

# PPTX 轉 AI 影片 自動化流程

使用 Google Gemini 2.5 與 HeyGen API 的零接觸工作流程

。

自動將靜態簡報轉化為生動的 AI 主播解說影片。

# | 系統架構



## 輸入

讀取本機 .pptx 檔案



## 轉檔

LibreOffice + Poppler -> PNG 圖片



## Gemini AI

視覺分析 -> 講稿生成



## HeyGen

虛擬主播影片 + 字幕

# | 前置需求與環境

## 使用環境&需額外下載軟體

- ✓ 核心執行環境: Python 3.12.7
- ✓ **LibreOffice**: 必須安裝，用於 PPT 轉 PDF
- ✓ Poppler for Windows (用於 PDF 轉圖片)

## 必要套件

- ✓ 參考 requirements.txt

```
≡ requirements.txt
1  google-genai>=1.0.0
2  requests
3  python-dotenv
4  pdf2image
```

# P案資料夾結構

## 根目錄 (Root)

heygen\_project\_PTT/ (專案資料夾)

## 關鍵檔案 (必須存在)

main.py : 自動化主程式

.env : API 金鑰設定檔

test.pptx : 欲報告之簡報

requirements.txt : 套件清單

```
HeyGen_Project_PPT/
├── .env                # [重要] 存放 API Key
├── main.py             # 主程式碼
├── requirements.txt    # 套件清單
├── test.pptx          # 輸入的簡報檔
└── outputs/           # (程式自動建立) 存放圖片與結果影片
```

# | Poppler (用於 PDF 轉圖片)

- 這是 pdf2image 套件背後的引擎，沒有它，程式會報錯 Poppler not found。
- 下載點: [Poppler Windows Releases \(GitHub\)](#)  
請下載最新版本的 .zip 檔 (例如 Release-24.02.0-0.zip)。
- 安裝步驟 (關鍵！):
  1. 下載後解壓縮。
  2. 將解壓出來的資料夾 (裡面會有 bin, include, lib 等資料夾) 搬到一個您不會刪除的地方，例如 C:\Program Files\poppler。
  3. 設定環境變數 (PATH):
    - 按 Win鍵 搜尋「編輯系統環境變數」。
    - 點選「環境變數」-> 在「系統變數」區塊找到 **Path** -> 點選「編輯」。
    - 點選「新增」，貼上 bin 資料夾的路徑 (例如：C:\Program Files\poppler\Library\bin 或 C:\Program Files\poppler\bin)。
    - 一路按確定儲存。
  4. 重啟 VS Code 或終端機讓設定生效。

# LibreOffice

1. 官方下載頁面：<https://zh-tw.libreoffice.org/download/libreoffice-fresh/>

安裝與設定步驟：

下載：進入網頁後，作業系統通常會自動偵測為 **Windows (64-bit)**，點擊「下載」即可。

安裝：執行下載的檔案，一路按「下一步」使用預設安裝即可。

關鍵檢查 (Check Path)：安裝完成後，請打開您的檔案總管，確認一下這個檔案是否存在：

C:\Program

Files\LibreOffice\program\soffice.exe

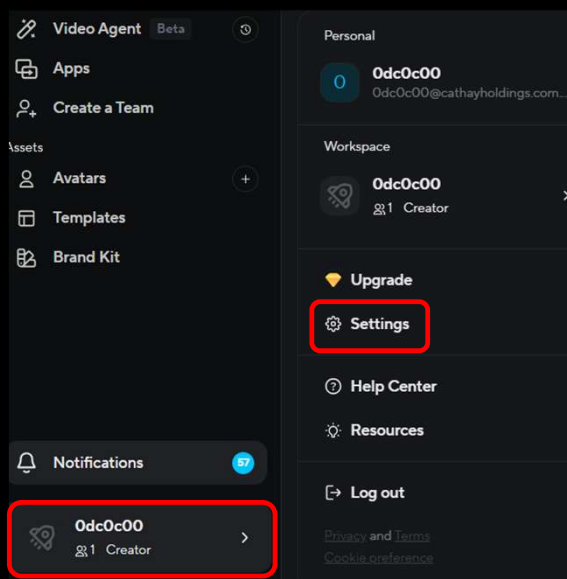
2. 如果存在：您的 Python 程式碼完全不用改，直接執行即可。

3. 如果不存在 (例如您安裝在 D 槽)：請找到 soffice.exe 的位置，並將程式碼中的 WINDOWS\_SOFFICE\_PATH 改成您實際的路徑。

