

# 五大 AI 模型多模態功能完全指南(2024-2025)

ChatGPT、Gemini、Claude、Copilot 與 Grok 在語音、影片、圖像處理的實戰能力與創意應用全面解析

AI 聊天機器人已從純文字工具進化為真正的多模態平台,但各模型的能力差異巨大。Gemini 是唯一能直接處理音訊檔案並翻譯的主流模型,Claude 無法原生處理語音,而 Grok 完全不支援轉錄功能。[Mac Performance Guide](#) 本報告深入比較五大 AI 在音訊轉錄、影片分析、檔案處理等核心需求的實際操作流程,並收錄 Reddit、X 等社群的進階技巧與隱藏功能,協助你選擇最適合的工具組合。

## 一、音訊處理能力:誰能真正轉錄你的錄音檔?

### 直接上傳音訊檔案的支援程度

Gemini 領先群雄,是唯一支援直接上傳 MP3/WAV/M4A 並進行轉錄翻譯的主流模型。在 Google AI Studio 中,用戶可上傳最高 2GB 的音訊檔案,總長度限制為 9.5 小時。Gemini 不僅能轉錄,還能自動偵測語言、標註說話者、識別情緒(開心/悲傷/憤怒/中性)。[google](#)

| 模型      | 直接上傳音訊        | 檔案大小限制 | 支援格式                           | 轉錄準確度      |
|---------|---------------|--------|--------------------------------|------------|
| Gemini  | ✅ 完整支援        | 2GB    | WAV, MP3, AIFF, AAC, OGG, FLAC | 高(含時間戳)    |
| ChatGPT | ⚠️ 僅 App/API  | 25MB   | MP3, WAV, M4A, WEBM            | 高(Whisper) |
| Claude  | ✗ 不支援         | —      | —                              | —          |
| Copilot | ⚠️ 需用 Word 轉錄 | 依功能而異  | MP3, WAV, M4A                  | 中等         |
| Grok    | ✗ 不支援         | —      | —                              | —          |

ChatGPT 的音訊處理需透過手機 App 或 macOS 桌面版,網頁版無法直接上傳音訊。背後使用 OpenAI 的 Whisper 模型,\*\* (OpenAI)25MB 檔案限制\*\* (OpenAI)約等於 15-20 分鐘的音訊。  
[getcockpit](#)

## 英文→繁體中文翻譯的最佳工作流程

### 推薦方案:使用 Gemini 2.0 Flash

#### 實際操作步驟:

1. 前往 [aistudio.google.com](https://aistudio.google.com)
2. 選擇 Gemini 2.0 Flash 模型
3. 上傳英文音訊檔案
4. 輸入 Prompt:  
「請將此音訊轉錄為英文逐字稿,並翻譯成繁體中文(台灣用語)。  
請同時輸出英文原文與中文翻譯,格式包含時間戳記。」
5. 複製結果,依需求調整格式

**免費 vs 付費差異:**Gemini 免費版限制音訊長度 **10 分鐘**,AI Pro 訂閱(\$19.99/月)可處理最長 **3 小時**的音訊。[\(google\)](#)

若使用 ChatGPT,需採兩步驟流程:先用 Whisper 轉錄為英文,再請 ChatGPT 翻譯。Whisper 只能翻譯「成英文」,無法直接翻譯成中文。[\(OpenAI\)](#)

---

## 二、影片處理能力:字幕生成與摘要功能

### 直接上傳 MP4 影片的支援情況

| 模型     | 直接上傳影片                                   | 長度限制            | 字幕格式     | YouTube 分析                                |
|--------|--|-----------------|----------|---|
| Gemini | <input checked="" type="checkbox"/> 完整支援 | 免費 5 分鐘/付費 1 小時 | SRT, VTT | <input checked="" type="checkbox"/> 直接貼網址 |

| 模型      | 直接上傳影片     | 長度限制          | 字幕格式     | YouTube 分析 |
|---------|------------|---------------|----------|------------|
| ChatGPT | ✓ 有限支援     | 25MB (~15 分鐘) | SRT, VTT | 需 GPT 外掛   |
| Claude  | ✗ 不支援      | —             | —        | ✗          |
| Copilot | ✓ M365 生態系 | 依平台而異         | 內建字幕     | ✓ Edge 整合  |
| Grok    | ✗ 不支援      | —             | —        | 僅 X 平台影片   |

