

TXT2PPTX實作與研究-Perplexity

Now I have comprehensive research data. Let me compose the final answer.

「文字→PPTX」LLM 簡報自動生成：問題分析、相關論文與開源專案全景

這是一個典型的**多模態結構化內容生成**問題，橫跨自然語言處理、版面設計、圖像檢索/生成等多個研究領域。以下針對你提出的五大核心問題，逐一梳理最相關的學術論文與開源專案，並進行架構層面的分析。

*

一、問題總覽與研究版圖

自動簡報生成（Automatic Presentation Generation）的 pipeline 可拆解為：

1. **內容擴充** (Content Expansion / Summarization)
2. **內容→分頁分配** (Content-to-Slide Allocation)
3. **文字排版定位** (Text Layout Positioning)
4. **圖片搜尋/生成** (Image Retrieval / Generation)
5. **圖文聯合排版** (Joint Text-Image Layout)

現有研究大致分為三個典範：**規則式 (rule-based)**、**模板式 (template-based)**、**編輯式 (edit-based)** 以及最新的 **Agent 式 (agentic)**。以下按照你的五個問題進行深度分析。

二、核心論文與對應問題映射

論文 / 專案	發表場所	主要解決的問題	核心方法
PPTAgent (Zheng et al.)	EMNLP 2025	②③④⑤ 全流程	兩階段 edit-based：分析參考簡報 → 編輯式生成
PreGenie (Xu et al.)	EMNLP 2025 Findings	①②③⑤ 全流程	Slidev + 多 MLLM 協作 + Code Review + Page Review
DocPres (Bandyopadhyay et al.)	INLG 2024	①② 內容擴充與分配	多階段分層摘要 + Slide-to-Section 映射 + CLIP 圖配
DOC2PPT (Fu et al.)	AAAI 2022	①②③⑤	階層式 seq2seq + 改寫模組 + Layout Prediction MLP
D2S (Sun et al.)	NAACL 2021	①② 內容選取	Query-based 摘要 + 圖片自適應選取
SlideCoder (Tang et al.)	EMNLP 2025	③ Layout ⑤ 還原	Layout-aware RAG + 色彩梯度分割 + 階層式代碼生成
SlideSpawn (Kumar et al.)	arXiv 2024	①②④	MLP 句子重要度 + ILP 選取 + 聚類標題分組
LayoutNUWA (Tang et al.)	ICLR 2024	③ Layout ⑤ 生成	Layout → HTML Code Generation → LLM 填充
PosterLlama (Seol et al.)	ECCV 2024	⑤ Content-aware Layout	VLM + HTML 排版 + 深度增強策略
		⑤ Content-	

SEGA (Wang et al.)	ICCV 2025	aware Layout	粗粒度→細粒度兩階段演進 + 設計先驗知識
LayoutTransformer (Gupta et al.)	ICCV 2021	③⑤	Self-attention 學習 layout 元素 間的上下文關係
LayoutGAN (Li et al.)	ICLR 2019	③⑤	GAN + wireframe discriminator 直接生成結構化排版
PPTC Benchmark (Guo et al.)	arXiv 2023	全流程評估	279 多輪指令 Benchmark，測試 LLM 操控 PPTX 能力
Slide Retrieval Study (Giouroukis et al.)	arXiv 2025	④ RAG 投影 影片檢索	ColPali / Caption-based / Hybrid 混合檢索比較

*

三、針對五大問題的深度分析

問題一：內容擴充

這是從簡短輸入文字擴展成完整簡報內容的問題，本質上是**受控文本生成 (Controlled Text Generation)** 和**多層摘要的逆過程**。

DocPres 是一個半結構化的簡報，為英文文件的「類別」(Category)——它是一串的一致性資訊，包含有 Class of Image 的類別名稱與其相關的子類別。這些資訊被用來解釋每個段落的內容，並對其進行分類。

PPTAgent 走不同路線：它不從零擴充，而是以「文件分段 + 參考簡報 schema 匹配」的方式，將文件內容對應到已有的投影片結構中。Stage I 會提取參考簡報的 schema (每頁的 Category/Description/Data 結構)，Stage II 的 Outline Generation 再將文件段落映射到對應的參考投影片。^[3]

