

# 利用 LLMs 自動從文字產生簡報的 Web AI Application 研究報告

## 執行摘要

基於大語言模型（LLMs）的自動簡報生成技術在 2025 年達到了前所未有的成熟度。最新研究顯示，PPTAgent 框架以 97.8% 的成功率和 3.67 的整體品質評分領先業界，而 PreGenie 等多模態代理系統在視覺設計方面表現卓越。這些 breakthrough 技術結合先進的 Web 應用架構，為企業和個人用戶提供了高效、智能的簡報創作解決方案。<sup>[1][2]</sup>

請見附錄-圖 1：利用 LLMs 自動從文字產生簡報的 Web AI Application 技術架構圖

## 技術現狀與最新發展

### SOTA 模型架構分析

2025 年的簡報生成領域出現了三個主要技術範式的突破：

#### 1. 編輯導向代理系統

PPTAgent 採用革命性的兩階段編輯範式，首先分析參考簡報以理解結構模式和內容架構，然後通過程式碼行動來生成投影片。此方法避免了從零開始創建複雜簡報的挑戰，成功率從傳統方法的 10% 提升至 97.8%。<sup>[1]</sup>

#### 2. 多模態代理框架

PreGenie 框架整合了五個專門的 LLM 和 VLM，包括文字摘要器、圖像標註器、程式碼生成器等組件。該系統基於 Slidev 框架，支援 Markdown 格式的專業簡報生成，在設計品質方面獲得 3.60 分的高分。<sup>[2]</sup>

#### 3. 直接整合系統

Talk-to-Your-Slides 透過 COM 通訊直接操作 PowerPoint 物件，實現了真正的即時編輯功能。系統採用兩層架構：高層 LLM 代理解釋指令並制定編輯計劃，低層 Python 腳本直接操作 PowerPoint 對象。<sup>[1]</sup>

(請見下頁表格)

模型/框架	技術類型	核心創新	內容品質 (1-5)	設計品質 (1-5)	多模態支援	實時編輯
PPTAgent (2025)	Edit-based Agent	兩階段編輯範式 + PPTEval 評估	3.28	3.27	✓ (圖像理解+生成)	✓
PreGenie (2025)	Agentic MLLM	Slidev 框架 + 迭代視覺檢查	3.45	3.6	✓ (多模態輸入)	✓
Talk-to-Your-Slides (2025)	Direct PowerPoint Integration	COM 通訊直接操作 PowerPoint	3.2	3.4	✓ (PowerPoint 對象)	✓
PASS (2025)	Document-to-Slides Pipeline	語音合成整合 + 通用文檔支援	3.15	3.25	✓ (文檔 + 圖像)	X
Multi-Staged LLM-VLM (2024)	Multi-staged End-to-End	LLM+VLM 組合多階段處理	3.24	3.3	✓ (LLM+VLM)	X
AutoPresent (2024)	Layout-aware Code Synthesis	預定義操作 API	2.98	2.33	X (純文字)	✓
Bandyopadhyay et al. (2024)	Multi-modal Approach	敘事結構重視	3.1	3.05	✓ (文字 + 圖像)	X
Baseline GPT-4o	Direct LLM Generation	端到端文字生成	2.98	2.33	X (純文字)	X

### 核心技術創新

#### PPTEval 評估框架

首個綜合性簡報評估系統 PPTEval 引入了三維評估標準：內容品質、視覺設計和邏輯連貫性。該框架採用 MLLM-as-a-Judge 範式，與人工評估的 Pearson 相關性達到 0.71，顯著優於傳統 ROUGE 指標。<sup>[3]</sup>

#### 多階段處理管線

最新的多階段 LLM-VLM 方法將簡報生成分解為內容提取、視覺整合和迭代優化三個階段。相比直接生成方法，這種 approach 在自動化指標和人工評估中都表現更優。<sup>[4][5]</sup>

(請見附錄-圖 1：文字到簡報生成的 AI 處理流程圖)

#### Web 應用架構設計

## 前端技術棧

React 19.0 + TypeScript 作為主流選擇，提供並發渲染和 Suspense 特性。Vite 6.1+作為建構工具，支援熱模組替換和 Tree-shaking 優化。UI 組件庫推薦 Material-UI v6 或 Chakra UI v3，確保一致的用戶體驗。<sup>[6][7]</sup>

