

PosterGen: Aesthetic-Aware Paper-to-Poster Generation via Multi-Agent LLMs

Zhilin Zhang^{1,2†}, Xiang Zhang^{3†}, Jiaqi Wei⁴, Yiwei Xu⁵, Chenyu You¹

zz10068@nyu.edu, xzhang23@ualberta.ca, jiaqi.wei@zju.edu.cn, ywxustat@ucla.edu,
chenyu.you@stonybrook.edu

¹Stony Brook University, ²New York University, ³University of British Columbia,

⁴Zhejiang University, ⁵University of California, Los Angeles

†Equal contribution

Multi-agent systems built upon large language models (LLMs) have demonstrated remarkable capabilities in tackling complex compositional tasks. In this work, we apply this paradigm to the paper-to-poster generation problem, a practical yet time-consuming process faced by researchers preparing for conferences. While recent approaches have attempted to automate this task, most neglect core design and aesthetic principles, resulting in posters that require substantial manual refinement. To address these design limitations, we propose **PosterGen**, a multi-agent framework that mirrors the workflow of professional poster designers. It consists of four collaborative specialized agents: (1) *Parser* and *Curator* agents extract content from the paper and organize storyboard; (2) *Layout* agent maps the content into a coherent spatial layout; (3) *Stylist* agents apply visual design elements such as color and typography; and (4) *Renderer* composes the final poster. Together, these agents produce posters that are both semantically grounded and visually appealing. To evaluate design quality, we introduce a vision-language model (VLM)-based rubric that measures layout balance, readability, and aesthetic coherence. Experimental results show that PosterGen consistently matches in content fidelity, and significantly outperforms existing methods in visual designs, generating posters that are presentation-ready with minimal human refinements.

Github: <https://github.com/Y-Research-SBU/PosterGen>

Website: <https://Y-Research-SBU.github.io/PosterGen>

Corresponding Author: chenyu.you@stonybrook.edu



1 Introduction

Academic communication typically falls into three types of mediums: papers, oral presentations, and posters. Among these options, the poster is often not a popular option. Many researchers find the process of creating a poster to be time-consuming and troublesome. They often suffer from spending significant efforts on the deliberate design of an academic poster. However, the importance of a good poster is often underestimated. Compared to an oral presentation, a poster can more effectively attract an audience that is specifically interested in the research work. It also provides a medium for direct, one-on-one dialogue between the audience and the authors (Faulkes, 2021). Therefore, an automatic, end-to-end method for generating academic posters is important to address this challenge, and promote more efficient communication during scientific conferences.

The automatic generation of commercial posters has received significant research attention (Chen et al., 2025a,b; Gao et al., 2025). In contrast, research on academic poster generation is less explored. Early works on academic

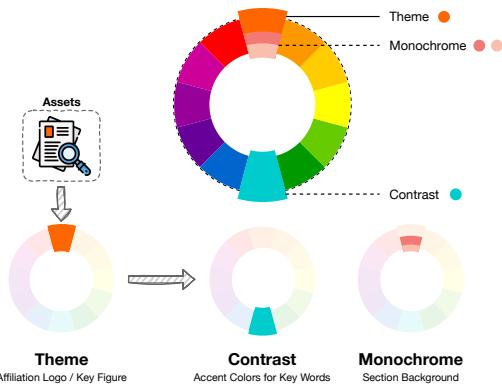


Figure 1 Overview of the color palette generation. A primary color is extracted from a source image to get monochromatic and contrast colors.

