

銘傳大學

人工智慧應用學系

專題研究初審報告

個人化 AI 助理系統研發

指導教授：葉生正 老師

專研學生：白東衢 許友銓 蔡承哲 林靄君 李宇芯

中 華 民 國 一 一 四 年 四 月 二 十 八 日

摘要

本研究使用 AI Agent 技術開發個人 AI 助理系統。結合行動裝置或穿戴式裝置如：手機、筆記型電腦、智慧型手錶和智慧眼鏡等設備，不侷限於單一裝置上，於日常生活中即時與使用者互動。有別於一般聊天機器人，此系統強調「個人化」且結合各種 API，根據使用者狀態、習慣或健康狀況給予建議及提醒。舉例來說，提醒行程規劃，讀取健康裝置數據給出提醒等等，在應用上更加靈活聰明。

本研究使用 Github、Hugging Face 作為開源平台，Github 儲存管理前後端程式碼、API 模組整合，Mongo DB 管理資料、帳密、對話紀錄，語音辨識系統做個人身分和聲音辨識，加入 Hugging Face 華語情緒辨識模型辨識與使用者互動時情緒狀態，Meta Glasses 語音轉文字與文字轉語音應用，Apple Watch 抓取健康資訊透過藍牙回傳。期望在此系統成為使用者日常生活中隨問隨答的貼心助理。

關鍵詞：Apple Watch、Github、Hugging Face、Meta Glasses

目錄

摘要.....	i
目錄.....	ii
表目錄.....	iii
圖目錄.....	iv
第一章 緒論.....	1
1.1 論文主體.....	1
1.2 圖表規範.....	2
1.3 參考文獻規範.....	3
參考文獻.....	10
附錄一 工作分配.....	11
附錄二 甘特圖.....	12

表目錄

表 2-1 傳統語音助理和個人化語音助理的比較

表 A-1 工作分配表

圖 1-1 ios 天氣 app 介面

圖 1-2 ios 行事曆 app 介面

圖 1-3 ios 健康 app 介面

圖 1-4 apple watch

圖 1-5 meta glasses

圖目錄

圖 2-1 組員四國語言音檔

圖 2-2 初始前處理畫面

圖 1-1 ios 天氣 app 介面

圖 1-2 ios 行事曆 app 介面

圖 1-3 ios 健康 app 介面

圖 1-4 apple watch

圖 1-5 meta glasses

圖 2-1 組員四國語言音檔

圖 2-2 初始前處理畫面

圖 2-3 前處理完成音檔

圖 2-4 語音辨識系統畫面

圖 2-5 Acc

圖 2-6 Loss

第一章 緒論

1.1 論文主體

頁次請直接使用 word 軟體直接產生，摘要至圖表目錄，以 i ii iii,... 等大寫羅馬數字連續編頁；論文第一章至附錄，均以 1,2,3,...等阿拉伯數字連續編頁。

章次編號使用一、二、...等數字；節次編號使用阿拉伯數字，且需搭配章次編號，例：第一章中的第二節其編號應為 1.2；段次編號使用阿拉伯數字，且需搭配章次及節次編號，例：第三章第二節第一段其編號為 3.2.1；非章節之標題而又需條列敘述時，可不依標題層次之限制，只需前後使用一致之序號即可。章次、章名稱位於打字版面頂端中央處。節次、段次均自版面左端排起。小段以下等號次及名稱，均以首行縮排兩字元表明層次。

中文為標楷體字型；英文為 Times New Roman 字型，標準字距，中英文皆採 1.5 倍行距。數學公式依出現次序連續編式號，並加()號標明於文中或公式後，如式(1)。

如有特殊事項論點等，可使用註腳(Footnote)說明。註腳依應用順序編號，編號標於相關文右上角以備參閱。註腳號碼及內容繕於同頁底端

版面內，與正文之間加劃橫線區隔，頁面不足可延用次頁底端版面。各章節內所使用之數學及特殊符號，均集中表列說明，以便參閱，表內各符號不須編號。

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x - \mu)^2} \quad (1)$$

1.2 圖表規範

圖、表應置於正文內，表號及表名置於表正上方，圖號及圖名置於圖正下方，並以阿拉伯數字區分不同之圖、表。圖、表資料來源及說明，或是內容需要（如表中之符號）作簡要說明時，請置於圖、表下方。參考結果如表 1-1 及圖 1-1 所示。