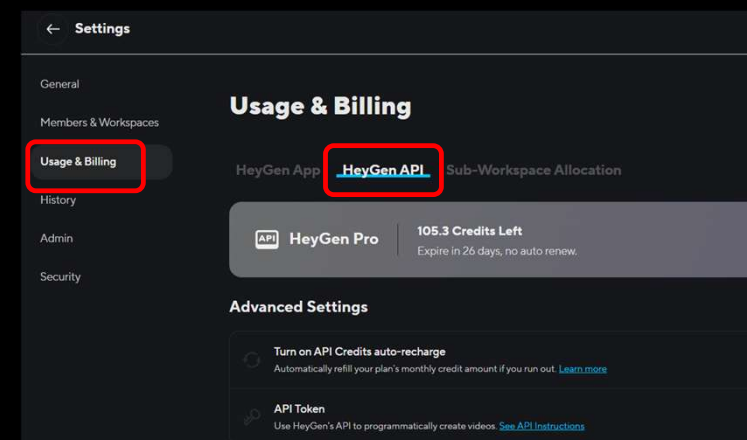
「請務必在 main.py 中修改  
**WINDOWS\_SOFFICE\_PATH** 路徑，使其指向您電腦中  
實際安裝的 LibreOffice 位置。」


# 如何獲取 HeyGen API 金鑰

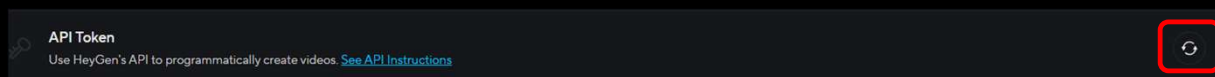
1. 左下角頭像 → Account Settings / Settings



2. 左邊側邊欄點選Usage & Billing 再點選 HeyGen API



3. HeyGen API 右邊有一個  符號，當忘記遺失 API 可重新產生 API



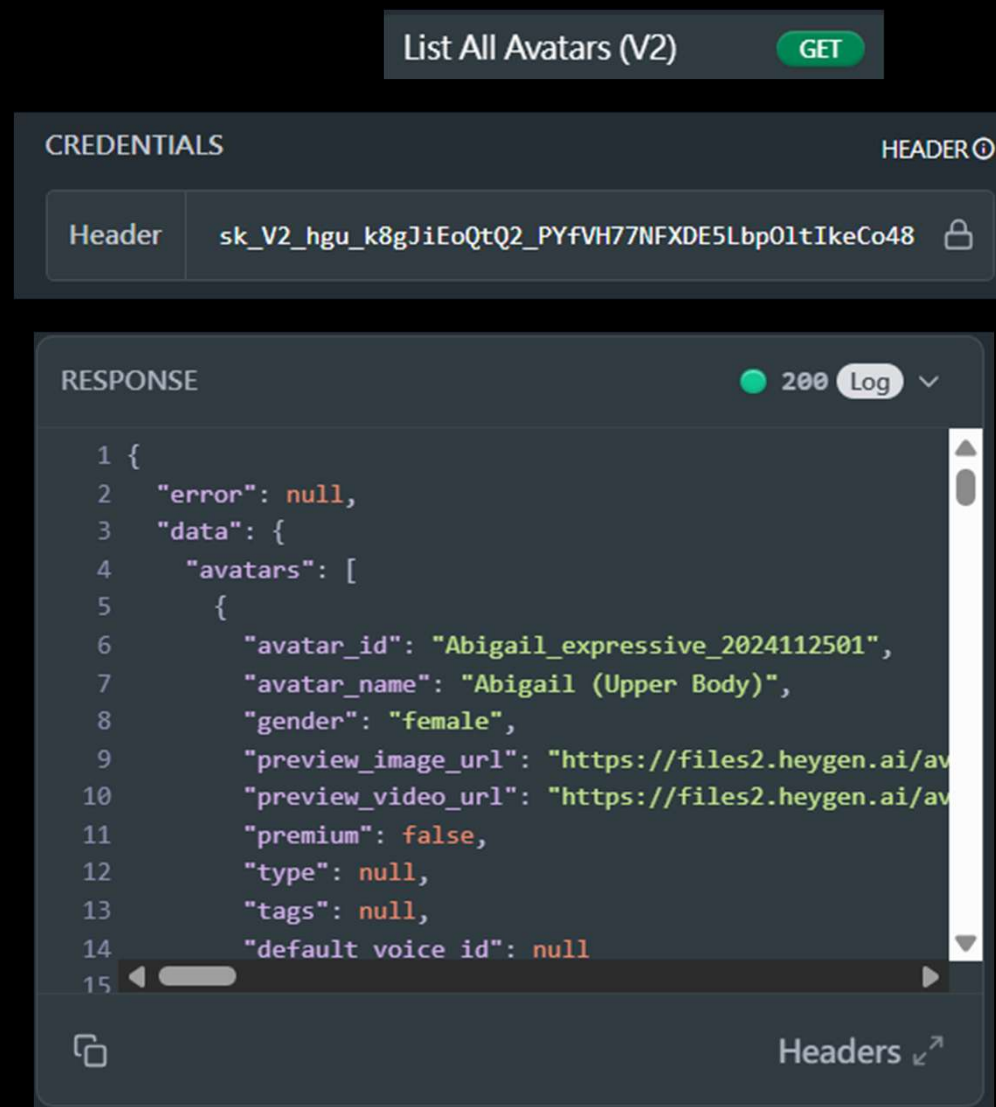
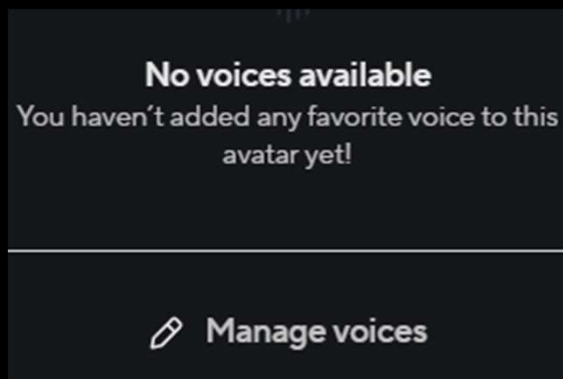
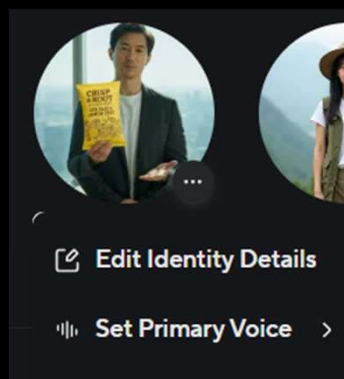
HeyGen API 如無法使用，請遵照上方步驟，將新取得 API 取代 .env 之 API