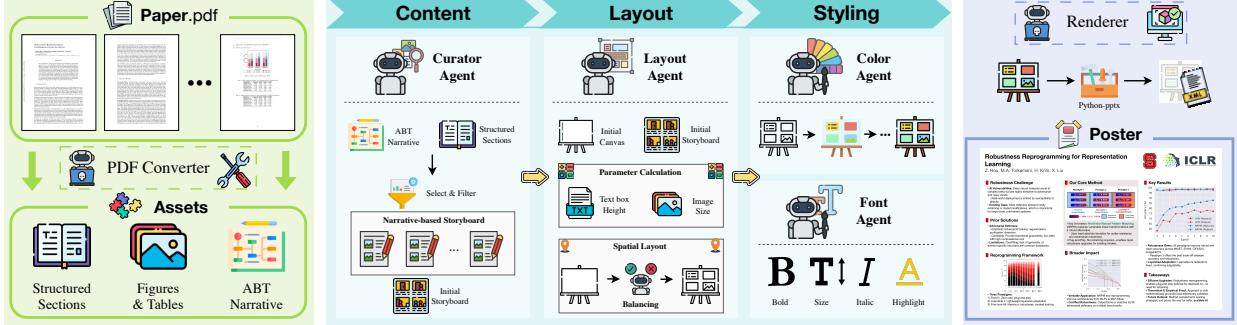


Figure 2 Overview of the **PosterGen** multi-agent framework. The process consists of three main stages: (1) The **Parser Agent** processes the input paper, extracting all text and visual assets and organizing them into a structured format focusing on an ABT narrative. (2) A series of design agents then transform this content into a styled layout: the **Curator Agent** creates a narrative-based storyboard, the **Layout Agent** calculates the precise spatial arrangement and balances the columns, and the **Styling Agents** apply a harmonious color palette and a hierarchical typographic system. (3) Finally, the **Renderer** module takes the styled metadata and produces the output poster.

posters often used neural models (Qiang et al., 2016, 2019; Xu and Wan, 2022). Other works focused on specific sub-tasks like layout generation (Wang et al., 2024) or text summarization (Saxena et al., 2025; Liu et al., 2022). These methods often produce posters with quality issues, such as severe visual overlap (Yao et al., 2025), and require further manual adjustment. Recently, LLM-powered multi-agent systems have shown strong performance on solving complex tasks. P2P (Sun et al., 2025) and PosterAgent (Pang et al., 2025) were the first to apply this approach to academic poster generation. However, their works do not sufficiently consider aesthetics and design principles, e.g., well-organized layout design that ensures natural reading flow, and styling choices for color and typography to present visual hierarchy.

In this work, we propose a new multi-agent framework that is guided by the design principles, which is further discussed in Section 4. Our multi-agent system adopts a workflow of specialist agents that mirrors a professional design process. It begins with a **Parser Agent** that extracts and structures all content from the source paper. Following this, the **Curator Agent** designs a narrative-based storyboard, which the **Layout Agent** then transforms into a spatially balanced, three-column layout. The **Styling Agents** subsequently apply a harmonious color palette and a hierarchical typographic system to ensure aesthetic coherence. This methodology is designed to generate a well-designed poster that minimizes the need for manual fine-tuning.

We further evaluate PosterGen against state-of-the-art method via VLM-as-Judge metrics. The evaluation leads to several key findings: (i) Text-to-image generation methods (GPT-4o), frequently suffer from content hallucination and fails to maintain fidelity to the source paper. (ii) Our PosterGen framework achieves content quality comparable to the human design and surpass state-of-the-art multi-agent method with significant and consistent improvements across a wide range of design and aesthetic metrics. (iii) The quantitative and qualitative results confirm that our design-centric approach is highly effective and capable of producing visually appealing and presentation-ready posters.

Our main contributions can be summarized as follows:

- We propose **PosterGen**, the first aesthetic-aware multi-agent framework for academic poster generation that embeds core design principles into multi-agent workflow.
- We introduce a comprehensive, VLM-based evaluation rubric specifically designed to assess the aesthetic and functional quality of generated posters, covering aspects from layout and color to typography.
- We provide a thorough evaluation demonstrating that our design-centric approach significantly outperforms existing methods in visual quality.

Aesthetic-driven Poster Layout

Visual and Design principles to follow for Poster Layout Design:

1. Multi-Column Layout
2. Left-aligned Titles
3. Eye-level Visual Anchor
4. Separate by White Space
5. Avoid distracting elements



Figure 3 A comparison of different layout structures. The vertically unaligned layout (left) results in a chaotic reading flow. In contrast, the vertically aligned grid (right), adopted by our Layout Agent, establishes a natural reading flow and places the visual anchor at eye-level for emphasis.

