## 技術架構與工具

* Python、Flask、Jinja2 模板引擎
* Selenium + BeautifulSoup 爬蟲技術
* jieba、pandas 中文文本處理與統計
* Bootstrap 5、Chart.js 前端框架與圖表套件
* OpenRouter (OpenAI 兼容接口) 串接 AI 生成內容

## 核心功能介紹

1. **Google Maps 自動評論爬取**：
   * Selenium 自動模擬瀏覽器，進入指定店家頁面。
   * 自動滾動評論區、點擊更多評論，直到擷取完成。
   * 過濾評論日期區間，控制資料品質。
2. **中文關鍵字統計分析**：
   * 使用 jieba 斷詞，處理中文文本。
   * pandas 統計詞頻，生成 Top 20 高頻關鍵字清單。
   * 結果透過 Chart.js 長條圖直覺呈現。
3. **AI 智能評論輔助分析**：
   * 使用者可針對評論提出問題，例如：「請總結優缺點」、「請判斷評價趨勢」。
   * 系統整合 OpenRouter (GPT 模型)，以中文專業回覆，模擬商業分析輔助工具。
4. **完整前後端流程展示**：
   * 表單輸入 → 爬蟲分析 → 結果視覺化 → AI 互動，完整體現後端主導開發流程。

# 🔄 專案完整流程

## 1. 使用者操作流程（從前端角度）

### (1) 首頁輸入資料

使用者打開網站 http://localhost:5000，看到一個表單：  
✅ 輸入 Google Maps 店家網址  
✅ 選擇起始年份（例如 2023）  
✅ 選擇結束年份（例如 2024，兩個年份必須相差 1 年）  
✅ 按下「開始分析」按鈕

### (2) 等待系統自動爬蟲

前端顯示等待，後端開始背景執行 Selenium，模擬瀏覽器自動操作：

## 2. 後端資料流程（Flask + 爬蟲 + 分析）

### (1) 接收到表單資料

Flask app.py 的 / 路由接收到 POST 請求：

* 解析網址、年份資料
* 確認格式沒問題（簡單驗證）

### (2) 啟動 Selenium 自動化爬蟲

呼叫 modules/scraper\_selenium.py 的 fetch\_google\_maps\_reviews 函式：

流程細節：  
✅ 開啟 Chrome 瀏覽器（可設 headless 模式）  
✅ 導航至 Google Maps 網址  
✅ 自動模擬點擊「評論」分頁  
✅ 自動滾動、點擊「更多評論」，直到撈滿指定評論數量  
✅ 萃取評論資料：作者、星等、時間、評論內容  
✅ 篩選符合指定年份的評論  
✅ 結果整理成 Python List 傳回

### (3) 關鍵字統計分析

呼叫 modules/analysis.py 的 keyword\_stats 函式：  
✅ 把所有評論文字合併  
✅ 使用 jieba 斷詞（中文分詞工具）  
✅ 統計詞頻，取得前 20 名高頻關鍵字  
✅ 轉換成 DataFrame 結構，方便前端顯示

### (4) 組合資料傳給前端

* 原始評論列表
* 星等統計（各星等數量）
* 關鍵字統計結果

傳給 results.html，進入結果頁面。

## 3. 結果頁前端展示（results.html）

### (1) 顯示以下資訊：

✅ 所有評論清單（作者、星等、時間、內容）  
✅ 各星等數量統計  
✅ 前 20 高頻關鍵字長條圖（Chart.js 繪製）

### (2) 提供 AI 問答功能

底部有輸入框：  
✅ 使用者可針對評論內容提出自訂問題（如「請幫我總結優缺點」）  
✅ 隱藏欄位內含所有評論資料（或後續優化為只傳編號）

## 4. AI 輔助分析流程（/ask 路由）

### (1) 接收到使用者提問

Flask 端解析：  
✅ 把評論資料與問題一起組成聊天上下文  
✅ 呼叫 OpenRouter API（GPT 相容模型），送出請求

### (2) AI 回傳結果

* 模型分析評論資料
* 用繁體中文回覆使用者的問題
* 回傳結果顯示於 answer.html

## 5. 繼續提問

在 answer.html 頁面：  
✅ 顯示 AI 回答結果  
✅ 保留評論資料（或優化為 session 快取）  
✅ 使用者可持續輸入新問題，系統再次呼叫 AI 回答

# 🔧 補充說明

* 爬蟲可自由調整滾動次數、延遲秒數，控制資料量
* AI 模型透過 OpenRouter，避免直接用 OpenAI 的 token 限制
* 前端使用 Bootstrap 統一介面風格，Chart.js 繪製圖表
* 爬蟲中已簡單處理 Google Maps DOM 結構變動，但仍需持續維護

# 📝 簡單總結版（面試時可講）

1. 使用者輸入 Google Maps 網址與年份
2. 系統自動用 Selenium 爬評論
3. 做中文分詞與關鍵字統計
4. 結果頁展示評論、星等、關鍵字圖表
5. 提供 AI 問答功能，進一步幫助評論理解