圖 1-1 人智系

表 1-1 圖表規範

	表	圖
標題	位於表格上方	位於圖片下方
樣式	套用「表標題」	套用「圖標題」

1.3 參考文獻規範

引用相關文獻時，請參照 IEEE 或 APA 格式撰寫，此論文以 IEEE 格式為例（APA 請參閱系網規範說明）。

於引用文句句尾處，按照內文出現順序標上[1]、[2]、[3]之標籤。本文結束後，在參考文獻列表處按照[1]、[2]等之順序標示清楚。易生混淆之單位，每次均用全名；簡單且廣為人知的單位，在內文中第二次出現後可用縮寫或簡稱，例如：APEC、WWW。範例請參照下章參考文獻。

第一章 研究動機與發想

1.1 研究動機與目的

作為一位單身且獨居大學生，日常瑣事往往都只能靠自己打理，生活能力再怎麼強，也是有疏忽或需要他人幫助的時候。為了更進一步的提升自己的生活品質和做事效率，於是有了開發結合物件的個人化 AI 助理的想法，要是能將此系統結合至眼鏡、手錶等物件，便能強化生活的便利性。

Apple Siri、Google Assistant 等智慧語音助理，這類語音助理通常用於手機和車載系統中，像是 Apple Siri 為蘋果設備內建（iPhone、Mac、Apple Watch），只有在擁有特定裝置的情況下，才可以使用，因此本團隊希望可以開發出一款沒有平台限制的跨設備整合 AI 助理。

現今的 AI 助理與使用者仍然偏向運用文字和語音居多，本團隊期許多加善用影像以及生理數據。



圖 1-1 ios 天氣 app 介面



圖 1-2 ios 行事曆 app 介面



圖 1-3 ios 健康 app 介面

為提升個人化體驗，本團隊提出將 AI 助理結合穿戴裝置，像是智

慧手錶¹、智能眼鏡²等設備，運用這些智能裝置收集個人生理數據和環境資料，藉此開發出一款提供使用者更智慧且更貼近的互動體驗之個人化 AI 助理，其功能不再是只有回應使用者的指令，而是根據使用者的狀況，主動判斷且提出建議，像是使用者有睡眠品質不佳的情況，AI 助理可以主動提出屬於使用者的專屬改善方案。



圖 1-4 apple watch



圖 1-5 meta glasses

本研究為針對整合不同數據來源和實時數據更新，從多個平台獲取數據，透過 API 整合不同應用程式，像是電子郵件、行事曆等，並確保數據的更新，提供個人化的建議，提升使用者的體驗品質。

本團隊希望能開發出具備同時處理文字、語音、圖像等多種形態數據的 AI 助理，更全面的提供使用者精確的反饋。本研究期望能根據使用者的個人行為和情緒反應，以滿足個性化需求。

¹ <https://www.apple.com/tw/>

² <https://www.meta.com/tw/>

1.2 研究問題

為了達成此研究，要解決的問題如下：

1. 如何設計語音身分辨識的登入介面。
2. 如何改善 AI 助理對用戶情感的理解與反應。
3. 如何整合不同平台的數據，以提升 AI 助理的個性化服務。
4. 如何從 apple watch 抓取數據。
5. 如何將 AI 助理結合 meta glasses。

第二章 研究方法及步驟

本團隊希望開發出一套使用者能夠不受平台限制且跨設備使用之系統，此章節將分成三個部分做研究（一）語音身分辨識訓練方法（二）情緒辨識的準確率及訓練方法。

2.1 語音身分辨識訓練方法

由於現代科技越發進步，即便再多複雜的密碼也難免於駭客入侵破解，為了減少這樣的狀況發生，因此採用語音辨識系統不單作為快速登入的身分驗證更是為了提高系統安全性。

首先請組員提供中文、英文、日文、韓文四國語言，每人每種語言採集四五種短句做為前處理之樣本，多國語言訓練模型能夠對不同的語音特徵有更佳的識別能力，提高辨識的準確性。