2 Related Work

Poster Generation. Recent research has explored the automatic generation of artistic and product posters in a broad sense. For example, some works utilize modular (Chen et al., 2025a) or unified (Chen et al., 2025b) frameworks to achieve a high aesthetic quality in generated posters. Other methods focus on precise generation control, such as layout structure control using language models (Seol et al., 2024), text accuracy (Gao et al., 2025), or handling multiple user-provided conditions (Zhang et al., 2025a), all of which excel at creating visually appealing posters for art or marketing purposes. However, a scientific poster has a different goal than an artistic poster, as its primary function is to convey complex research ideas and results precisely and clearly within a limited physical space. Early works used neural models to generate posters from papers (Qiang et al., 2016, 2019; Xu and Wan, 2022). Recently, (Yao et al., 2025; Wang et al., 2024; Saxena et al., 2025) proposed several benchmarks for this task; however, their methods suffer from several limitations, such as severe visual overlap (Yao et al., 2025), and restrictions to layout generation (Wang et al., 2024) or text summary (Saxena et al., 2025) only.

Recent studies show that LLM-powered multi-agent frameworks can outperform single models on complex multimodal tasks (Guo et al., 2024; Li et al., 2023; Yin et al., 2023; Liu et al., 2023; Jin et al., 2024; Zhang et al., 2024a,b; Wu et al., 2024; Cao et al., 2025a) by letting agents take on specialized roles and coordinate through mechanisms like self-reflection (Bo et al., 2024; Wei et al., 2025). P2P (Sun et al., 2025) and PosterAgent (Pang et al., 2025) were the first works to apply this multi-agent solution to scientific poster generation. However, these methods lack a thorough consideration of design principles and aesthetics, and require extensive manual adjustments before they are ready for use in a conference poster session.

3 Design Principles

Academic posters are visual communication tools that require deliberate design to initiate conversations (Faulkes, 2021). Our framework embeds four core principles, i.e., narrative, layout, color, and typography into agent designs to ensure effectiveness.

Narrative. A coherent narrative is the foundation of a design-aware poster. Following a schema widely adopted in scientific writing, we adopt the “And, But, Therefore” (**ABT**) structure (Olson, 2019) to distill the paper’s core message, which establishes context (And), identifies problems (But), and presents solutions (Therefore). This narrative then guides the creation of specific, content-driven section titles.

Layout Structure. Since a poster is a two-dimensional space with width and height, a three-column grid is a common and effective method to ensure a natural reading flow, as shown in Figure 3 (right). This structure strategically places a key visual anchor at the eye-level hot zone (top of the center column) and utilizes white space to separate elements and reduce visual clutter.

Color Design. Color is used to create hierarchy and ensure accessibility. One optimal approach is to employ a restrained, theme-based, and monochromatic palette to maintain visual harmony (Figure 1). This framework establishes a three-tier system: theme color for primary emphasis, monochromatic variants for section backgrounds and a high-contrast accent color for highlights, and all text must adhere to the WCAG 4.5:1 contrast ratio to guarantee readability.

Typography Design. Typography works with color to create clarity from a standard viewing distance. We prioritize legible sans-serif typefaces and establish two types of hierarchy: a visual hierarchy using different font sizes (title, headings, body), and a semantic hierarchy using formatting such as bolding, italics, or a contrast color.

4 PosterGen Agent

In this work, we propose a novel multi-agent framework for generating functionally effective and aesthetic-aware scientific posters, as shown in Figure 2. Our framework implements the design criteria from Section 3 by embedding them as core logic within each specialized agent. This architecture establishes a cascade of structured design constraints throughout the generation process.

The PosterGen workflow consists of four specialist agents or modules: Parser and Curator Agents (Section 4.1), Layout Agent (Section 4.2), Color and Font Agents (Section 4.3), and Renderer (Section 4.4).

4.1 Parser and Curator Agents

Parser Agent. Given a research paper in PDF format, the parser agent initiates the workflow and is responsible for extracting the raw text and all available visual assets (e.g., figures and tables). To accomplish this, we utilize an external PDF converter tool, Marker (Paruchuri, 2025), which converts the paper’s content into Markdown format and saves all identified visual assets as PNG images.