PreGenie 則用 Text Summarizer (LLM) 提取關鍵資訊 (摘要、標題、作者等)，搭配 Image Captioner (VLM) 為圖片標記標題與描述，兩者共同組成完整的內容素材。^[5]

建議方案：採用 DocPres 的分層摘要策略作為內容擴充的骨幹，先建立全文的結構化理解，再讓 LLM 以 CoT 方式逐步展開各主題要點。

問題二：擴充後文字如何分配到各分頁

這是文件-投影片結構映射 (Document-to-Slide Structural Mapping) 的問題。

DOC2PPT 用更底層的方式解決：它設計了一個三層階層式 RNN (PTsec → PTslide → PTobj)，模型自動學習文件章節 → 投影片 → 頁面元素的對應關係，並透過 attention 機制決定哪些句子歸到哪頁。[7]

PPTAgent 的 Outline Generation 模块由 LLM 生成結構化大綱，每個大綱條目將確定「參考哪段投影片」和「使用文件的哪些段落」，形成一對一的映射關係。

SlideSpawn 使用 ILP (整數線性規劃) 來選取句子，以最大化重要性分數為目標，同時約束字數上限和句子間冗餘度，再用語意聚類將選中句子分組為不同投影片。[9]

建議方案：先用 LLM 生成投影片大綱（標題列表），再用語意相似度將文件段落映射到各標題。PPTAgent 的 schema-guided outline 或 DocPres 的 section mapping 都是可行路線。

問題三：每頁文字排版定位

這是版面生成 (Layout Generation) 的核心子問題。

LayoutNUWA 首創將 layout 生成視為代碼生成任務：把版面元素量化為 HTML/SVG 的 `<rect>` 標籤，用 LLM 透過 Code Instruct Tuning 預測各元素的 x, y, width, height。在多個 dataset 上達到 50%+ 的性能提升。[11]

DOC2PPT 的 Layout Prediction 模組用兩層 MLP 預測每個投影片物件的 `{lx, ly, lw, lh}` (位置和尺寸)，輸入是物件層級的隱藏表示和章節級別的 attention 向量。[8]

PPTAgent 視頻了從零生成 layout 的困難——它從參考投影片 `read_ppt`，通過 API (`replace_slide`, `add_slide`, `close_paragraph` 等) 修改內容並從頭開始佈局。HTML 這張圖讓 LLM 從結構拆分元素位置。

SlideCoder 提出了 Color Gradient-based Segmentation 演算法，先從參考圖像分割出版面區塊，再用 Hierarchical RAG 為各區塊生成對應的代碼。它還發布了 SlideMaster (7B 模型) 用於投影片代碼生成。[13]

LayoutTransformer 用 self-attention 學習 layout 元素之間的上下文關係，支援自回歸生成（從空集合或 partial layout 完成），且能自動捕捉元素的語義屬性。

建議方案：如果使用模板/參考投影片，PPTAgent 的 edit-based 方式最穩定。如需從零生成，LayoutNUWA 的 HTML 代碼生成範式與你的 python-pptx 工作流最相容，可將 HTML 座標轉換為 PPTX 座標。

問題四：搜尋或生成適合每頁的圖片

這涉及**圖像檢索 (Image Retrieval)** 和**圖像生成 (Image Generation)** 兩個方向。

DOC2PPT 引入了 text-figure matching objective，鼓勵相關的文字-圖片配對出現在同一頁投影片上，透過訓練讓模型學會圖文配對。^[7]

PreGenie 在純文字輸入的情境下，用 LLM 生成圖像提示詞，呼叫外部 text-to-image 模型產生插圖；在文件輸入場景下，Image Captioner (VLM) 為每張圖生成標題、描述和原文位置，供後續匹配。

PPTAgent V2 在最新版本加入了 AI 圖像生成、Web 搜尋和 Agent Sandbox (30+ 工具)，可自主建立視覺素材。^[16]

投影片檢索研究 (Giouroukis et al., 2025) 系統比較了多種 RAG 方式：ColPali (視覺 late-interaction) 達到 86.9% NDCG@10 但存儲需求大；VLM captioning + hybrid retrieval 在精度與效率間取得最佳平衡 (83.9% NDCG@10，存儲需求少 67%)。^[18]