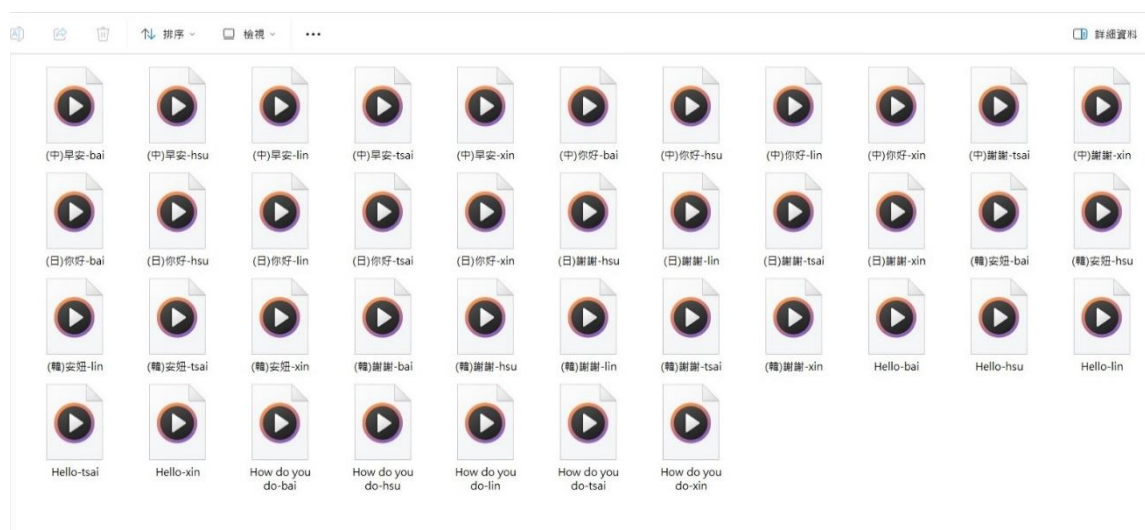


圖 2-1 組員四國語言音檔

起初使用 Audacity 對音訊做前處理動作，進行至一半發現太土法煉鋼，未來用於系統上也不可能一個新用戶進來一個個做處理沒有效率，因此使用 Visual Studio Code 寫程式進行語音前處理。再來進行降噪使音檔在後續進行訓練時能夠減少不必要干擾和誤差的狀況，提高語音辨識準確度，此外也運用數據增強技術，提高音量、增加背景噪音擴展資料樣本多樣化，在有限的資料中產生最佳效益。

