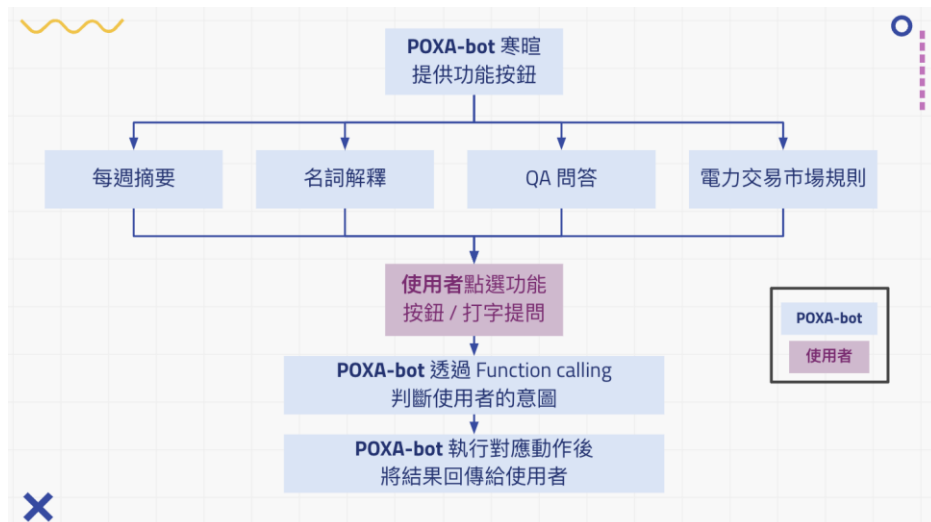
```
_id: ObjectId('66f053597285e7df57077c7f')
term: "dReg"
vocabulary: "調頻服務"
```

- PDFDocument.definitions
 - 用於儲存電力交易相關之專有名詞及其定義

`_id: ObjectId('66d93c2219a0b57411ce8f23')`
`term: "參與容量"`
`definition: "於日前輔助服務市場係指合格交易者之交易資源已完成註冊登記程序之容量；於備用容量市場係指合格交易者之交易資源參與交易媒合之容量"`

(3) POXAGPT Controller 之設計與實作

A. 對話流程設計



- 機器人會打招呼、介紹自己能提供的功能。所有功能包含 **每週摘要**、**名詞解釋**、**QA 問答**、**電力交易市場規則**。
- 使用者可以透過點選按鈕、打字向機器人提問。
- 機器人會透過 Function Calling 來判斷使用者的意圖。
- 機器人確認意圖（想要執行的功能）後，會執行對應的動作，再回覆給使用者。
每個功能如何執行的細節請參考其他章節。

B. 摘要功能之設計與實作（基於 Blog 資料與 Function Calling）



- **請 GPT 拿到跟時間有關的描述**：當 Function Calling 判斷為該意圖時，會先請 OpenAI 拿到跟時間有的描述。其中，「若未給定...，請...」，這句說明非常重要！

第一次 Prompt 的內容如下：

```

"type": "string",
"description": f""跟時間有關的描述，不要推測使用者未提供的數據。
若有未給定年 or 月，請使用與「{datetime.today().strftime('%Y/%m/%d')}」對應的數字。
若未給定日期，請保持空白。
""

```


- 請 GPT 將其轉換為確切日期

當拿到描述後，會再請 Open AI 將描述轉換成確切的日期！下面 Prompt 的描述都是必要的，因 Open AI 對日期的敏感程度極低，因此需要透過描述來提高正確率。

第二次 Prompt 的內容如下：

```
{"role": "system", "content": f"""  
你是一個日期轉換工具，只會輸出八位數字（%Y%m%d），請不要輸出除了數字之外的東西  
若未提供年份，請使用{today}對應的年份  
若未提供年份與月份，請使用{today}對應的年份與月份  
若未提供日期，但有給定第n週，請使用7*n  
若未提供任何能推測日期之描述，請使用 6  
.....  
},
```