建議方案：採用 CLIP 相似度做圖片匹配 + text-to-image model 做圖片生成的雙軌策略。RAG 部分可參考 captioning-based 檢索方式。

問題五：圖文聯合排版——最大難點

這是**Content-Aware Layout Generation** 問題，也是學界目前最活躍的研究方向。

PosterLlama (ECCV 2024) 將版面元素格式化為 HTML code，利用 VLM 理解背景圖的語意後生成排版，搭配深度增強策略保證視覺美觀。支持無條件生成、條件生成、Layout 補全等多種模式。[20]

SEGA (ICCV 2025) 提出 Stepwise Evolution 範式：先用粗粒度模組估計初步排版，再用精調模組迭代細化。關鍵創新是將**設計先驗知識 (design prior)** 注入模型，顯著提升複雜版面的成功率。[22]

PreGenie 的 Page Review 機制最直接解決圖文配合問題：Code Generator 先生成初版排版，Page Reviewer (VLM) 遵真檢查渲染結果——發現圖片溢出就調整尺寸、發現文字堆疊就改為分欄佈局、發現留白不均就重新組織為 bullet point 格式。平均需要 3.8 輪視覺審查迭代。

Illustration Layout for Slides (ACM MM 2025) 專門處理投影片中插圖的佈局問題，用 pixel-based diffusion model 預測插圖應該放置的位置和大小，是目前唯一直接針對投影片插圖排版的研究。[24]

建議方案：圖文聯合排版建議採用**兩階段策略**——先用 LLM/LayoutNUWA 生成初步版面，再用 VLM (如 Qwen2.5-VL) 對渲染結果做視覺審查並迭代修正。PreGenie 的 Code Review + Page Review 雙迴圈是目前效果最好的工程方案。

*

四、推薦的開源專案

專案	Stars	特色	適用場景
PPTAgent	~2,000+	V2 支援 Deep Research + AI 圖片生成 + 30+ 工具沙箱	最完整的端到端方案
SlideCoder	—	Layout-aware RAG + SlideMaster 7B 模型	Layout 還原與代碼生成
txt2pptx	—	LangChain + RAG on python-pptx docs	輕量快速原型
Powerpointer-For-Local-LLMs	—	本地 LLM + python-pptx，7 種設計模板	本地部署、隱私優先
python-pptx	3,200+	PPTX 讀寫基礎庫	底層 PPTX 操控
Layout-Generation	—	LayoutVAE + LayoutTransformer + LayoutGAN 基線實現	版面生成研究
LayoutNUWA	—	Layout 作為 Code Generation 任務	學術研究 / Layout 模組嵌入

PPTAgent V2 是目前功能最完整的開源方案，其 Agent 架構包含 Planner、Content Organizer、Editor、Coder、Layout Selector 等多個角色，值得作為你的系統架構參考。[26]

*

五、針對你的 RAG 應用的架構建議

綜合所有論文的方法論，建議你的系統採用以下 pipeline：



關鍵工程建議：

- 中間表示格式：PPTAgent 將 HTML 與 PPTX 的 XML，腳本與 LLM 理解對應 (PPT XML 彙總上平行，HTML 表示可靈活互動數十行)
- Self-correction 優勢：PPTAgent 和 ProGenie 都證明了 R2P1 式的迭代修正對提升成功率至關重要 (PPTAgent 從 74.0% 提升到 93%)
- 視覺驗證迴圈：ProGenie 的實驗證明 Page Review 在設計美觀和頁面一致性上帶來最顯著提升，是區分「能用」與「好用」的關鍵
- 評估標準：建議採用 PPTLevel 的三維度評估 (Content, Design, Cohherence)，搭配 CLIP/LongCLIP 評估函數相關性

來源

[1] [Literature Review] Enhancing Presentation Slide Generation by ...