# 挑選聲音以及數字人

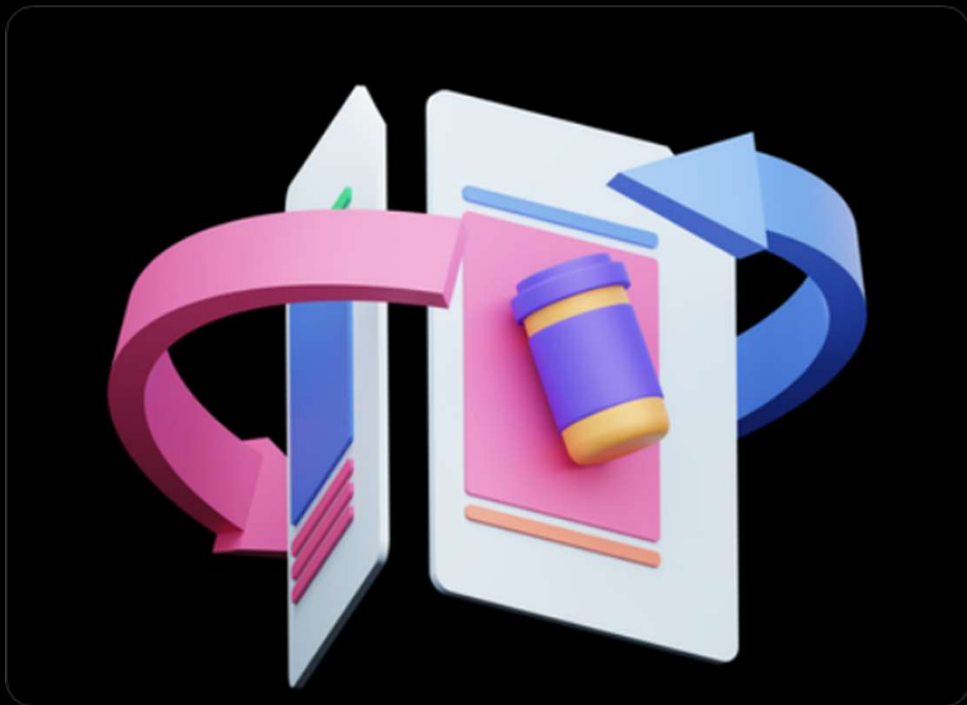
<https://docs.heygen.com/reference/list-avatars-v2> 查詢可使用之數字人ID