您好，這個專案是我獨立設計開發的，主要是用來展示我後端開發能力、資料分析能力，以及串接 AI 應用的整合實戰作品，專案名稱叫做「Google Maps 評論分析系統」。

整體來說，這個系統可以自動幫助使用者抓取指定店家在 Google Maps 上的公開評論，並針對這些評論做關鍵字統計、星等分佈分析，最後還提供 AI 輔助，讓使用者可以直接針對評論提出進一步的問題，像是總結優缺點、判斷評論趨勢等等。

## 技術面我主要用到：

* **後端**：Flask 負責整個後端流程控制
* **爬蟲**：Selenium + BeautifulSoup 自動操作 Google Maps 抓取評論
* **資料分析**：jieba 中文斷詞、pandas 做關鍵字統計
* **前端**：Bootstrap 統一介面，Chart.js 畫圖表
* **AI 整合**：OpenRouter（GPT 相容模型）回答自訂問題

## 🛠️ 整個操作流程：

1. 使用者輸入 Google Maps 店家網址與年份區間
2. 系統自動啟動 Selenium 爬蟲，模擬人工操作，把評論資料抓下來
3. 篩選符合指定年份的評論，避免資料太雜
4. 做中文分詞、關鍵字統計，分析最常見的詞
5. 結果頁顯示所有評論、星等分佈，以及關鍵字長條圖
6. 使用者可以針對評論內容，直接輸入問題，系統會呼叫 AI 幫忙回答
7. 可以無限次繼續提問，整個分析流程不需要重新輸入資料

## 🌟 這個專案的亮點：

* 從資料抓取到分析、前端展示，完整流程自己開發
* 結合爬蟲與實際資料來源，模擬真實商業場景
* 中文文字處理完整，非單純英文 demo
* AI 功能不只是玩具，是實際輔助商業判斷用
* 程式碼結構清晰，前後端資料流設計有考量效能與安全性

這個專案的目的是模擬實務情境下，怎麼結合後端、資料分析跟 AI 技術來解決實際問題，也希望透過這個作品讓面試官了解我的技術能力與思考邏輯。

這個專案主要是我想展示我自己整合能力，包含後端 Flask 開發、爬蟲、資料處理、AI 串接的實際應用。

它的功能就是，使用者只要輸入 Google Maps 上面店家的網址，選一個年份區間，系統就會自動去抓這家店的公開評論，然後統計裡面最常出現的關鍵字、星等分佈，還可以直接用 AI 去幫忙回答問題。

技術上我用 Selenium 自動操作 Google Maps、jieba 做中文分詞、pandas 算關鍵字，前端用 Bootstrap 跟 Chart.js 畫圖表，然後最後用 OpenRouter 去串 GPT 模型，讓 AI 幫忙總結評論內容或分析趨勢。

可能被問的問題:

### 爬蟲相關

* Selenium 為什麼用這個，而不是 requests 之類的？
* 爬 Google Maps 有遇到什麼反爬蟲問題嗎？怎麼處理？
* Google Maps 的 DOM 結構變動怎麼因應？
* headless 跟非 headless 模式差別？

### 資料處理相關

* jieba 斷詞準確嗎？有沒有處理過亂碼或英文？
* 停用詞是怎麼設計或規劃的？
* 關鍵字統計如果資料很多，效能會不會有問題？

### AI 串接相關

* OpenRouter 跟 OpenAI 的差別是什麼？為什麼用這個？
* AI 回答有沒有做安全性過濾或錯誤處理？
* prompt 設計有沒有注意上下文或繁體中文？

# 系統設計與架構類

考官會看你有沒有整體思考：

* 資料量變多時，JSON 傳遞會有什麼風險？怎麼優化？
* 如果多人同時使用，session 或快取資料怎麼設計？
* Flask 如果要改成 API 服務給前端 Vue 或 React 用，會怎麼改？
* 為什麼不用資料庫？未來擴充怎麼考慮？
* 有沒有想過用 Docker 打包，部署上雲端？

這類不看程式碼，看你邏輯與思考能力：

* 這個系統實際商業價值是什麼？
* 如果 Google Maps 改版導致爬蟲失效，你會怎麼處理？
* AI 給錯誤回答，使用者抱怨，該怎麼改進？
* 你認為這個專案還有哪些缺點？
* 除了 Google Maps，你有沒有想過其他資料來源？

如果面試官比較資深，可能額外追問：

* Selenium 有沒有做過效能優化（像隱藏圖片、減少不必要載入）？
* 前端 Chart.js 可不可以做互動式功能（例如點擊某個關鍵字顯示對應評論）？
* AI 回答如果被濫用，怎麼做權限或使用次數限制？
* 有沒有做過自動化測試？Flask 有沒有寫 pytest 測試過？