狀態管理採用 Zustand 或 React Query 處理複雜的資料同步需求。這些工具特別適合處理 AI 生成過程中的異步狀態更新和快取管理。<sup>[7]</sup>

## 後端架構

FastAPI 0.115+ 因其異步處理能力和自動文檔生成功能成為首選。支援 WebSocket 連接以實現即時的生成進度反饋，以及背景任務處理耗時的 AI 推理過程。<sup>[6][7]</sup>

資料庫設計採用 PostgreSQL 16 存儲結構化資料，配合 Redis 7.2 提供毫秒級快取服務。檔案處理使用 python-pptx 庫進行 PowerPoint 格式操作。<sup>[8][9][7][10]</sup>

## API 設計模式

RESTful 架構結合 OpenAPI 3.1 規範確保類型安全和標準化。關鍵端點包括：<sup>[7]</sup>

- POST /api/v1/presentation/generate - 主要生成接口<sup>[11]</sup>
- GET /api/v1/presentation/{id}/status - 進度查詢<sup>[11]</sup>
- PUT /api/v1/presentation/{id}/edit - 即時編輯<sup>[11]</sup>

請求參數支援多種自定義選項：語調設定、詳細程度、投影片數量、模板選擇等。<sup>[11]</sup>

## 技術實現細節

### LLM 整合策略

多模型支援架構允許開發者選擇適合的 LLM 提供者：OpenAI GPT-4o、Google Gemini、Anthropic Claude 或本地 Ollama 模型。模型路由器根據任務複雜度自動選擇最適合的模型。<sup>[11]</sup>

Prompt Engineering 採用結構化模板系統，針對不同簡報類型（商業、學術、教育）優化提示詞。少樣本學習技術提高特定領域的生成品質。<sup>[1][2]</sup>

## 容器化部署

Docker Compose 配置支援 GPU 加速和多服務編排。Docker Offload 技術允許計算密集的 AI 推理任務無縫轉移到雲端 GPU 資源。<sup>[12][13][14]</sup>

```
services:
  api:
    build: ./backend
    ports: ["8000:80"]
    environment:
      - LLM=openai
      - OPENAI_API_KEY=${OPENAI_KEY}
  frontend:
    build: ./frontend
    ports: ["3000:80"]
```

## 效能最佳化

快取策略包括 LLM 回應快取、模板快取和圖像資源快取。批次處理能力允許同時處理多個簡報請求，提高資源利用效率。<sup>[11]</sup>

CDN 整合確保全球用戶快速訪問生成的簡報檔案。自動擴展功能根據需求動態調整計算資源。<sup>[15]</sup>

## 評估與品質保證

### 評估指標體系

PPTEval 框架建立了業界首個標準化評估體系：<sup>[3]</sup>

- 內容品質：文字清晰度、資訊適當性、視覺支援度
- 設計品質：色彩協調、視覺元素、整體美感
- 邏輯連貫性：結構發展、背景資訊完整性

TSBench 資料集提供 379 個多樣化編輯指令，用於系統性評估簡報編輯代理的能力。<sup>[1]</sup>

### 品質控制機制

多層驗證包括語法檢查、格式驗證和內容審核。自我修正機制允許系統偵測執行錯誤並提供即時回饋進行迭代改進。<sup>[1]</sup>

A/B 測試框架支援不同模型和參數配置的效果比較。使用者回饋整合持續優化生成品質。<sup>[3]</sup>

## 商業應用與案例

### 企業級部署

私有化部署解決方案確保敏感資料不離開企業內部。SSO 整合和權限管理支援企業級安全需求。<sup>[16]</sup>

品牌客製化功能自動應用企業視覺識別系統，保持品牌一致性。模板庫管理允許組織建立和分享標準化簡報模板。<sup>[16]</sup>

### 垂直應用領域

教育科技：自動課程簡報生成，支援多語言和個性化學習路徑<sup>[17][18]</sup>

商業諮詢：快速提案簡報，整合資料視覺化和市場分析<sup>[19][20]</sup>

醫療健康：病例簡報和研究發表，符合專業規範和隱私要求<sup>[17]</sup>

### 技術挑戰與解決方案

#### 現有限制

複雜佈局處理：嵌套群組圖形的解析仍是瓶頸，影響結果一致性。視覺一致性：儘管有所改進，仍無法充分利用視覺提示來優化樣式一致性，可能出現元素重疊等設計缺陷。<sup>[1][21]</sup>