- 在網頁右上角選取 Python 輸入 API 即可從 RESPONSE 知道可使用之數字人ID
- 可將回應複製到 WORD，接著利用快捷鍵 CTRL+F 來查找自己想要的ID

聲音ID 部分請到 UI 介面右邊點選 Avatars，隨便一個數字人，接著選擇 Set Primary Voice，接著點選 Manage voices，下一步複製想要之聲音 ID



# | 步驟 1 : PPTX 轉圖片



## 兩階段轉檔

由於難以直接讀取 PPTX 視覺內容，我們先轉為 PDF，再轉為高解析度 PNG。

## 步驟 2：Gemini 視覺分析

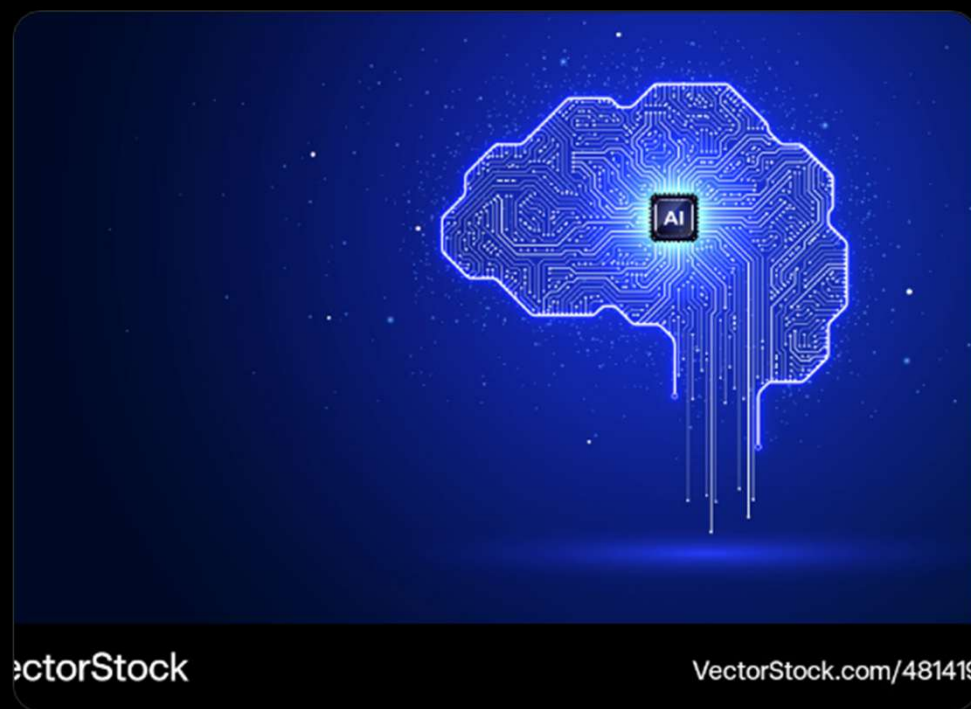
### 「看懂」簡報內容

我們使用最新的 google-genai SDK 上傳簡報圖片，並生成約 25 秒的繁體中文口語講稿。

```
client = genai.Client(api_key=KEY)
file_ref = client.files.upload(file=img_path)

prompt = "你是專業講師...生成約 25 秒的口語講稿"

response = client.models.generate_content(
    model="gemini-2.5-flash-lite",
    contents=[file_ref, prompt]
)
```

