

TaiSoc

ATK【繁體中文AI開源實踐專案】



Agenda

01 專案簡介

02 學期進度

- 前端
- 後端

03 未來展望



TaiSoc

專門回答台灣社會人文與地理知識問題的語言模型



問題定義

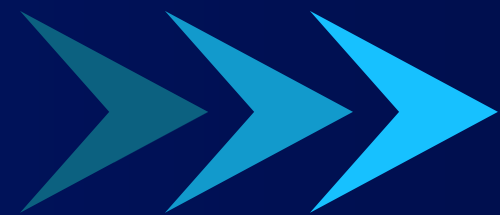
現有主流語言模型多以全球化或通用性為設計目標



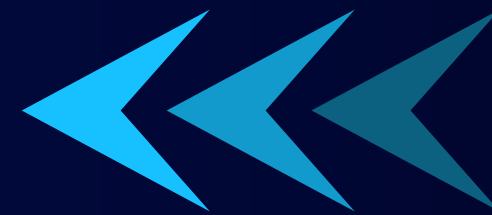
問題定義

模型	回覆準確率	回覆平均信心值
GPT-4o	70.35%	75.11%
Claude	53.76%	85.44%
Gemini	32.68%	83.34%

Taiwan



TaiSoc



Social

目標客群

產品定位於專業與通用之間，涵蓋廣泛客群

01

教育與學術機構

- 教師與學生
- 研究人員

02

公共部門與政策制定者

- 政府機構
- 非營利組織

03

商業與文化產業

- 旅遊業者
- 文創與出版業

04

媒體與新聞業

- 記者與編輯

05

普通消費者

- 本國居民
- 外國人士

前端介面設計-----Streamlit

1. 安全性設計：

- 使用者需登入，透過 SHA-256 加密驗證密碼，確保帳戶資料安全。
- API 請求通過 Bearer Token (API Key) 進行授權，防止未經授權的存取。

```
USER_DB = {  
    "admin": "8c6976e5b5410415bde908bd4dee15dfb167a9c873fc4bb8a81f6f2ab448a918"  
}  
  
def hash_password(password: str) -> str:  
    return hashlib.sha256(password.encode()).hexdigest()  
  
def authenticate(username: str, password: str) -> bool:  
    return username in USER_DB and USER_DB[username] == hash_password(password)
```

2. 介面設計：

- 使用 Streamlit 建構，用戶可以直接在聊天介面輸入問題，查看聊天記錄和即時回應，界面簡單直觀。

```
def render_chat() -> None:  
    st.title("🇹🇼 台灣知識問答")  
  
    # Initialize chat history  
    if 'chat_history' not in st.session_state:  
        st.session_state.chat_history = []  
  
    # Display chat history  
    for msg in st.session_state.chat_history:  
        with st.chat_message(msg["role"]):  
            st.markdown(msg["content"])  
  
    # Handle new messages  
    if prompt := st.chat_input("請輸入問題"):  
        user_msg = {"role": "user", "content": prompt}  
        st.session_state.chat_history.append(user_msg)  
  
        with st.chat_message("user"):  
            st.markdown(prompt)  
  
        with st.chat_message("assistant"):  
            with st.spinner("思考中..."):  
                response = call_rag_api(prompt, st.session_state.chat_history)  
                st.markdown(response)  
                st.session_state.chat_history.append(  
                    {"role": "assistant", "content": response}  
                )
```