To minimize token usage for downstream agents (particularly for lengthy papers), the parser agent concurrently performs several processing functions. (a) It distills the paper’s core narrative into the ABT structure (see Section 3), to establish a guiding framework for all subsequent content organization; (b) it restructures the raw text into logical sections that focus on main content and essential details, under rigid limitation of maximum 1000 words per section; and (c) it classifies the extracted visual assets into distinct categories based on their narrative role: a single “key_visual” representing the core research; visuals for “problem_illustration” and “method_workflow”; figures depicting “main_results” and “comparative_results”; and all other visual elements as “supporting” material.

Curator Agent. The Curator Agent functions not as a simple content organizer but as a spatial narrative designer. Its primary design consideration is to orchestrate all parsed content elements tightly around the ABT narrative. This narrative-centric approach ensures that the poster’s structure is fluid and engaging, rather than rigid or monotonous. By establishing a strong narrative foundation early, this agent also minimizes the need for unnecessary content and visual refinements in later stages.

Operating on the ABT structure and structured sections provided by the Parser, the Curator Agent performs the initial strategic placement of content. It maps the narrative onto a preliminary three-column storyboard. To follow the narrative and visual strategy, the agent enforces a strict limit of five to eight sections for the entire poster. This constraint guarantees that the three-column layout is fully utilized while preventing content overflow. This structured placement process imitates a typical human design pattern, which progresses logically from introduction and methods to results and discussion.

4.2 Layout Agent

The Layout Agent is designed to implement the storyboard provided by the Curator Agent spatially. It transforms the conceptual plan into a metadata structure, including precise coordinates and sizes for each element. Operating under the constraint of a three-column layout with fixed column width, it systematically places every element onto the canvas, as shown in Figure 3. A vital tool for this process is the precise calculation of each element’s height, which is used to maximize the use of vertical space in each column, and effectively

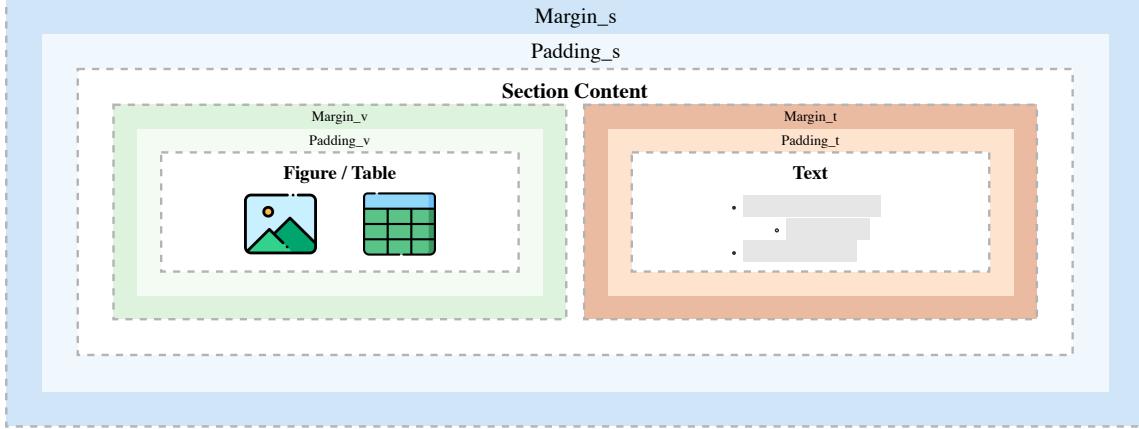


Figure 4 An illustration of the CSS-like box model used to control the spacing between poster elements.

Algorithm 1: Optimized TextFrame Height Estimation

Input: T (text), w (width), f (font attributes), ε (precision)

Data: Initial bounds for binary search

Output: h^* (estimated height)

```

1  $h_{\min}, h_{\max} \leftarrow$  initial bounds;
2 while  $h_{\max} - h_{\min} > \varepsilon$  do
3    $h_{\text{test}} \leftarrow (h_{\min} + h_{\max})/2;$ 
4    $B \leftarrow \text{SimulateTextbox}(T, w, h_{\text{test}}, f);$ 
5   if IsOverflowing( $B$ ) then
6      $| h_{\min} \leftarrow h_{\text{test}};$ 
7   else
8      $| h_{\max} \leftarrow h_{\text{test}};$ 
9   end if
10   $\text{Delete}(B);$ 
11 end while
12  $h^* \leftarrow h_{\max} + \text{NewlineOffset}(T, f.\text{size});$ 
13 return  $h^*;$ 
```

prevent both element overflow and space underutilization. To ensure a visually balanced and aesthetically pleasing composition, the layout agent also employs a CSS-like box model to handle the white space between elements.