跨平台相容性：COM 通訊限制系統僅能在 Windows 環境運行，限制了使用者可及性。<sup>[1]</sup>

#### 解決策略

混合方法：結合底層物件存取和 UI 視覺驗證，提高佈局品質。跨平台 API：開發基於 Web 的 PowerPoint API 或平台無關的中介軟體解決方案。<sup>[1]</sup>

視覺回饋整合：納入渲染後的投影片圖像，實現更強健的視覺品質控制。<sup>[1]</sup>

### 未來發展趨勢

#### 技術演進方向

多模態融合：整合語音、視頻和互動元素的下一代簡報系統。即時協作：支援多人同時編輯和 AI 輔助的智能協作平台。<sup>[22]</sup>

個性化引擎：基於使用者偏好和歷史資料的高度個性化內容生成。跨語言支援：原生多語言生成和即時翻譯功能。<sup>[23][24]</sup>

## 產業標準化

開放格式支援：擴展至更多簡報格式，包括 Web 標準和新興格式。API 標準化：建立業界統一的簡報生成 API 規範。<sup>[11][25]</sup>

倫理 AI 整合：確保生成內容的準確性、公平性和透明度。<sup>[1]</sup>

## 技術建議與最佳實踐

### 開發建議

1. 採用微服務架構：分離 AI 推理、檔案處理和用戶介面服務，提高系統可維護性<sup>[7]</sup>
2. 實施快取策略：多層快取設計減少 LLM API 呼叫成本和延遲<sup>[11]</sup>
3. 建立監控系統：全面的效能監控和錯誤追蹤確保服務穩定性<sup>[15]</sup>

### 部署策略

1. 容器化優先：使用 Docker 和 Kubernetes 實現一致的開發和生產環境<sup>[12][13]</sup>
2. 雲原生設計：支援多雲部署和自動擴展能力<sup>[15][13]</sup>
3. 安全防護：實施 API 限流、身份驗證和資料加密<sup>[16]</sup>