補充：最原始的設計，其實是將「拿到時間描述」、「轉換成確切日期」放在同一個 Prompt 中，但發現非常容易出錯，所以才拆成兩層。

C. 電力市場 QA 功能設計（基於 Blog 資料與 Function Calling）

- 一般問題

1. 問題分類：可分成數據型、敘述型(事實性、意見性或推理性問題)。
2. 問題的時間點判斷：需判斷問題是否想要最新的資訊，如果是詢問當前或目前等時間點，就從資料庫拿取最新的文章資訊。
3. 問題回答

- 數據型問題：執行 bert 模型，去計算問題與每篇文章的 cosine 相似度，找到最高分的文章而後去使用 gpt 根據所匹配的文章去回答問題。
- 敘述型問題：先對問題進行關鍵字分析(使用 gpt 對問題本身分析關鍵字)，根據分析出的關鍵字去找尋匹配的文章，再依據匹配出的文章去回答問題。

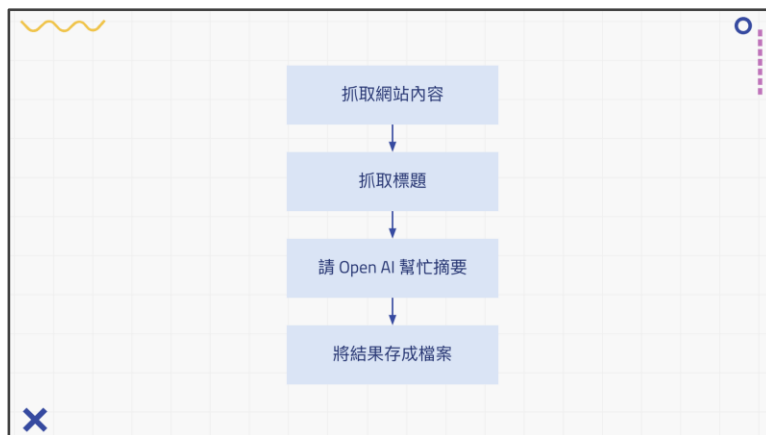
- ETP (Exchange-Traded Products)問題

1. 問題分析：對問題進行分析，決定欲查詢元素的前、後綴。
 - 字首：當問題提到 E-dReg，prefix=edreg；提到調頻備轉，prefix=reg；提到即時備轉，prefix=sr；提到補充備轉，prefix=sup。
 - 字中：midfix=Offering，當問題提到得標量，midfix=Bid；提到非交易，midfix=BidNontrade；提到結清價格，midfix=Price。
 - 字尾：當問題中提到民營，suffix=Qse。若未提到則無 suffix。
2. 檔案讀取判斷：根據分析出的元素決定讀取哪份檔案的資料。
(poxa-info.etp_offering.json & poxa-info.etp_settle_value_query.json)
3. 依時間搜尋資料：依據查詢時的日期去尋找最近的資料內容。

D. 能源政策 QA 功能設計

1. **建立 Vector Store**：建立 Open AI Vector Store，將上傳的檔案以向量方式儲存，用來存放問答功能需要的背景知識。
2. **上傳檔案**：將電力交易市場規則相關資料文件上傳到已建立的 Vector Store。
3. **建立 Assistant**：建立 Open AI Assistant，讓 Assistant 使用已上傳至 Vector Store 的資料，並設定成使用資料作為背景知識的問答助手。
4. **File Search**：當使用者提出問題時，Assistant 利用 File Search 功能找到適合回答問題的文件(目前限定為三份)，並分別以三份不同文件提供三種答案給使用者。