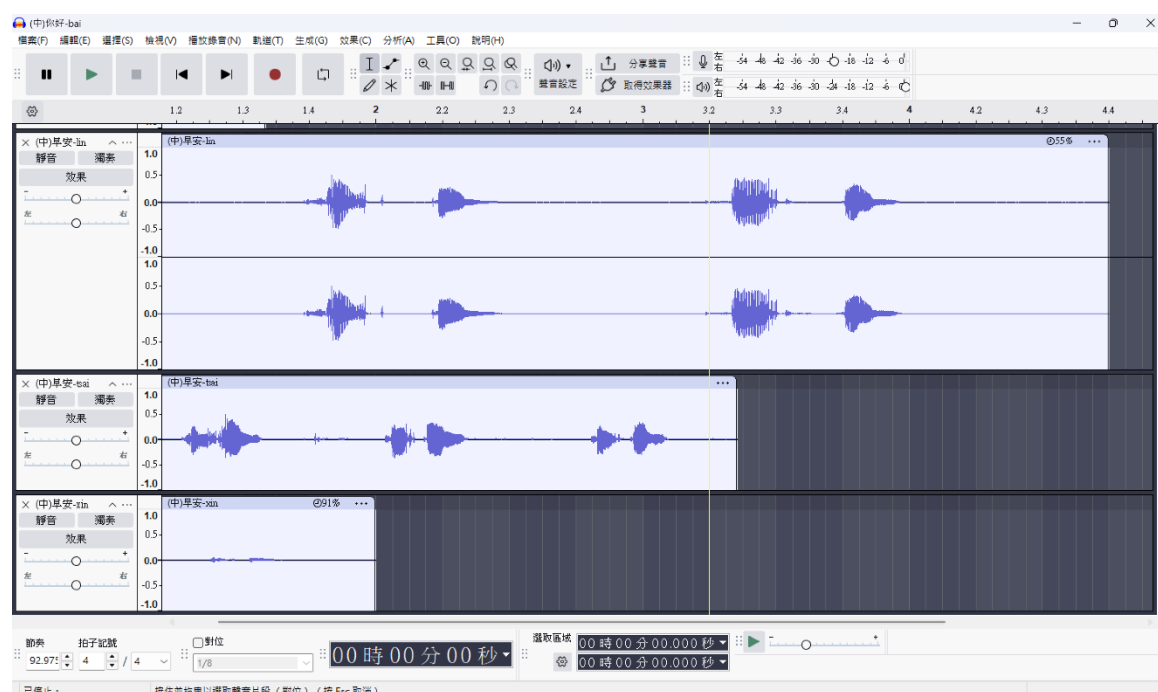


圖 2-2 初始前處理畫面

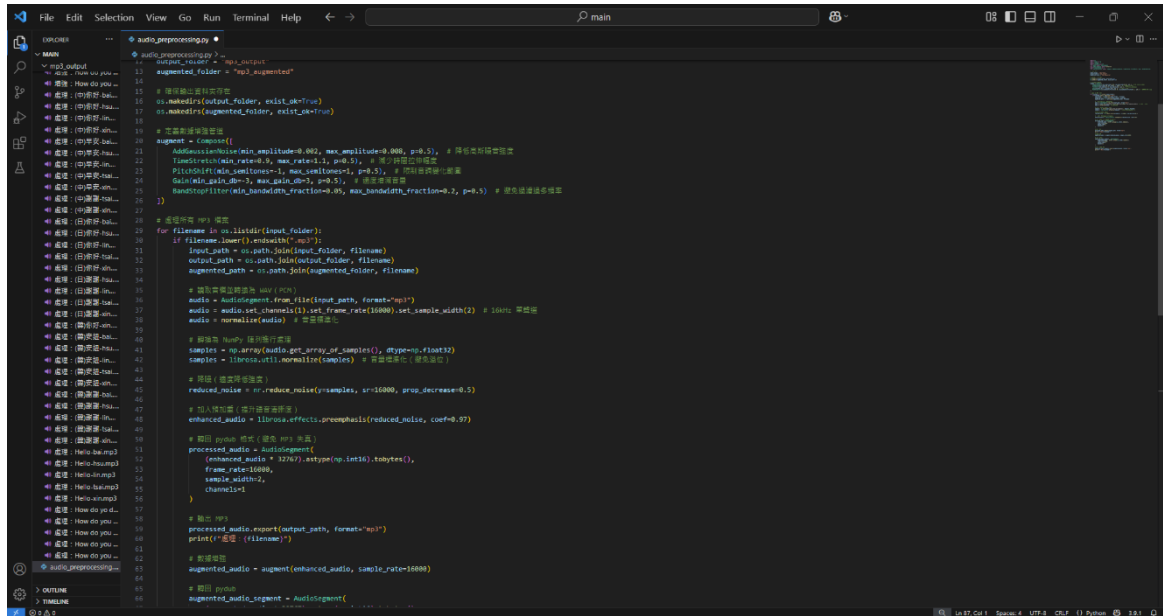


圖 2-3 前處理完成音檔

前處理完畢後進行特徵提取，從各音檔提取聲學特徵，使用 `librosa.feature.mfcc()` 函數中取出 40 維的 MFCC 特徵去統計，並轉為數字特徵用來訓練模型。訓練模型選用了三種方法, SVM 、 Random Forrest, XGBoost ，從中找出表現最佳的模型做使用。

