## 使用說明

## 功能簡介

主要用於批次處理 input 資料夾內的圖片或 PDF 檔案，透過AI進行資料萃取，並將結果以 JSON 格式輸出至 output 資料夾。支援自動搬移處理後的檔案、錯誤檔案管理、日誌記錄、動態 prompt 對應等。

## 支援檔案格式

* 圖片：.jpg, .jpeg, .png, .bmp
* PDF：.pdf

## 主要流程

1. 掃描 input 資料夾（含子資料夾）所有支援的檔案
2. 依據檔案類型自動選擇 prompt（可依子資料夾對應不同 prompt）
3. 呼叫 API 進行資料萃取
4. 將結果以 JSON 格式輸出至 output 資料夾（保留原有子資料夾結構）
5. 處理成功的檔案自動搬移至 old 資料夾，失敗則搬移至 error 資料夾
6. 產生 summary log

## 參數與設定

* API\_KEY： API 金鑰（由開發者設定）
* API\_URL： API 端點（由開發者設定）
* INPUT\_FOLDER：輸入檔案資料夾（預設 input）
* OUTPUT\_FOLDER：輸出 JSON 資料夾（預設 output）
* OLD\_FOLDER：成功處理後檔案搬移資料夾（預設 old）
* ERROR\_FOLDER：失敗檔案搬移資料夾（預設 error）
* PROMPT\_DIR：prompt 對應目錄（可依子資料夾自訂 prompt）
* PROMPT\_FILE：預設 prompt 檔案
* LOG\_FILE：日誌檔案路徑
* LOG\_TO\_FILE：是否寫入日誌檔
* MOVE\_PROCESSED\_FILE：是否自動搬移處理後檔案

## 執行方式

python Extract

### 支援參數（可選）

* --input\_folder 指定輸入資料夾
* --output\_folder 指定輸出資料夾
* --old\_folder 指定 old 資料夾
* --error\_folder 指定 error 資料夾
* --prompt\_file 指定 prompt 檔案
* --log\_file 指定 log 檔案
* --move\_processed\_file / --no\_move\_processed\_file 控制是否搬移檔案
* --log\_to\_file / --no\_log\_to\_file 控制是否寫入日誌檔

範例：

python --input\_folder ./input --output\_folder ./output --move\_processed\_file

## 資料夾結構建議

AI/

input/ # 原始檔案（可有多層子資料夾）

output/ # 萃取結果 JSON（保留子資料夾結構）

old/ # 處理成功後搬移的檔案

error/ # 處理失敗搬移的檔案

prompt/ # 可依 input 子資料夾自訂 prompt.txt

prompt.txt # 預設 prompt

extract\_log.txt # 日誌檔

## 常見問題

* **API 金鑰錯誤或配額不足**：請確認 API\_KEY 正確且有配額
* \*\* 回傳格式非 JSON\*\*：程式會嘗試自動去除 markdown 標記，若仍無法解析，會寫入空陣列
* **檔案搬移失敗**：請確認 old/error 資料夾有寫入權限
* **prompt 對應**：可在 prompt/ 下建立與 input 子資料夾同名的 .txt 以自訂 prompt

## 其他

* 執行時會自動產生 config.json（若不存在）
* 產生 summary log 記錄成功與失敗檔案
* 支援 logging 記錄詳細處理過程