前端介面設計-----即時通信

3. 回答生成流程：

- 問題和聊天歷史會以 JSON 格式發送到後端 RESTful API。
- 後端的 AI 模型根據上下文理解問題並生成回應，返回給前端顯示。

4. 即時通信與錯誤處理：

- 系統透過 HTTP 請求與後端互動，並對超時或無效回應進行錯誤捕捉，確保穩定性。

```
# Configuration
API_CONFIG = {
    "url": "http://localhost:3000",
    "key": "AM5MX0A-PR14GY0-JJW"
}
```

```
def call_rag_api(message: str, history: List[Dict] = None) -> str:
    try:
        headers = {
            "accept": "application/json",
            "Authorization": f"Bearer {API_CONFIG['key']}",
            "Content-Type": "application/json"
        }

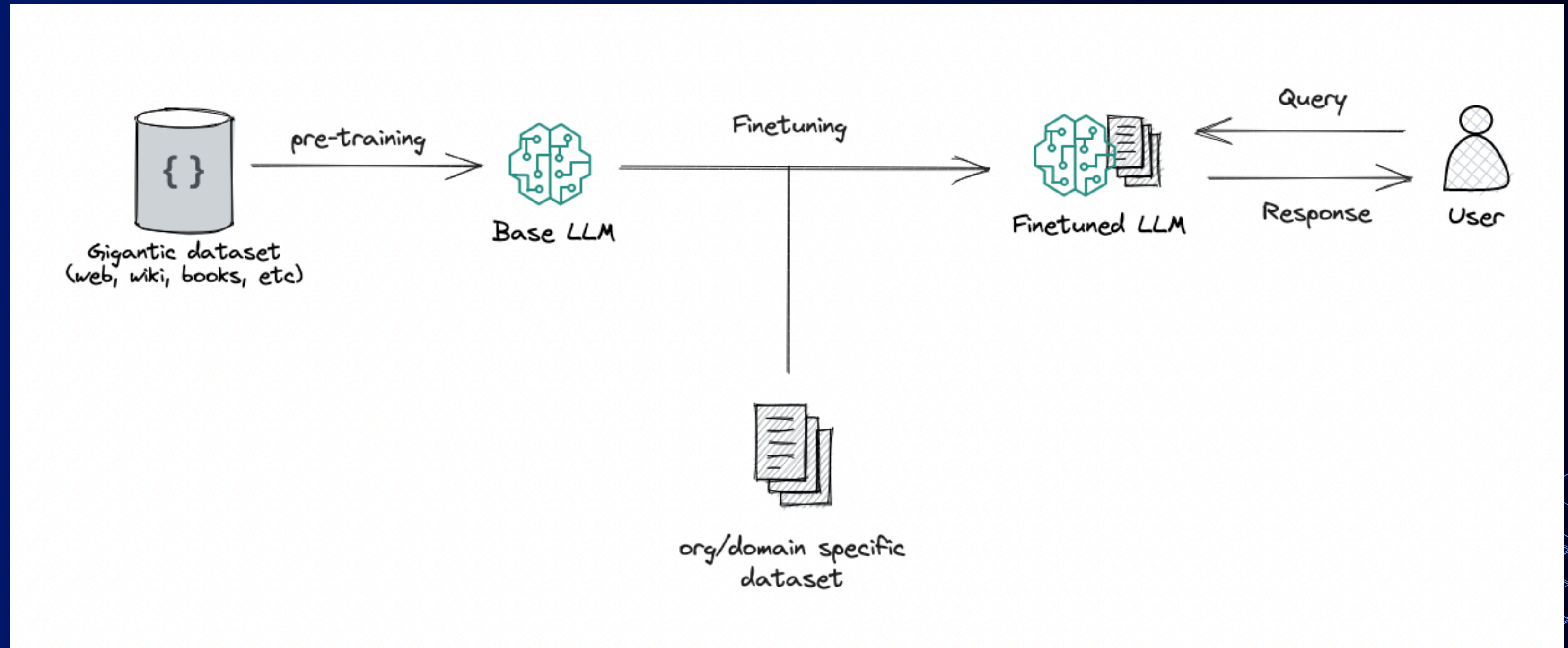
        payload = {
            "message": message,
            "mode": "query",
            "history": history or []
        }

        response = requests.post(API_CONFIG['url'], json=payload, headers=headers)
        response.raise_for_status()

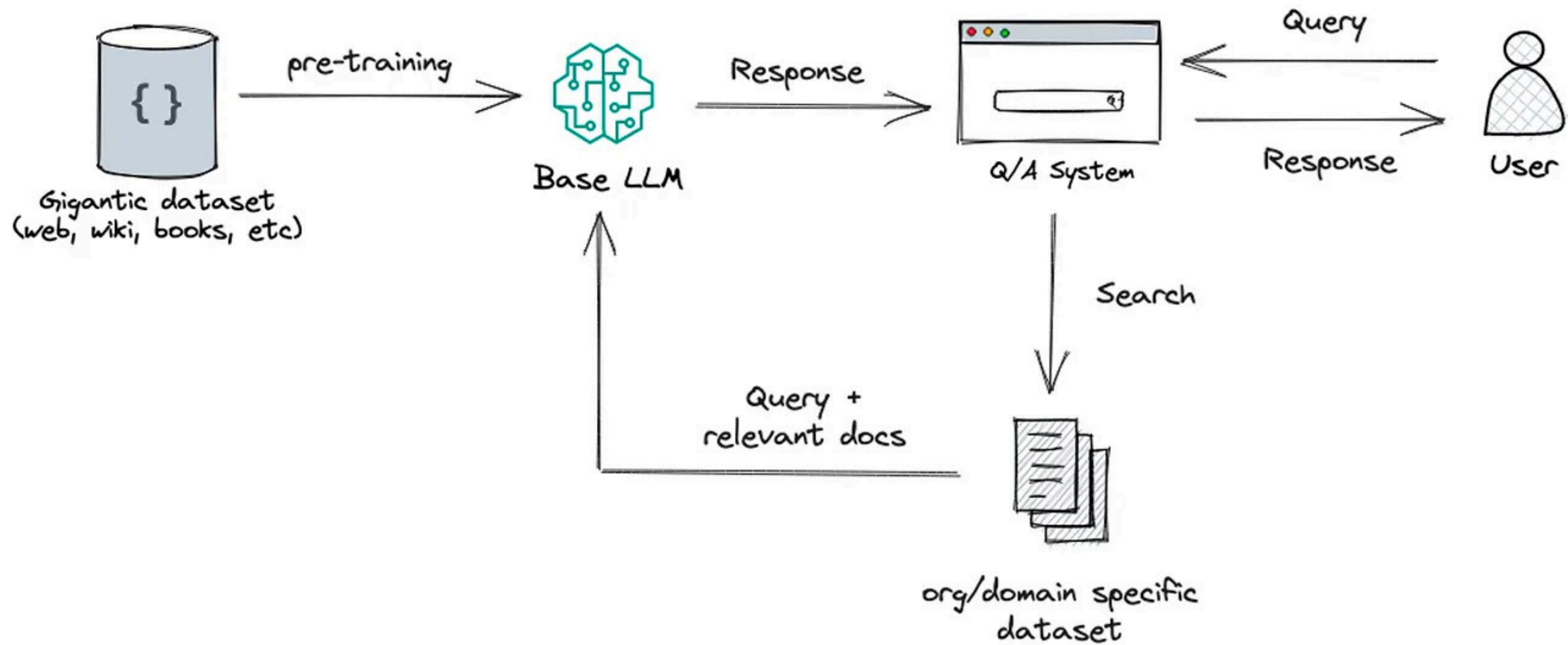
        data = response.json()
        return data.get("textResponse", "無法獲取回應")

    except Exception as e:
        return f"錯誤: {str(e)}"
```