**Gemini 的獨家優勢**是可以直接貼上 YouTube 網址進行分析。只需輸入影片連結並詢問「這部影片在講什麼？」,Gemini 會自動提取字幕並生成帶有時間戳記的回覆,點擊即可跳轉到影片相關段落。 TechRadar

## MP4 影片生成 SRT 字幕的完整流程

使用 Gemini(推薦):

### 操作步驟:

1. 進入 [aistudio.google.com](https://aistudio.google.com)
2. 點選「上傳檔案」→ 選擇 MP4 影片
3. 輸入 Prompt:  
「請為這部影片生成 SRT 格式字幕:  
  - 使用標準 SRT 時間戳格式(00:00:00,000)
  - 每條字幕限制在合理長度
  - 確保時間戳與音訊精確對齊
  - 同時提供英文原文與繁體中文翻譯」
4. 複製輸出內容,存檔為 .srt 檔案

**長度限制注意事項:**Gemini 單次請求約可處理 **8-10 分鐘**影片(受 8,000 token 輸出限制)。超過此長度的影片建議切割為 5 分鐘片段分別處理。

**成本估算:**透過 API 使用約 **\$0.10/小時**影片內容,免費版在 AI Studio 有每日額度限制。

## Claude 用戶的替代方案

由於 Claude 不支援影片處理,需先用外部工具提取音訊:

```
bash
```

```
# 使用 FFmpeg 提取音訊  
ffmpeg -i video.mp4 -vn audio.mp3
```

然後使用 Whisper 或 AssemblyAI 轉錄,再將逐字稿上傳至 Claude 進行翻譯與潤飾。Claude 的優勢在於後處理品質極高,特別適合需要精修翻譯或生成摘要的場景。

---

## 三、檔案格式支援與上傳限制總覽

### 各模型檔案處理能力比較

| 格式         | ChatGPT | Gemini  | Claude          | Copilot | Grok   |
|------------|---------|---------|-----------------|---------|--------|
| PDF        | ✓ 文字提取  | ✓ 視覺+文字 | ✓ 最強 (100頁視覺分析) | ✓       | ✓      |
| Word/DOCX  | ✓       | ✓       | ✓               | ✓ 原生整合  | ✓      |
| Excel/XLSX | ✓ ~50MB | ✓ 需付費   | ⚠ 需啟用分析工具       | ✓ 最強    | ✓      |
| PowerPoint | ✓       | ✓       | ⚠ 有限            | ✓ 最強    | ⚠      |
| 圖片         | ✓ 20MB  | ✓ 100MB | ✓ 30MB          | ✓       | ✓ 25MB |
| CSV        | ✓       | ✓       | ✓               | ✓       | ✓      |

## 檔案大小與數量限制

### ChatGPT:

- 單檔硬性限制:**512MB** ([OpenAI Help Center](#))
- 圖片:**20MB** ([OpenAI Help Center](#))
- 文字/文件:**2M tokens** ([OpenAI Help Center](#))
- Plus 用戶:每 3 小時最多 **80 個檔案** ([Medium](#)) ([Data Studios](#))
- 免費用戶:每天 **3 個檔案** ([Medium](#)) ([OpenAI Help Center](#))

### Gemini:

- 影片/音訊:**2GB** ([Data Studios](#))
- 其他檔案:**100MB** ([New School](#)) ([Google Support](#))
- 免費版 Context:**32,000 tokens**(約 50 頁)
- AI Pro:**1M tokens**(約 1,500 頁) ([9to5Google](#))

### Claude:

- 單檔限制:**30MB** ([Data Studios](#)) ([CometAPI](#))
- 每次對話:**20 個檔案** ([Data Studios](#))
- PDF 視覺分析:**最多 100 頁** ([Anthropic +2](#))
- Context window:**200,000 tokens** ([Claude](#))