While calculating the height for visual assets is straightforward due to their fixed aspect ratios, determining the height for textFrames is much more complex for PPTX. This challenge derives from a discrepancy between the `python-pptx` library, which acts as an XML editor, and the final rendering engine (e.g., Microsoft PowerPoint) that determines the actual appearance. To bridge this gap, we propose an estimation algorithm, which is detailed in Algorithm 1, to estimate the final rendered height accurately. The algorithm first employs a binary search to identify the minimum text box height that avoids any font size reduction by the rendering engine. It then applies a corrective offset, calculated from the number of newline characters, to compensate for subtle deviations in the engine's behavior.

Also, we treat white space as a critical design element that provides natural separation and serves as a visual pause for the reader. To fully control this design pattern, we implement a Python version of the CSS-like box model, as illustrated in Figure 4. In this way, every element, whether text, a figure, or a table, is encapsulated within a box model. The class allows for distinct ‘margin’ and ‘padding’ settings to enable fine-grained control over the spacing surrounding each element. This approach significantly narrows the layout capability gap that typically exists between automated HTML-based and PPTX-based layout generation methods.

4.3 Styling Agents

Once the spatial layout is determined, the styling agents apply the visual and typographic details to generate styled layouts. This stage consists of two specialized components: a color agent and a font agent. Rather than the simple assignment of colors and fonts, we highlight the importance of a design thinking process rooted in the principles of poster aesthetics. This perspective is based on a core understanding that in academic posters, color and typography are not merely decorative; instead, they serve as essential media for both visual and semantic hierarchy.

Color Agent. The color agent focuses on creating a suitable and harmonious color palette, as shown in Figure 1. The agent first searches for the author’s affiliation logo. If it exists, a VLM is adopted to analyze the image and extract a dominant theme color. This method leverages the institution’s official branding to ensure an official appearance. For a fallback plan, the agent can also analyze the key figure from the paper to identify a suitable theme color. After selecting the primary theme color, the next step for the color agent is to generate a complete color scheme strictly following color theory principles. For instance, given the theme color, the color agent will create the following color scheme:

- monochromatic shades for backgrounds and accents, e.g., monochromatic light and dark;
- a high-contrast color that is used specifically for highlighting important keywords.

In this way, the color agent generates a limited color palette that ensures aesthetic cohesion and high readability.

Font Agent. The Font Agent manages typography and works to establish a clear visual hierarchy and emphasize key information within the text. It operates in a two-stage process: the agent first employs one LLM call to analyze the summarized text of the paper, which extracts a list of important keywords for each section. Next, the agent applies styling by using a set of predefined interfaces to assign different font families and sizes. The font agent also highlights the keywords identified in the previous stage via the contrast color from the Color Agent. To avoid a tedious appearance, we adopt several different highlighting styles, i.e., **bolding** and *italics*, to make the poster more visually engaging.

4.4 Renderer

The renderer takes charge of producing the output poster files. It takes the fully styled layout metadata from the previous agents and renders a standard PPTX file via the `python-pptx` library. Additionally, it attaches conference and affiliation logos to the top-right corner of the poster. In the final step, the renderer uses LibreOffice (headless mode) to convert this presentation file into a high-resolution png image for visual inspection and refinement.

5 Experiments

5.1 Metrics

We evaluate the generated posters via a comprehensive rubric, which is detailed in Table 1. This rubric is administered by a Vision-Language Model (VLM) to serve as an expert judge. The evaluation is divided into two fundamental domains: **Poster Content** and **Poster Design**.