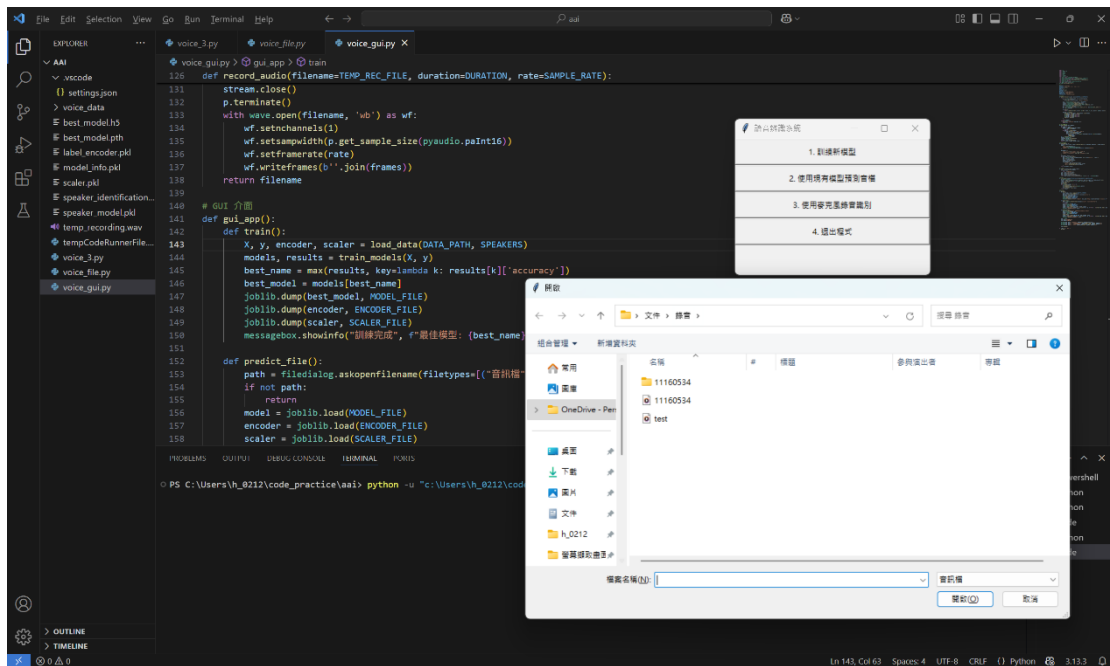


圖 2-4 語音辨識系統畫面

表 2-1 傳統語音助理和個人化語音助理的比較

特徵	傳統語音助理	個人化語音助理
互動模式	主要基於語音指令, 反應固定	支援語音 文字和影像的整合
個性化服務	提供的服務較為通用, 缺乏個人化體驗	根據使用者的喜好和需求提供量身定制的个人化服務
數據互通性	數據來源有限 通常依賴單一平台或服務	能夠從多元化的資料來源中獲取數據 提供更準確的答案

技術基礎	基本的語音辨識和自然語言處理技術	結合大型語言模型和機器學習
使用者體驗	互動較為生硬 缺乏人性化的交流	能夠理解情感 更自然且流暢的互動體驗

2.2 情緒辨識的準確率及訓練方法

情緒辨識模組我們採用 Hugging Face 開源平台上的華語情緒識別(hubert-base-ch-speech-emotion-recognition)模型，利用 CASIA 資料集中 1200 位演員表演六種情緒，分別為開心、憤怒、悲傷、驚訝、恐懼、無意見，用來作為資料訓練。

識別過程中大致分為三步驟，先將音檔進行前處理成可用的檔案，如同 2.1 節語音前處理，再來 Hubert 模型採用非監督式學習進行特徵提取後開始進行訓練，第三步進行情緒分類，以最高機率類別作為結果。AI Agent 能夠根據當時使用者情緒狀況，自動做出調整給予更佳人性化的回覆，提升與使用者之互動性。