### Copilot:

- 無授權:**1MB**

- M365 Copilot 授權:**512MB** ([Microsoft Learn](#)) ([Medium](#))
  - 最佳摘要長度:**150 萬字以下** ([Data Studios](#))
- 

## 四、五大模型殺手級應用與獨特優勢

### ChatGPT:Custom GPTs 生態系 + DALL-E 3 整合

**Custom GPTs 數據:**超過 300 萬個 Custom GPTs 已被創建, ([OpenAI](#)) GPT Store 有 **15 萬+公開 GPTs**。  
([BYU](#))

#### 推薦 GPTs:

- **Video Insights:**輸入 YouTube 網址自動提取逐字稿並分析
- **Screenshot To Code GPT:**截圖轉 HTML/Tailwind/JS 程式碼
- **AI Meme Generator:**輸入主題自動生成 Meme 圖

#### DALL-E 3 整合:

- 直接在對話中生成圖片 ([OpenAI](#))
- 文字渲染成功率超過 **95%** ([Tom's Guide](#))
- 免費用戶:每天 **2 張圖**
- 2025 更新:GPT-4o 接手圖片生成,支援透明背景 PNG

#### Advanced Data Analysis(Code Interpreter):

- 沙盒環境執行 Python 程式碼 ([Codecademy](#))
- 分析上傳的數據檔案 ([Codecademy](#))

- 生成視覺化圖表(Matplotlib)
- 檔案限制:100MB,執行逾時:60秒

## Gemini:超長上下文 + Google 生態系深度整合

### 1M Token Context Window 實際應用:

- 約等於 1,500 頁文字
- 30,000 行程式碼 (Google Support)
- 1 小時影片(預設解析度)
- 11 小時音訊

### Google Workspace 整合(2025 年 1 月起納入 Business/Enterprise):

- Gmail:摘要郵件串、AI 建議回覆 (Google Support)
- Google Docs:草稿生成、圖片插入 (Data Studios)
- Google Sheets:建立公式、樞紐分析表、Smart Fill (Google Workspace)
- Google Meet:自動會議記錄、行動項目、多語言支援 (Data Studios)

### YouTube 影片分析 Prompt 範例:

「分析這部影片的主要論點,並以繁體中文列出:

1. 核心觀點(附時間戳)
2. 支持論據
3. 潛在反駁
4. 行動建議」

### 旅遊規劃整合(Google Maps/Flights/Hotels):

「規劃東京 5 天行程,幫我找下個月從台北出發的機票、  
預算 5000 台幣以下的飯店,並標註每個景點的交通方式」

## Claude:PDF 分析王者 + Artifacts 互動應用

### PDF 處理是 Claude 的最強項:

- **雙模態處理:**同時進行文字提取 + 視覺分析 (Medium)
- 每頁轉換為圖片進行全面理解 (THE DECODER +2)
- 支援圖表、表格、手寫文字(OCR) (Towards Data Science)
- Token 成本:每頁 1,500-3,000 tokens (Claude) (THE DECODER)

### Artifacts 功能(2024 年 6 月推出):

- 即時預覽生成的程式碼 (Anthropic)
- 互動式單頁應用(HTML/CSS/JS) (Simon Willison)
- React 元件即時渲染
- Mermaid 圖表(流程圖、心智圖)
- 數據視覺化(Plotly.js) (DataCamp)
- 支援版本控制與發布分享 (LogRocket)

### 程式碼品質領先:

- SWE-Bench 測試:Claude Opus 4 達 72.5%(GPT-4 為 54.6%) (IntuitionLabs)
- 支援 30+ 小時自主編碼工作階段 (Anthropic)

## Microsoft Copilot:Office 365 原生整合的生產力巨獸

### Word Copilot:

- 從提示或參考檔案生成文件 (Microsoft Support)
- 摘要長文件 (think-cell)
- 調整語氣、風格、清晰度 (Microsoft Blogs) (think-cell)
- Agent Mode:自動從檔案、郵件、會議選擇素材 (Petri)