<https://www.themoonlight.io/en/review/enhancing-presentation-slide-generation-by-langs-with-a-multi-staged-end-to-end-approach>

[2] Enhancing Presentation Slide Generation by LLMs with a ... <https://arxiv.org/abs/2406.06556>

[3] PPTAgent: Generating and Evaluating Presentations ... <https://arxiv.org/html/2501.03936v3>

[4] [Literature Review] PPTAgent: Generating and Evaluating ...
<https://www.themoonlight.io/en/review/pptagent-generating-and-evaluating-presentations-beyond-text-to-slides>

[5] PreGenie: An Agentic Framework for High-quality Visual ... <https://arxiv.org/html/2505.21660v2>

[6] Paper page - PreGenie: An Agentic Framework for High-quality Visual Presentation Generation
<https://huggingface.co/papers/2505.21660>

[7] DOC2PPT: Automatic Presentation Slides Generation from ... <https://arxiv.org/abs/2101.11796>

[8] DOC2PPT: Automatic Presentation Slides Generation from ...
https://tsujuifu.github.io/pubs/aaai22_doc2ppt.pdf

[9] [Revue de papier] SlideSpawn: An Automatic Slides Generation System for Research Publications
<https://www.themoonlight.io/fr/review/slidespawn-an-automatic-slides-generation-system-for-research-publications>

[10] SlideSpawn: An Automatic Slides Generation System for ... <https://arxiv.org/abs/2411.17719>

[11] LayoutNUWA: Revealing the Hidden Layout Expertise of Large Language Models - Microsoft Research
<https://www.microsoft.com/en-us/research/publication/layoutnuwa-revealing-the-hidden-layout-expertise-of-large-language-models/>

[12] LayoutNUWA: Revealing the Hidden Layout Expertise of ... <https://arxiv.org/abs/2309.09506>

[13] SlideCoder: Layout-aware RAG-enhanced Hierarchical Slide ... <https://aclanthology.org/2025.emnlp-main.458/>

[14] SlideCoder: Layout-aware RAG-enhanced Hierarchical Slide Generation from Design
<http://arxiv.org/abs/2506.07964>

[15] Layout Generation and Completion with Self-attention <https://arxiv.org/abs/2006.14615>

[16] icip-cas/PPTAgent: An Agentic Framework for Reflective ... <https://github.com/icip-cas/PPTAgent>

[17] Releases · icip-cas/PPTAgent <https://github.com/icip-cas/PPTAgent/releases>

[18] [Literature Review] What's the Best Way to Retrieve Slides? ...
<https://www.themoonlight.io/en/review/whats-the-best-way-to-retrieve-slides-a-comparative-study-of-multimodal-caption-based-and-hybrid-retrieval-techniques>

[19] What's the Best Way to Retrieve Slides? A Comparative ... <https://arxiv.org/abs/2509.15211>

[20] PosterLlama: Ability to design language models and ... <https://ai-scholar.tech/en/articles/layout-llama>

gen/posterllama

- [21] PosterLlama: Bridging Design Ability of Language Model to ... <https://arxiv.org/abs/2404.00995>
- [22] A Stepwise Evolution Paradigm for Content-Aware Layout ...
<https://iccv.thecvf.com/virtual/2025/poster/1121>
- [23] A Stepwise Evolution Paradigm for Content-Aware Layout ... <https://arxiv.org/abs/2510.15749>
- [24] Illustration Layout Generation for Slide Enhancement with ...
<https://dl.acm.org/doi/abs/10.1145/3746027.3754818>
- [25] Accepted Regular Papers <https://acmmm2025.org/accepted-regular-papers/>
- [26] PPTAgent 代码库整体结构：设计理念，核心模块，数据流与 ...
https://blog.csdn.net/qq_42540492/article/details/149483549
- [27] [Literature Review] Slide2Text: Leveraging LLMs for ...
<https://www.themoonlight.io/en/review/slides2text-leveraging-llms-for-personalized-textbook-generation-from-powerpoint-presentations>
- [28] SlidesAI: Free AI Presentation Maker & Google Slides Generator <https://www.slidesai.io>
- [29] Convert Text to PPT Online | AI Text to PowerPoint Generator <https://www.presentations.ai/tools/text-to-ppt>
- [30] Evaluating Large Language Models for PowerPoint Task Completion
<https://huggingface.co/papers/2311.01767>
- [31] AI Presentation Maker | Create Slides & Decks in Minutes <https://www.beautiful.ai/presentation-maker>
- [32] [P] We finally got Text-to-PowerPoint working!! (Generative AI for Slides ✨)
https://www.reddit.com/r/MachineLearning/comments/zxab4m/p_wefinally_got_texttopowerpoint_working/
- [33] Automated Powerpoint Presentation Generation from PDF ...
<https://www.ijfmr.com/papers/2024/6/31042.pdf>
- [34] Smallppt: AI PowerPoint Generator & Slideshow Maker <https://smallppt.com>
- [35] Text to Presentation - AI Slides from Text Input <https://presenti.ai/text-to-presentation/>
- [36] Build a PowerPoint Generator AI Agent with LLMs with an IITian | ML/AI Project
<https://www.youtube.com/watch?v=xgSGGFDzu98>
- [37] AI presentation slide designer with Copilot in PowerPoint <https://powerpoint.cloud.microsoft/create/en/ai-presentation-designer/>
- [38] The 8 best AI presentation makers in 2026 <https://zapier.com/blog/best-ai-presentation-maker/>
- [39] Free AI Presentation Maker <https://slidesgo.com/ai/presentation-maker>
- [40] The Top 10 Best AI Presentation Makers in 2026 <https://www.beautiful.ai/blog/best-ai-presentation-makers>