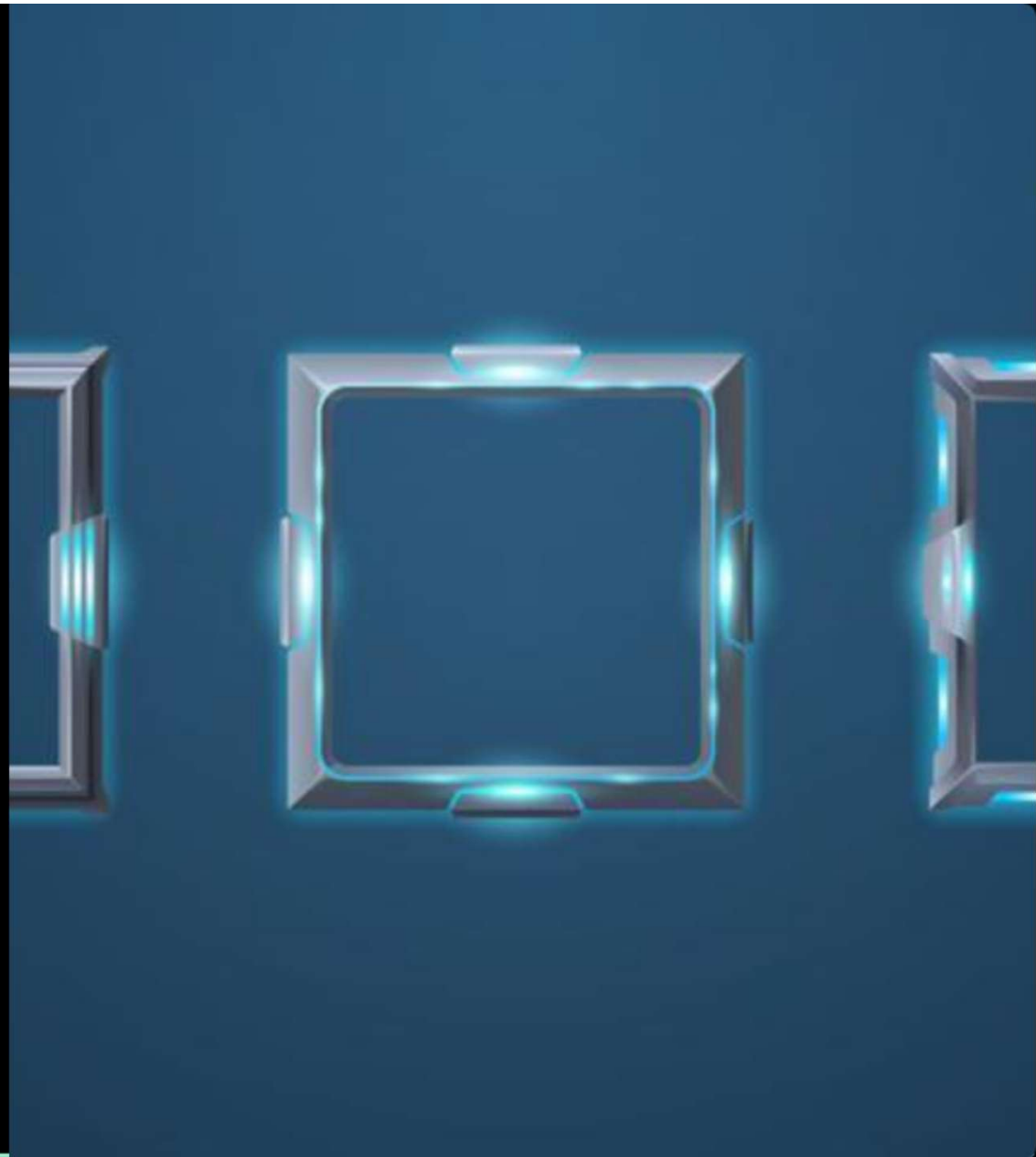


## 步驟 3：影片合成

### Talking Photo 模式

我們為每張投影片建立場景物件。talking\_photo 模式能讓靜態人像配合講稿生成臉部動畫。

```
scene = {  
  "character": {  
    "type": "talking_photo",  
    "talking_photo_id": TALKING_PHOTO_ID  
  },  
  "voice": { "input_text": script },  
  "background": { "image_asset_id": bg_id }  
}
```



## 步驟 4：字幕自動化



### 處理延遲

HeyGen 影片常有微小的聲音延遲。我們下載字幕，將 ASS/VTT 轉為 SRT，並套用負值的時間偏移量修正。

```
def convert_ass_to_srt(path, offset=-1.3):  
    # 解析 ASS 時間戳記  
    start_dt = parse_ass_time(start_str)  
    # 套用時間偏移  
    new_start = start_dt + timedelta(seconds=offset)  
    # 寫入 SRT 格式  
    ...
```

```
$ python main.py
```

```
[1/5] 正在處理投影片: test.pptx... OK
```

```
[2/5] Gemini 正在看圖說故事... OK
```

```
[3/5] 建立影片任務 (Talking Photo)... ID: 12345
```

```
[4/5] 下載影片與字幕... Done
```

```
[5/5] 格式轉換完成！
```

```
>> final_output.mp4 Ready.
```

# 開始製作吧!

準備開始自動化內容創作~