### Excel Copilot:

- 自然語言查詢:「上個月最大的支出是什麼?」 (Microsoft)
- AI 建議公式
- 自動建立圖表 (Microsoft) (Microsoft)
- Python 整合(進階分析)
- 新功能:選擇使用 Anthropic 或 OpenAI 模型 (Petri)

### PowerPoint Copilot:

- 從 Word 文件或大綱生成簡報 (Microsoft)
- Designer 整合生成 AI 圖片
- 建立講者備註 (think-cell)
- Agent Mode(Frontier 早期存取)

### Teams 會議 Copilot:

- 即時會議摘要與行動項目 (Microsoft Support)

- 「Catch me up」功能(遲到補進度)
- Audio Recap:AI 生成播客風格會議總結 ([Microsoft Learn](#))

## Grok:X/Twitter 即時數據存取 + 無審查風格

### 獨家優勢:即時 X 平台數據

- 唯一能存取 X 即時公開貼文的 AI 模型 (X)
- DeepSearch 技術:同時搜尋網路與 X 平台
- 趨勢追蹤、情緒分析、品牌監控 ([Latenode](#))
- 可設定自動每日/每週新聞摘要

### Aurora 圖片生成模型:

- 逼真渲染能力優異
- **關鍵優勢:**可生成真實人物(名人、政治人物)
- 圖片內文字渲染準確
- 較少內容限制

### Grok Imagine(2025 年 7 月):

- 文字轉影片:生成 6-15 秒短片
- 支援場景轉換、物件動畫、環境音效
- 圖片轉影片功能
- 需 SuperGrok 或 Premium+ 訂閱

### 訂閱價格:

- X Basic:\$3/月(有限 Grok 存取)
  - X Premium:\$8/月(Grok 存取)
  - X Premium+:**\$40/月**(完整 Grok 3/4)
  - SuperGrok:\$30/月(獨立訂閱)
  - SuperGrok Heavy:**\$300/月**(多代理人版本)
- 

## 五、社群創意用法與隱藏功能精選

### Reddit 熱門進階 Prompt 技巧

#### 1. Clarification Champion(釐清大師):

「請持續問我釐清問題,直到你有 95% 信心能成功完成這個任務。  
僅使用可驗證的來源,不要推測或包含虛構內容。」

#### 2. Red Team Protocol(紅隊檢視):

「對這個想法進行紅隊分析。它有什麼問題?有什麼弱點?」

#### 3. 時空對話(重大決策輔助):

「我想讓你扮演三個版本的我:  
- 6 個月前的我(在考慮這個改變之前)  
- 5 年後的我(如果我做了這個改變)  
- 5 年後的我(如果我沒做這個改變)  
讓他們就這個決定進行對話。」

#### 4. Socratic 學習法:

「用蘇格拉底式教學法教我 [主題]。在合理的地方使用第一原理思考。」

Reddit 用戶回饋: 「用這方法學習卷積神經網路架構效果極佳, 它讓我主動建構知識而非被動接收。」

#### 5. Gordon Ramsay Treatment(殘酷評論):

「用 Gordon Ramsay 的方式評論這個 [作品/想法/計畫]」

#### 隱藏功能揭密

##### ChatGPT Memory 功能:

- 路徑: Settings → Personalization → Reference saved memories
- 跨對話記住個人資訊
- 測試: 「記住我不喜歡香菜」
- 查看記憶: 「給我這次對話中所有記憶, 以分類結構化數據呈現」 (Medium)

##### Gemini 隱藏指令:

- 「Make this a power prompt」 : AI 自動優化你的提示詞
- @mentions: 使用 @docs、@youtube、@maps 跨平台搜尋
- Gems: 建立記住你目標的客製化 Gemini

##### Copilot Context IQ:

- 使用「/」搜尋文件、會議或聯絡人 (Microsoft)

- 範例: 「摘要我與 /[會議名稱] 的會議並建立行動項目」
- Word 中可參考最多 10 個來源(Word、PPT、PDF、會議記錄、郵件)