- [41] [AAAI'22] DOC2PPT <https://doc2ppt.github.io>
- [42] Generating and Evaluating Presentations Beyond Text-to ... <https://arxiv.org/html/2501.03936v1>
- [43] D2S: Document-to-Slide Generation Via Query-Based Text ... <https://aclanthology.org/2021.naacl-main.111.pdf>
- [44] PPTAgent: Generating and Evaluating Presentations ... <https://arxiv.org/abs/2501.03936>
- [45] Automatic Slide Generation for Scientific Papers <https://ceur-ws.org/Vol-2526/paper2.pdf>
- [46] DOC2PPT: Automatic Presentation Slides Generation from Scientific Documents <http://arxiv.org/pdf/2101.11796.pdf>
- [47] Related papers: PPTAgent: Generating and Evaluating Presentations Beyond Text-to-Slides https://fugumt.com/fugumt/paper_check/2501.03936v1_enmode
- [48] SlideSpawn: An Automatic Slides Generation System for ... <https://arxiv.org/pdf/2411.17719.pdf>
- [49] DOC2PPT: Automatic Presentation Slides Generation from Scientific ... <https://aaai.org/papers/00634-doc2ppt-automatic-presentation-slides-generation-from-scientific-documents/>
- [50] Daily Paper Cast | PPTAgent: Generating and Evaluating Presentations Beyond Text-to-Slides <https://dailypapercast.transistor.fm/episodes/pptagent-generating-and-evaluating-presentations-beyond-text-to-slides>
- [51] Illustration Layout Generation for Slide Enhancement with ... <https://dl.acm.org/doi/10.1145/3746027.3754818>
- [52] D2S: Document-to-Slide Generation Via Query-Based Text ... <https://arxiv.org/abs/2105.03664>
- [53] SEGA: A Stepwise Evolution Paradigm for Content-Aware ... https://openaccess.thecvf.com/content/ICCV2025/papers/Wang_SEGA_A_Stepwise_Evolution_Paradigm_for_Content-Aware_Layout_Generation_with_ICCV_2025_paper.pdf
- [54] blackbyte7/txt2pptx: AI Agent based tool to convert text ... <https://github.com/blackbyte7/txt2pptx>
- [55] LayoutGAN: Generating Graphic Layouts with Wireframe Discriminators <http://arxiv.org/abs/1901.06767>
- [56] Best AI Presentation Maker for Professional Decks | Beautiful ... <https://www.beautiful.ai>
- [57] scanny/python-pptx: Create Open XML PowerPoint ... <https://github.com/scanny/python-pptx>
- [58] Layout Generation and Baseline implementations <https://github.com/Layout-Generation/layout-generation>
- [59] Building Text to PPT Generation App using Generative AI https://www.youtube.com/watch?v=DLV_o1FiCKY
- [60] [LG] LayoutGAN: Generating Graphic Layouts with Wireframe Discriminators <https://bloomberry.github.io/LayoutGAN/>
- [61] SlideWizard <https://app.slidewizard.ai>