The Content domain verifies that the poster is an accurate, concise, and coherent narrative of the source paper, free from factual errors or overly dense text. The Design domain evaluates the poster’s visual execution based on the principles outlined in Section 3, which includes the application of foundational principles (Alignment, Proximity, Repetition, Contrast), effective spatial organization that avoids flaws like element overlap, and the establishment of a clear information hierarchy. Furthermore, it assesses the use of a limited and consistent color palette with prominent accents and a disciplined, legible typographic system.

5.2 Baselines

We choose two types of baselines: an end-to-end text-to-image generation method (GPT-4o), and the state-of-the-art multi-agent poster generation method.

Table 1 VLM-as-Judge evaluation criteria for poster content and design, all on a 1-5 scale.

Domain	Focus Area	Dimension
Poster Content	Content Components	Fidelity, Conciseness, Richness, Synergy
	Narrative Strategy	Narrative, Balance
Poster Design	Design Principles	Principles
	Spatial Organization	Harmony, Alignment, Flow
	Info Hierarchy	Hierarchy
	Color Design	Theme, Readability, Accent
	Typography	Font, Consistency

GPT-4o Image Generation is directly based on the ChatGPT web interface. We provide the GPT-4o model with the source PDF file, along with a text prompt that instructs it to generate an academic poster of a given size. This method produces the final poster as a single image in an end-to-end way, without explicit intermediate generation stages.

PosterAgent (Pang et al., 2025) proposes a top-down, multi-agent pipeline that consists of a (1) Parser to distill the source paper into a structured asset library; and a (2) Planner agent that arranges assets into a binary-tree layout, which is subsequently refined by a (3) Painter-Commenter loop that leverages VLM feedback to correct layout issues. Although this baseline provides a solid technical solution for poster generation, it does not sufficiently incorporate aesthetic and design principles into its agent workflow, which marks a key difference from our approach.

5.3 Quantitive Results

Table 2 and Table 3 present the VLM-as-Judge evaluation results, which compare our PosterGen with baseline methods. To ensure a comprehensive assessment, we employed two distinct VLMs, GPT-4o (Achiam et al., 2023) and Claude Sonnet 4, as evaluators.

The results indicate that the text-to-image GPT-4o method consistently underperforms compared to the multi-agent approaches on this paper-to-poster task. Across nearly all content and design metrics from both VLM judges, its scores are significantly lower, which demonstrates the limitations of direct image generation for this complex, content-sensitive task.

In terms of content metrics, Table 2 shows that PosterGen achieves a performance comparable to PosterAgent, with nearly identical average scores. When evaluated by GPT-4o, the average scores differ by only **0.02**. While PosterAgent scores higher in ‘Summarization & Conciseness’ by a margin of **0.4**, PosterGen shows slight advantages in ‘Visual Richness’ and ‘Narrative’. Under the Claude Sonnet 4 evaluation, both methods achieve an identical average score of **3.70**. PosterAgent leads in ‘Summarization & Conciseness’ (+**1.4** difference), while PosterGen performs better in ‘Narrative’ (+**0.6** difference). The results of content evaluations confirm their similar capabilities in content processing.

On the other hand, the results for design metrics in Table 3 demonstrate a clear and consistent improvement by PosterGen. Under the GPT-4o evaluation, PosterGen’s average score is higher by **0.18** points (**4.44** vs. **4.26**), with significant gains in key areas such as ‘Theme Coherence’ (+**0.5**), ‘Style Consistency’ (+**0.4**), and achieves a peak score of **4.90** in the ‘Font Legibility’ metric. The Claude Sonnet 4 judge reinforces this finding, with PosterGen surpassing PosterAgent by **0.17** in the average score. The most notable improvements can be noticed in ‘Theme Coherence’, where PosterGen scores **0.8** points higher, and in ‘Readability’, with a slight advantage of **0.4**. These results confirm that PosterGen’s design-centric agents improve upon the state-of-the-art method across multiple dimensions of design and aesthetics.

In summary, the quantitative results demonstrate that PosterGen effectively enhances the design and aesthetic quality of posters compared to the state-of-the-art method, without sacrificing the core content fidelity. The performance of the GPT-4o text-to-image method suggests that a collaborative, multi-agent workflow is a more

Table 2 VLM-as-Judge results on **content metrics** across different poster generation methods, with scores averaged over 10 posters rated on a 1-5 scale.