### 快速指令詞彙:

- 「TLDR:」 — 摘要重點
  - 「Humanize:」 — 讓 AI 文字更自然
  - 「Jargonize:」 — 讓文字更專業/企業風格
  - 「Continue」 — 當 ChatGPT 中斷時,輸入此詞繼續
- 

## 六、創意應用分類整理

### 1. 影音處理

| 應用           | 最佳工具              | 實際 Prompt                 |
|--------------|-------------------|---------------------------|
| 英文音訊→中文逐字稿   | Gemini            | 「轉錄此音訊並翻譯成繁體中文,輸出 SRT 格式」 |
| YouTube 影片摘要 | Gemini            | 直接貼網址+「摘要這部影片的五大重點」       |
| 會議錄音整理       | Copilot (Teams)   | 自動生成摘要、行動項目、決策紀錄          |
| Podcast 逐字稿  | ChatGPT (Whisper) | 透過 API 上傳,支援多語言           |
| 影片字幕翻譯       | Gemini            | 「生成雙語字幕(英文+繁體中文)」         |

## 2. 圖片處理

| 應用          | 最佳工具   | 說明  |
|-------------|--|---|
| 截圖→HTML/CSS | Claude Sonnet 3.7 + screenshot-to-code<br><a href="#">GitHub</a> | 支援 Tailwind、React、Vue<br><a href="#">GitHub</a> |
| 手繪草圖→UI     | Uizard / Google Stitch   | 掃描手繪→數位 Mockup, 可匯出 Figma                       |
| 圖表數據分析      | Julius AI / ChartPixel   | 自然語言問答  |
| 產品照→行銷文案    | Anyword + ChatGPT<br>案   | 理解圖片內容生成描述                                      |
| Meme 生成     | ChatGPT Custom GPT   | 100+ 模板, AI 生成文案                                |

## 3. 文件處理

| 應用       | 最佳工具  | 優勢   |
|----------|---|--|
| PDF 深度分析 | Claude  | 視覺+文字雙模態, 100 頁內最強<br><a href="#">Data Studios</a><br><a href="#">Medium</a> |
| 合約風險標註   | Icertis<br><a href="#">Icertis</a> / LexCheck<br><a href="#">Lexcheck</a> | 自動標註高/中/低風險條款<br><a href="#">Lexcheck</a>                                    |
| 學術論文問答   | Elicit + Claude   | 138M+ 論文庫<br><a href="#">Elicit</a> + 深度分析                                   |
| 多文件比較    | Claude Projects   | 上傳多份 PDF 建立知識庫   |

| 應用     | 最佳工具       | 優勢             |
|--------|------------|----------------|
| 表格數據提取 | Claude API | 輸出 JSON/CSV 格式 |

## 4. 程式開發

| 應用          | 最佳工具                         | 特色   |
|-------------|------------------------------|--|
| Code Review | Qodo / CodeRabbit            | 71.2% SWE-bench, 深度理解 codebase                   |
| 程式碼生成       | Claude Code / GitHub Copilot | 終端機整合, 自主多步驟工作流                                  |
| 除錯與重構       | Claude                       | SWE-Bench 72.5% 領先 <a href="#">IntuitionLabs</a> |
| API 文件撰寫    | ChatGPT                      | 清晰結構化輸出  |

## 5. 數據分析

| 應用         | 最佳工具  | Prompt 範例                       |
|------------|---|---------------------------------|
| Excel 公式生成 | Gemini in Sheets ( <a href="#">Google Workspace</a> ) / Copilot | 「寫一個公式計算 A 欄含『訂閱』且 B 欄>100 的總和」 |
| 數據視覺化      | ChartPixel / Julius AI  | 「為這份數據建議最適合的圖表類型」               |
| 異常值檢測      | Gemini / Copilot  | 「標出這份數據中的異常值」                   |
| 預測分析       | Gemini Advanced   | 「根據歷史數據預測下季表現」                  |

## 6. 學習教育

| 應用      | Prompt 範例                                   |
|---------|---|
| 概念學習    | 「用蘇格拉底式教學法教我量子力學,從基礎開始」                     |
| ELI5 解釋 | 「ELI5: 區塊鏈」                                 |
| 面試模擬    | 「你是 [公司名] 的 [職位] 面試官,對我進行行為面試」 (Novoresume) |
| 語言學習    | 「糾正我的英文文法並解釋錯誤原因」                           |