- [62] A beautiful powerpoint generator which uses GPT 3.5 to ... <https://github.com/CyberTimon/Powerpointer>
- [63] CyberTimon/Powerpointer-For-Local-LLMs <https://github.com/CyberTimon/Powerpointer-For-Local-LLMs>
- [64] LAYOUTGAN: GENERATING GRAPHIC LAYOUTS
<https://openreview.net/pdf/6bea64344b5868c50f04a043422f06436b39c251.pdf>
- [65] An Agentic Framework for High-quality Visual Presentation ... <https://aclanthology.org/2025.findings-emnlp.165/>
- [66] PosterLlama: Bridging Design Ability of Language Model to Content ... <https://lait-cvlab.github.io/PosterLlama/>
- [67] PreGenie: An Agentic Framework for High-quality Visual ... <https://arxiv.org/abs/2505.21660>
- [68] Versatile Advertising Poster Layout Generation with MLLMs <https://openreview.net/forum?id=viX7rUMzwg>
- [69] Toolerific.AI <https://toolerific.ai/ai-tools/opensource/icip-cas-PPTAgent>
- [70] PreGenie: An Agentic Framework for High-quality Visual ... <https://aclanthology.org/2025.findings-emnlp.165.pdf>
- [71] PosterO: Structuring Layout Trees to Enable Language ...
https://openaccess.thecvf.com/content/CVPR2025/papers/Hsu_PosterO_Structuring_Layout_Trees_to_Enable_Language_Models_in_Generalized_CVPR_2025_paper.pdf
- [72] Activity · icip-cas/PPTAgent <https://github.com/icip-cas/PPTAgent/activity>
- [73] PreGenie: An Agentic Framework for High-quality Visual ... <https://arxiv.org/html/2505.21660v1>
- [74] SEGA: A Stepwise Evolution Paradigm for Content-Aware ... <https://arxiv.org/html/2510.15749v1>
- [75] SlideCraft: Synthetic Slides Generation for Robust Slide Analysis <https://travisseng.github.io/slidecraft/>
- [76] An Automatic Slides Generation System for Research Publications
<https://www.emergentmind.com/papers/2411.17719>
- [77] arXiv:2502.15412v1 [cs.CL] 21 Feb 2025 <https://arxiv.org/pdf/2502.15412.pdf>
- [78] [論文評述] What's the Best Way to Retrieve Slides? A ... <https://www.themoonlight.io/tw/review/whats-the-best-way-to-retrieve-slides-a-comparative-study-of-multimodal-caption-based-and-hybrid-retrieval-techniques>
- [79] Create professional slide layouts with Designer <https://support.microsoft.com/en-us/office/create-professional-slide-layouts-with-designer-53c77d7b-dc40-45c2-b684-81415eac0617>
- [80] 2411.17719v1 - Artificial Intelligence <https://www.scribd.com/document/860777098/2411-17719v1>
- [81] Multi-modal RAG on slide decks <https://blog.langchain.com/multi-modal-rag-template/>
- [82] 1 <http://arxiv.org/pdf/2411.17719.pdf>

- [83] What's the Best Way to Retrieve Slides? A Comparative ... <https://arxiv.org/html/2509.15211v1>
- [84] PosterLlama: Ability to design language models and generate content-aware layouts <https://aischolar.tech/en/articles/layout-gen%2Fposterllama>
- [85] SEGA: A Stepwise Evolution Paradigm for Content-Aware Layout ...
https://openaccess.thecvf.com/content/ICCV2025/supplemental/Wang_SEGA_A_Stepwise_ICCV_2025_supplemental.pdf
- [86] 2506.07964 <https://papers.cool/arxiv/2506.07964>
- [87] PosterLlama: Bridging Design Ability of Langauge Model to ... <https://chatpaper.com/paper/98115>
- [88] SEGA: A Stepwise Evolution Paradigm for Content-Aware ... <https://brucew91.github.io/SEGA.github.io/>
- [89] SlideCoder: Layout-aware RAG-enhanced Hierarchical Slide Generation... <https://openreview.net/forum?id=pml6dgJ6gs>