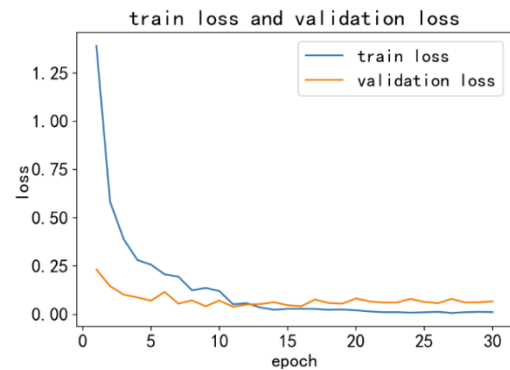
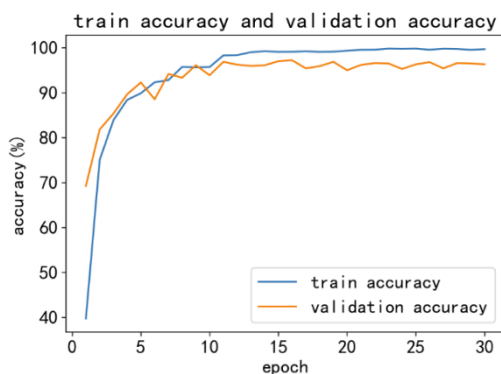


圖 2-5 Acc

圖 2-6 Loss

第三章 系統架構

本團隊希望開發出一套使用者能夠不受平台限制且跨設備使用之系統，此章節將分成三個部分做研究：（一）前端設備，（二）後端應用，（三）雲端數據管理。

3.1 前端設備

前端針對不同平台分別進行適應開發，確保各平台核心功能一致性，為 Android 、iOS、Mac 等行動裝置提供完整的操作體驗，包含語音輸入、聲音識別及資料顯示。Apple Watch 、Meta glasses 穿戴式裝置，著重於特定用途如抓取生理健康資訊、AI 語音功能等。

健康資訊利用 HealthKit 授權健康資料存取步數、睡眠和心律心跳查詢，傳至 MongoDB 。 MongoDB 是一種非關聯式資料庫，不僅高效能還能處理大量資料，自由添加欄位無限擴展儲存資料。Meta glasses 由於 SDK 未公開因此用藍牙與行動裝置連接，接收語音後將指令發送至後端處理後，最後傳回至智慧型眼鏡中進行播放。

3.2 後端處理應用

後端處理主要分成三類，API 處理、GPT 處理、數據管理三種。API 處理負責接收前端的請求，並根據其請求做處理；GPT 處理用於語義理解後並生成回應；數據管理則是負責資料的儲存及檢索。

本系統在語義理解方面，使用了最新的大型語言模型 (LLM) 技術：雙層 GPT-4o-mini，一層負責進行初步語義解析，將輸入的文字轉為標準化的格式，另一層則是進行整合及生成回應。利用此技術原因在於使系統在處理和計算時能夠更加精準判斷。

系統中使用多種 API 如天氣查詢和新聞資訊功能，加入 Open Weather API 提供即時及未來天氣預報，新聞則是整合多家媒體，根據使用者查詢關鍵字進行篩選推薦。利用 API 的優點在於未來持續擴展應用時不用對主程式做修改，只要加入所需 API 即可。