## 7. 商業應用

| 應用       | 最佳工具                               | 說明            |
|----------|------------------------------------|---------------|
| 簡報製作     | Copilot (PowerPoint)               | 從 Word 文件一鍵生成 |
| 會議記錄     | Copilot (Teams) / Gemini (Meet)    | 自動摘要、行動項目     |
| Email 撰寫 | Copilot (Outlook) / Gemini (Gmail) | AI 建議回覆、語氣調整  |
| 競品分析     | Grok + Gemini                      | X 即時數據 + 深度研究 |

## 8. 創意寫作

| 應用   | 最佳工具                        | 說明                             |
|------|-----------------------------|--------------------------------|
| 小說寫作 | Sudowrite / Claude (Medium) | 客製化 LLM,無內容限制 (StorytellingDB) |
| 角色發展 | Claude                      | 自然對話,風格捕捉能力強                   |
| 劇本創作 | ChatGPT                     | 結構化輸出,格式清晰                     |

| 應用   | 最佳工具   | 說明        |
|------|--------|-----------|
| 歌詞創作 | Claude | 韻律感與文學性較佳 |

## 七、價格與功能對照總表

| 模型             | 免費方案                   | 付費方案   | 最佳用途                          |
|----------------|------------------------|--|-------------------------------|
| <b>ChatGPT</b> | GPT-4o 有限使用,3<br>檔/天   | Plus \$20/月、<br><a href="#">Content Whale</a> Pro \$200/月<br><a href="#">IntuitionLabs</a> | 全方位、Custom GPTs 生態系           |
| <b>Gemini</b>  | 32K context,5 分鐘<br>影片 | AI Pro \$19.99/月、<br><a href="#">GofP</a> Ultra \$249.99/<br>月<br><a href="#">Nexos</a>    | YouTube 分析、超長文件、<br>Google 整合 |
| <b>Claude</b>  | Sonnet 4,~20 訊息/<br>天  | Pro \$20/月、Max \$100-200/月<br><a href="#">ScreenApp</a>                                    | PDF 分析、程式碼、創意寫作               |
| <b>Copilot</b> | 基本功能                   | Pro \$20/月、M365 \$30/用戶/月  | Office 重度用戶、企業                |
| <b>Grok</b>    | ~10 提示/2 小時            | Premium+ \$40/月、SuperGrok \$30/月   | X 即時數據、趨勢追蹤                   |

## 結論:選擇最適合你的 AI 工具組合

**音訊轉錄翻譯首選 Gemini**——免費版已足夠處理 10 分鐘以內的錄音,且能一步完成轉錄+翻譯,無需額外工具。ChatGPT 需透過 App 操作,[Microsoft Learn](#) Claude 和 Grok 則完全不支援原生音訊處理。[Microsoft Learn](#) [Mac Performance Guide](#)

**影片字幕生成同樣推薦 Gemini**,它是唯一支援直接上傳 MP4 並生成 SRT 字幕的主流模型,YouTube 網址直接貼上即可分析更是獨家功能。

**PDF 深度分析當選 Claude**——雙模態處理(視覺+文字)、100 頁視覺分析能力、優異的圖表理解,讓它在學術論文與合約審閱場景中無可替代。

**Office 生產力無疑選 Copilot**——若你的工作流程圍繞 Word、Excel、PowerPoint、Teams,Copilot 的 (Microsoft Learn)原生整合提供無縫體驗,是唯一能「在應用程式內」直接使用 AI 的方案。

**即時社群趨勢則看 Grok**——它對 X 平台數據的獨家存取權,使其成為社群監控、輿情分析、即時新聞追蹤的最佳選擇。

**建議採用組合策略:**以 Gemini 處理影音、Claude 分析文件、Copilot 提升 Office 生產力、ChatGPT 作為多功能後備。根據具體任務選擇最適合的工具,而非期待單一模型滿足所有需求。