VLM Model	Method	Content Metrics						
		Fid. \uparrow	Conc. \uparrow	Rich. \uparrow	Syn. \uparrow	Nar. \uparrow	Bal. \uparrow	Avg. \uparrow
GPT-4o	GPT-4o-Image	2.20	2.90	3.30	3.00	2.90	2.70	2.83
	PosterAgent	4.90	4.50	4.70	3.70	4.50	3.80	4.35
	PosterGen	4.80	4.10	4.80	3.70	4.60	4.00	4.33
Claude Sonnet 4	GPT-4o-Image	2.80	2.00	2.60	2.30	2.90	2.00	2.43
	PosterAgent	4.10	4.80	3.60	3.20	3.90	2.60	3.70
	PosterGen	4.00	3.40	4.10	3.00	4.50	3.20	3.70

Table 3 VLM-as-Judge results on **design metrics** across different poster generation methods, with scores averaged over 10 posters rated on a 1-5 scale. The best scores for each design metric are **bolded**.

VLM Model	Method	Design Metrics										
		Prin. \uparrow	Harm. \uparrow	Align. \uparrow	Flow \uparrow	Hier. \uparrow	Theme \uparrow	Read. \uparrow	Accent \uparrow	Font \uparrow	Cons. \uparrow	
GPT-4o	GPT-4o-Image	3.40	3.50	3.00	3.60	3.90	3.60	3.80	3.50	2.90	2.70	3.39
	PosterAgent	4.30	4.40	3.60	4.60	4.10	4.00	4.20	4.50	4.50	4.40	4.26
	PosterGen	4.30	4.40	3.70	4.70	4.20	4.50	4.10	4.80	4.90	4.80	4.44
Claude Sonnet 4	GPT-4o-Image	2.60	2.50	2.40	2.70	3.90	2.70	4.40	2.40	3.50	3.00	3.01
	PosterAgent	4.00	3.50	3.70	3.90	4.00	3.10	4.40	2.80	4.10	4.00	3.75
	PosterGen	4.00	3.90	3.10	4.00	4.10	3.90	4.80	3.40	4.10	3.90	3.92

effective strategy for the complex task of generating academic posters that are both visually appealing and grounded in design principles. Among the ten cases, we will discuss two representative examples in the next section. Please refer to the other cases in Appendix D.

5.4 Qualitative Results

Besides VLM-based evaluation, we conducted a qualitative comparison on two representative papers ((Zhang et al., 2025d) and (Sarkar et al., 2025)), as shown in Figure 5. A close examination of the baseline methods reveals significant limitations. As observed in Figure 5(a), while the generated poster presents a clear two-column layout at a macro level, its content suffers from critical flaws. These include sections with gibberish text, duplicated content blocks, and the hallucination of visual assets not present in the source paper. This indicates that text-to-image generation methods struggle with content consistency for this complex task.

In comparison, PosterAgent Pang et al. (2025) achieves a significant improvement in content fidelity through its multi-agent workflow, but still limited in design and aesthetic principles. Its layout suffers from element overlap and text overflow, and it fails to establish a logical reading flow. Furthermore, its stylistic choices are monotonous, using identically sized, black bullet points that do not create a visual hierarchy or emphasize key information.

In contrast, the posters generated by our method, **PosterGen** (c), exhibit superior design quality. The title bar is more aesthetically refined, featuring differentiated fonts for the title, authors and sections to create a appealing hierarchy, and incorporates conference and affiliation logos in the top-right corner. Instead of relying on rigid borders, PosterGen establishes an intuitive reading order through colored section blocks, strong left alignment, and the deliberate use of white space, which constitutes over 25% of the canvas to reduce cognitive load. The textual content is enriched through a two-level bullet point structure, where several key phrases are strategically highlighted using contrasting colors, bolding, and italics to guide the reader’s focus. In addition, PosterGen draws attention to a main anchor section by applying a light monochromatic background color, which provides emphasis without being visually distracting.