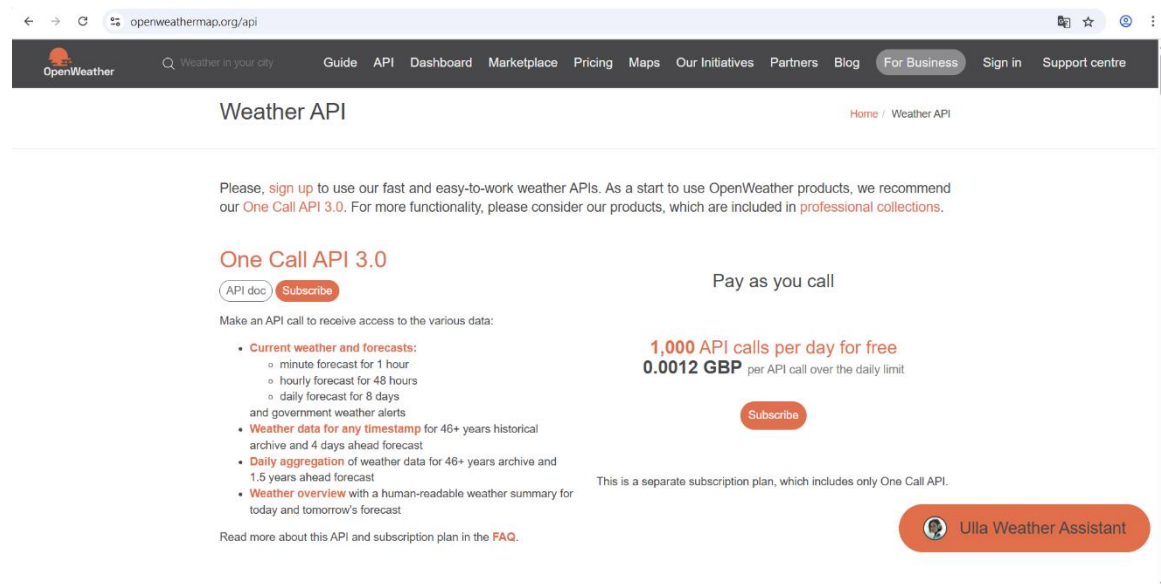


圖 3-? Open Weather API 網站

3.3 雲端數據管理

本系統雲端資料庫使用 MongoDB, 屬於一種非關聯式資料庫, 適合大量儲存或檢索非結構化的對話數據及半結構化的使用者資料, 本系統分為四大儲存庫: (一) 使用者資料訊息、(二) 裡使對話紀錄、(三) Apple Watch 健康數據、(四) 全域設定。

第四章 結論

(貼完成介面照片眼鏡圖等等 說明介紹能幹嘛)

4.1 預期結果

4.2 未來展望

本團隊期望未來的 AI 助理能達到主動理解使用者，並且預測使用者需求的智慧助理，根據歷史紀錄和偏好設定，提供更為精準和個人化的建議和反饋。相較於傳統的 AI 助理提供反應，預測型 AI 為未來的發展趨勢，預先判斷或推測使用者需求，主動提出即時建議。

隨著 AI 技術的進步，結合情緒辨識的 AI 助理，會更受使用者喜愛，提供更加人性化的服務，提升使用者的體驗滿意度。開發出支持語音、文字、圖像等互動模式，將會有數據安全和隱私保護的問題需要考量和評估，在提供個人化服務的同時，也拗確保使用者的數據安全性，並提供透明的數據使用政策，讓使用者了解到數據是如何收集和使用，以增加使用者對此系統的信賴。

參考文獻

- [1] Xiangang Zuo, Zhixia Zhang and Jianping Xie, “Detection on Large Amount of Web Pages Based on a Highly Active Way of Site-Level Template,” International Journal of Hybrid Information Technology, vol. 8, no. 4, pp. 259-266, 2015.
- [2] Alpa K. Oza and Shailendra Mishra, “Elimination of Noisy Information from Web Pages,” International Journal of Recent Technology and Engineering, vol. 2, issue. 1, ISSN: 2277-3878, 2013.
- [3] Chien-Li Chou, Hua-Tsung Chen and Suh-Yin Lee, “Pattern-Based Near-Duplicate Video Retrieval and Localization on Web-Scale Videos,” Multimedia, vol. 17, no. 3, pp. 382-395, 2015.

附錄一 工作分配

劃分工作內容，並列出小組之工作分配。

表 A-1 工作分配表

工作項目 組員	白東衢	許友銓	蔡承哲	林靄君	李宇芯
資料蒐集	100%	100%	100%	100%	100%
文件整理	100%	100%	100%	100%	100%
介面設計	100%	100%	100%	100%	100%
程式撰寫	100%	100%	100%	100%	100%
圖表製作	100%	100%	100%	100%	100%
文件撰寫	100%	100%	100%	100%	100%

附錄二 甘特圖

列出各項工作內容隨著時間進展之情形。