## 結論

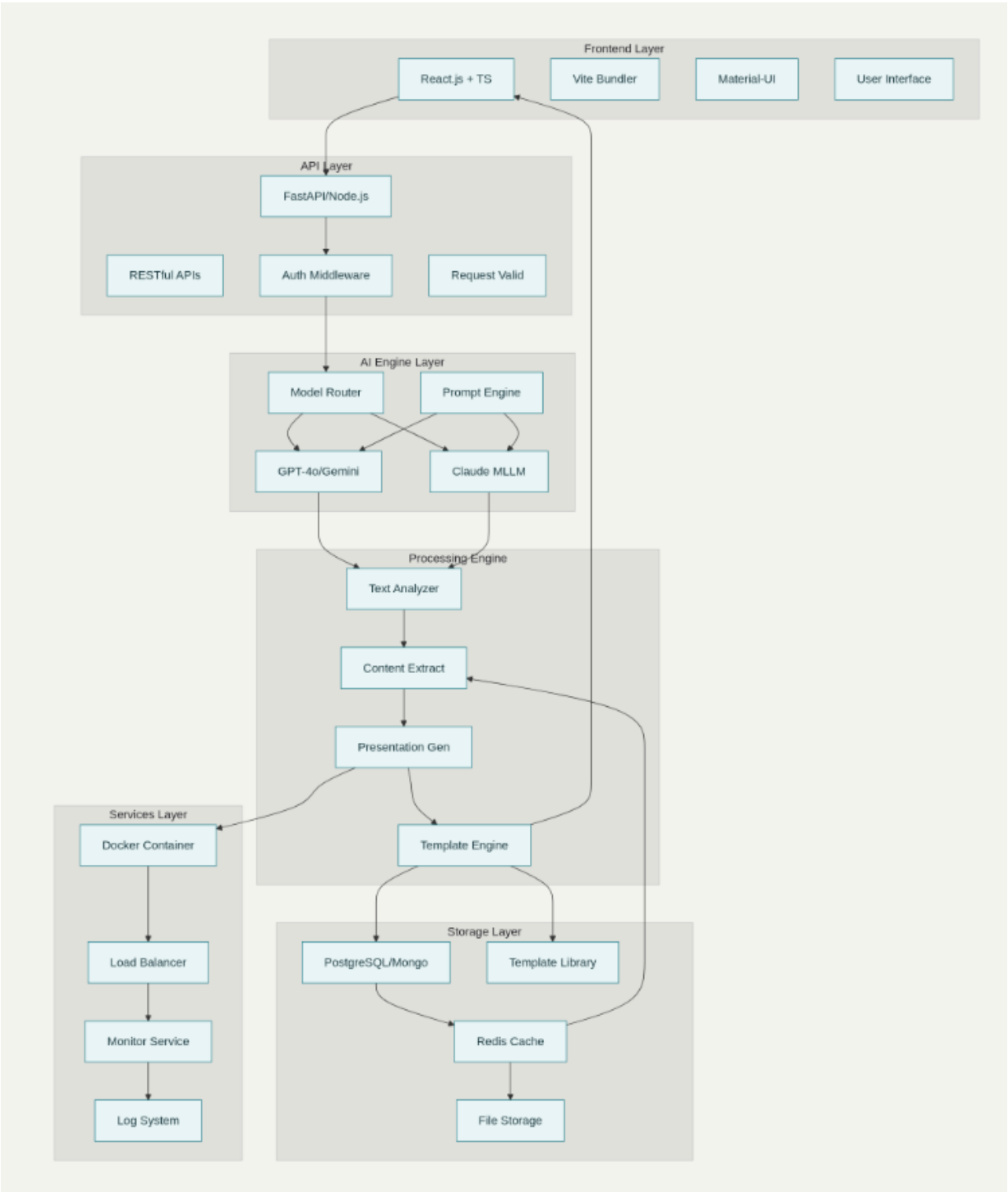
利用 LLMs 自動從文字產生簡報的 Web AI 應用已進入成熟期，以 PPTAgent 和 PreGenie 為代表的 SOTA 系統展現了令人印象深刻的技術能力。透過採用編輯導向範式、多模態代理架構和綜合評估框架，這些系統在成功率、內容品質和視覺設計方面都取得了顯著突破。

Web 應用架構的最佳實踐結合 React、FastAPI、Docker 和雲原生技術，為開發者提供了構建可擴展、高效能簡報生成服務的完整解決方案。隨著技術持續演進，預計將看到更多創新應用和商業模式的出現，推動整個產業向更智能、更人性化的方向發展。