(4) POXA 自動摘要功能



1. **抓取網站內容**：使用 **selenium** 並透過 CSS Selector 抓取網站內除「本週摘要」區塊外的所有 p 標籤內的純文字，其中「市場最新動態」還需要多一步驟：使用 selenium 模擬使用者點選按鈕，使得網頁元素能夠動態加載，加載完畢後，再抓取 p 標籤內的純文字。
2. **抓取標題**：抓取所有 **h2** 標籤當作摘要結果的**標題**，抓取所有 **h3** 標籤當作摘要結果的**子標題**。
3. **請 Open AI 幫忙摘要**：首先，需要準備一份歷史摘要，當作範例給 Open AI 參考。接著，將範例、上述兩步驟的資料統整並寫入 Prompt，Prompt 的實際內容如下，主要包含四部分：
 - **定義 Open AI 的身份**（例如：您是一個重點摘要的助手）
 - **提醒事項**（例如：每個子標題包含...）
 - **預期的輸出格式**
 - **實際的數據**

```
"content": f"""
您是一個直到獲取所有資訊後，才摘要重點的助手，會按照期望的輸出格式給予回覆，請依照以下標題 {h2_titles} 進行摘要。
{temp}
其中「市場最新動態」還須包含四個子標題「調頻服務」、「E-dReg」、「即時備轉」、「補充備轉」，每個子標題需包含以下內容：
「平均結清價格(required)」、「平均結清價格較上週上升或下滑多少(required)」、「本週參與容量(required)」、「
參與容量較上週上升或下滑多少(required)」、「來自哪裡(optional)」，請用一句話描述，不需要分段。
其中「台電電力供需資料」一定會提到「再生能源占比(required)」、「滲透率(required)」。
```

期望的輸出格式如下（它是過去的歷史資料，這只是給你參考輸出的格式，並非實際的數據，請不要參考其中的數據內容）：

```
{output_sample}。
```

而實際的數據如下：

```
{plain_text}
"""
```

4. 將結果存成檔案：部分結果如下圖所示：

```
1  [📰 台電最新公告] (#📰-台電最新公告)
2
3  - 本週台電沒有公告，POXA會持續追蹤最新公告。
4
5  [🔴 本週主題分析] (#🔴-本週主題分析)
6
7  - 本週**調頻備轉**擴增至784.7MW，價格連續234天維持0元，平均得標率降至63.7%，持續影響整體收益；**E-dReg**本週維持
8
9  [📊 市場最新動態] (#📊-市場最新動態)
10
11 - 「調頻服務」平均結清價格維持在0元/Mwh，參與容量784.7MW，較上週**增加1.9MW**，來自倍捷能源科技有限公司；
12 - 「E-dReg」平均結清價格維持500元/Mwh，參與容量341.6MW，無新增合格交易者；
13 - 「即時備轉」平均結清價格從247.15元/Mwh下滑至243.21元/Mwh，**下滑1.59%**，參與容量140.1MW，無新增合格交易者；
14 - 「補充備轉」平均結清價格從209.51元/Mwh上升至214.43元/Mwh，**上升2.35%**，參與容量333.1MW，無新增合格交易者。
15
16 [📊 台電電力供需資料] (#📊-台電電力供需資料)
17
18 - 本週再生能源佔比平均約9.1%，滲透率最高達23.47%，較上週有所減少，主要因風力發電的變化影響。
19
20 [poxa-energy-專屬分析] (#poxa-energy-專屬分析)
21
22 - POXA Energy將持續分析電網頻率及供需變化對dReg/sReg儲能充放電排程的影響，並提供最新的市場動態。
```

實作時，觀察到一些給定 Prompt 的技巧如下：

- **不要先統整好所有資訊包含哪些!!**

容易出錯的 Prompt：接下來的資料會包含: 1 輸出範例 2 實際數據，...

不易出錯的 Prompt：輸出範例如下: ... ，而實際數據如下: ... 。

- **先後順序非常重要!!**

例如：先告訴 Open AI 要做的事，再給定實際資料、先給定參考格式，再給定實際資料，若順序顛倒，輸出的資料容易被參考格式之資料干擾。

- **描述的方式，盡量像程式語言一樣!!**

例如：哪些輸出的資訊是 required，哪些是 optional。

4. 目標達成情形

規劃交付日期	規劃交付項目	達成情形
第一期 (2024/9/30 日前)	● 基本功能程式碼與設計文件。	已完成基本功能雛形，包含使用者聊天/問答介面 (POXAGPT Front-End UI)、問答控制器 (POXAGPT Controller)、資料處理和儲存模組 (POXAGPT Data Manager)。
第二期 (2024/12/31 日前)	● 核心功能程式碼與設計文件。	已完成資料轉換工具。
第三期 (2025/2/28 日前)	● 服務管理與進階功能程式碼與設計文件。 ● 系統測試報告、軟體安裝文件與使用說明文件。	尚未開始執行。