後端-----Finetuning 原理



後端-----RAG 原理



後端-----實操



我國護照共有145個地方免簽證可進入，2022年7月更榮獲「全球護照指數」全球第34、亞洲排行第5。但是世界上有一個國家「喬治亞」，由於該國內部政治情勢，以及來自中國壓力等原因，明文禁止持中華民國護照者入境，不過此舉仍無法阻止熱愛旅遊的背包客們不惜冒著被遣返的風險前往這個神秘的國度。下列敘述何者正確？A.我國護照遭喬治雅拒絕入境的事件，顯示出在中共打壓下，我國的主權無法受到喬治雅政府的承認B.我國外交不積極爭取出國免簽證，係因為免簽證待遇不只象徵我國合法主權，更可增加國人出國便利C.我國背包客為了到喬治亞旅遊，改持中國旅行證或歐盟居留證試圖進入，此舉有國家認同混淆疑慮D.我國政府因應喬治亞政府不友善，可以透過體育競賽和民間文化交流的方式，建立我國之對內主權



選項C是正確的。它提到的是背包客們為了到喬治亞旅遊，改持中國旅行證或歐盟居留證試圖進入，這個舉動有國家認同混淆疑慮的問題。



選項C: 我國背包客為了到喬治亞旅遊，改持中國旅行證或歐盟居留證試圖進入，此舉有國家認同混淆疑慮



A、B



A、B、C



後端-----實操



我國護照共有145個地方免簽證可進入，2022年7月更榮獲「全球護照指數」全球第34、亞洲排行第5。但是世界上有一個國家「喬治亞」，由於該國內部政治情勢，以及來自中國壓力等原因，明文禁止持中華民國護照者入境，不過此舉仍無法阻止熱愛旅遊的背包客們不惜冒著被遣返的風險前往這個神秘的國度。下列敘述何者正確？A.我國護照遭喬治雅拒絕入境的事件，顯示出在中共打壓下，我國的主權無法受到喬治雅政府的承認B.我國外交不積極爭取出國免簽證，係因為免簽待遇不只象徵我國合法主權，更可增加國人出國便利C.我國背包客為了到喬治亞旅遊，改持中國旅行證或歐盟居留證試圖進入，此舉有國家認同混淆疑慮D.我國政府因應喬治亞政府不友善，可以透過體育競賽和民間文化交流的方式，建立我國之對內主權