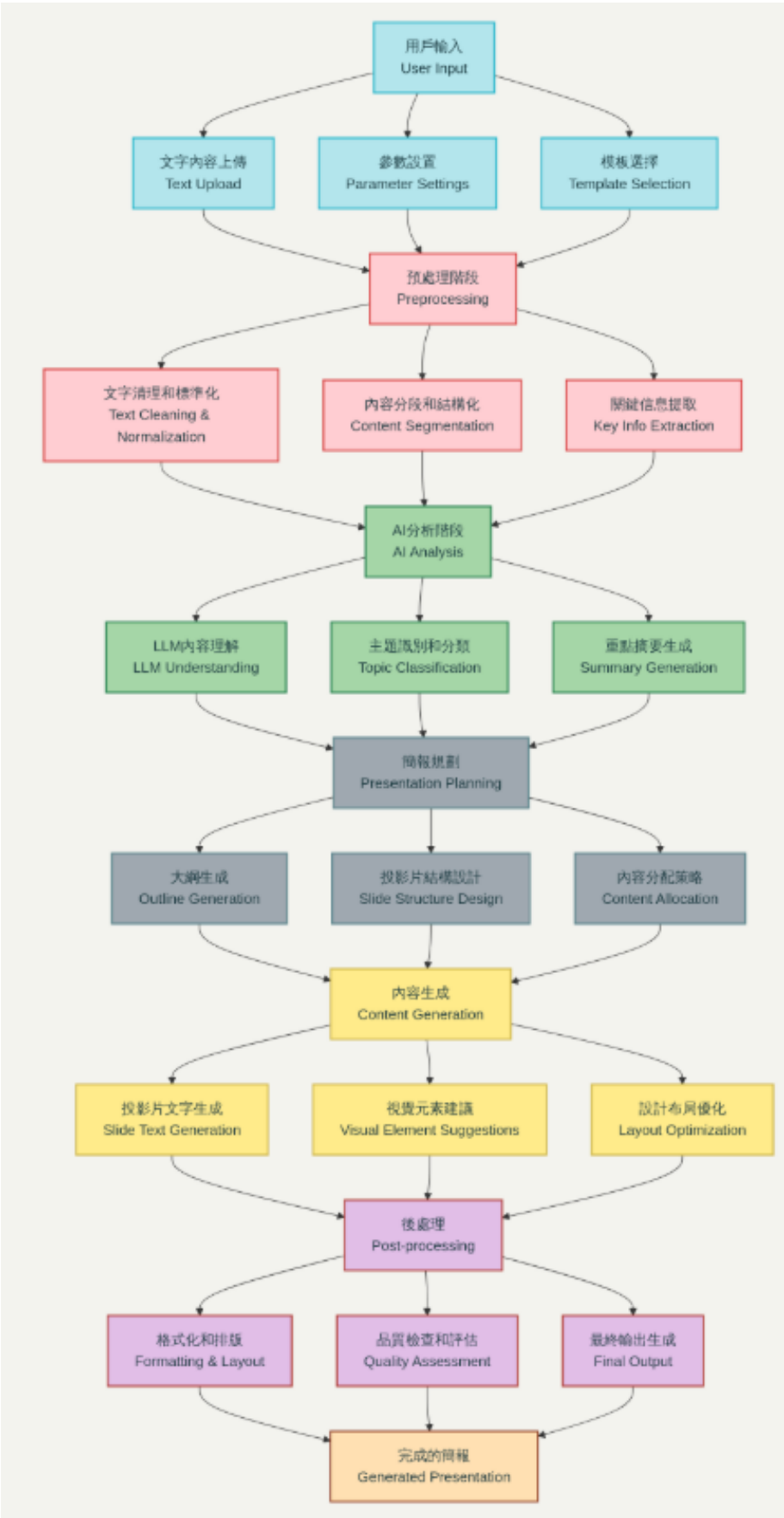
對於希望進入此領域的開發者和企業，建議採用開源框架如 Presenton 作為起點，結合最新的 LLM 技術和 Web 開發最佳實踐，構建符合特定需求的定制化解決方案。關鍵成功因素包括正確的技術架構選擇、穩健的品質評估機制，以及持續的用戶體驗優化。<sup>[11]</sup>

附錄一：圖表

圖一：利用 LLMs 自動從文字產生簡報的 Web AI Application 技術架構圖



圖二：文字到簡報生成的 AI 處理流程圖





## 附錄二：用於產生簡報檔內容的模型

### SOTA 開源圖像生成模型推薦

- Stable Diffusion XL (SDXL) 完全開源，支援多語言 Prompt，可生成寫實、扁平、卡通、插圖、線稿等各種簡報用圖片。 特色：高度可自訂，支援分辨率調整與風格選擇，生態最豐富，各大開源工具（DreamStudio、AUTOMATIC1111 等）均可快速集成。
- DeepFloyd IF 具備文本理解力強，對簡報主題、標語、文字元素生成品質佳，適合需要結合教學重點、圖表說明等設計情境。
- OpenJourney 風格化美術表現力突出，適合教育簡報、STEM/STEAM 主題封面、活動海報等。
- DreamShaper SD 衍生版本，主打精細美學、通用插圖，由開源社群持續更新，常用於教師創意簡報視覺設計。
- Chroma、Neta-Lumina 2025 年新興模型，強調通用性與運算效率，高度自訂但目前資源相對較少，適合多元風格開發。
- FLUX.1、HunyuanImage-2.1 近年大量封測的通用型開源圖像生成，支援多領域、教育及展示用途