C. 我國背包客為了到喬治亞旅遊，改持中國旅行證或歐盟居留證試圖進入，此舉有國家認同混淆疑慮



A、B



正確答案是B。根據文中所述，因為免簽待遇不只象徵我國合法主權，更可增加國人出國便利，所以我國外交不積極爭取出國免簽證。



正確答案是C：我國背包客為了到喬治亞旅遊，改持中國旅行證或歐盟居留證試圖進入，此舉有國家認同混淆疑慮。這個選項敘述的是事實上發生的情況，描述了因為喬治亞對中華民國護照的限制而引起的問題和挑戰。



後端-----llama 3.2 回覆比較

我國護照共有 145 個地方免簽證可進入，2022 年 7 月更榮獲「全球護照指數」全球第 34、亞洲排行第 5。但是世界上有一個國家「喬治亞」，由於該國內部政治情勢，以及來自中國壓力等原因，明文禁止持中華民國護照者入境，不過此舉仍無法阻止熱愛旅遊的背包客們不惜冒著被遣返的風險前往這個神秘的國度。下列敘述何者正確？ A.我國護照遭喬治雅拒絕入境的事件，顯示出在中共打壓下，我國的主權無法受到喬治雅政府的承認 B.我國外交不積極爭取出國免簽證，係因為免簽待遇不只象徵我國合法主權，更可增加國人出國便利 C.我國背包客為了到喬治亞旅遊，改持中國旅行證或歐盟居留證試圖進入，此舉有國家認同混淆疑慮 D.我國政府因應喬治亞政府不友善，可以透過體育競賽和民間文化交流的方式，建立我國之對內主權

Q | llama3.2:latest

答案：C

解釋：根據文章所述，我們可以知道以下幾點：

- 145 個地方免簽證可進入：指的是各國的旅行政策規定。
- 「全球護照指數」全球第 34、亞洲排行第 5：指的是我國護照在全球及亞洲等地的 Ranking。

接下來分析每個選項：

- A. 因為喬治雅政府不友善，我國主權無法受到承認：雖然文章中提到喬治雅政府對我國民眾不予友善，但這並不是因為我國主權無法受到承認。
- B. 我國外交不積極爭取出國免簽證：根據文獻，雖然有多種原因影響各國的旅行政策規定，但是文章中並沒有提到該原因是「免簽待遇不只象徵我國合法主權，更可增加國人出國便利」。
- C. 我國背包客為了到喬治亞旅遊，改持中國旅行證或歐盟居留證試圖進入：這是正確的。根據文章所述，我國民眾不受喬治雅政府禁止，我國民眾仍然可以通過改用其他Travel Documents進行旅行。
- D. 我國政府因應喬治亞政府不友善，可以透過體育競賽和民間文化交流的方式，建立我國之對內主權：根據文章所述，我國政府因應喬治雅政府不友善，並不會透過體育競賽和民間文化交流的方式來維護我國之對外主權。

✎ □ 🔍 🌐 📄 🗨️ 🔄

後端-----實操



在國立台北大學宿舍訪客時間內，未經登記，於宿舍內容留訪客；訪客為他棟宿生，亦同要扣多少點數？



扣6~10點



在國立台北大學宿舍破壞環境整齊、清潔然後又在國立台北大學宿舍使用公共設備（施）未立即還原總共要扣多少點



扣6~10點



在國立台北大學宿舍破壞環境整齊、清潔然後又在國立台北大學宿舍使用公共設備（施）未立即還原然後又在國立台北大學宿舍未將機踏車輛停放於規定區域內總共要扣多少點



扣11~15點



後端-----總結

- 問題點

RAG	資料集為選擇題時，出現偏誤
Finetuning	資料集為開放式問題時，出現偏誤

- 解方

1. 藉由生成式AI產生類似問題的同樣答案
2. 將原有選項拆分為不同的論述

登入系統

歡迎使用台灣知識問答系統

使用者名稱

admin

密碼

..... Press Enter to submit form

登入

你的密碼強度：弱

使用建議的強式密碼

rjGLyyTF_NkUmc5 重新整理

將會儲存您的密碼並將裝置填滿。管理您的密碼 設定

未來展望



- 優化

前端：介面優化

後端：雲端操作

- 新發想

因應實習需求增加，實習資訊分散

嘗試整合實習資訊於同一平台，方便
學生應徵及企業徵才