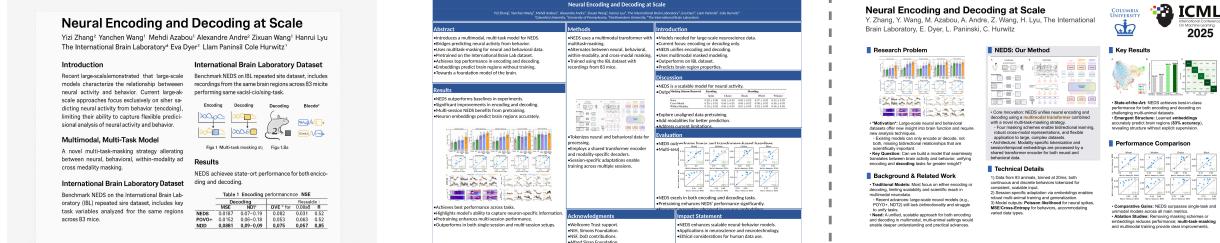


Figure 5 Qualitative comparison on two representative papers. (a) Posters generated by GPT-4o-Image; (b) Posters generated by PosterAgent from Pang et al. (2025); (c) Posters generated by PosterGen (ours).

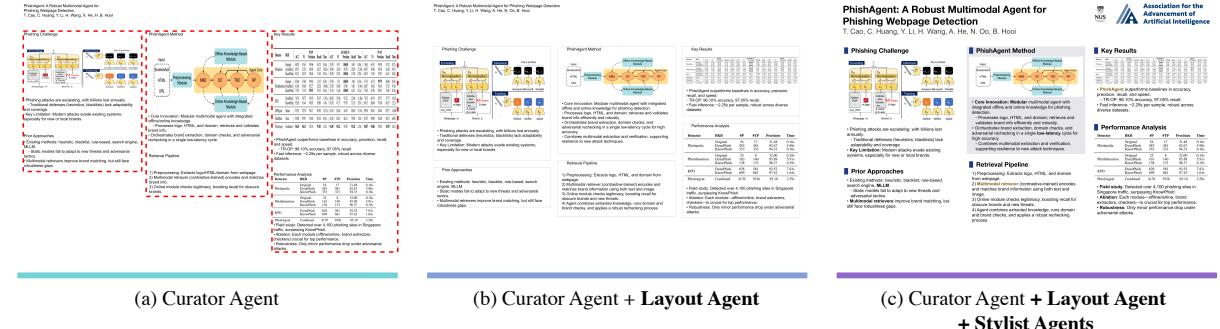


Figure 6 Ablation study on Cao et al. (2025b). (a) Output of Curator Agent. Chaotic layouts are highlighted in the red dashed box. (b) Output of Layout Agent. Layout Agent applies spatial adjustments and balance columns. (c) Output of the entire multi-agent pipeline. Stylist Agents apply visually appealing colors and font elements to the poster.

5.5 Ablation Study

We conduct an ablation study to isolate the contributions of the three core agents in our PosterGen pipeline: the Curator Agent, the Layout Agent, and the Stylist Agents. Experimenting on the same paper (Cao et al., 2025b), we sequentially intercept the output poster of each agent during the generation process. The results of this ablation analysis are visualized in Figure 6.

Figure 6(a) shows the output after the Curator Agent stage. This agent is responsible for generating the initial storyboard, which contains all the necessary textual and visual assets extracted from the source paper. However, without any spatial adjustments within the three-column format, the resulting poster exhibits severe layout issues (as highlighted in red dashed boxes). These include improper spacing, imbalanced columns, and significant content overflow and underutilization.

The output of the Layout Agent, shown in Figure 6(b), addresses these layout flaws. This agent applies the box model to make appropriate spatial adjustments to all text and visual elements. Then, it executes a balancing loop to resolve column imbalances. During this stage, Layout Agent handles layout issues, including underutilization and overflows.

The Stylist Agents apply the final aesthetic layer at the final step. The Color Agent introduces a theme-based color to section titles, applies a light monochromatic background to the most important section to draw attention, and uses a high-contrast color to highlight key terms. The Font Agent adjusts font families and sizes

to enrich the overall visual hierarchy, and further emphasizes keywords through the strategic use of bold, italic, and the contrast color.

6 Conclusion

In this work, we present **PosterGen**, a novel aesthetic-aware multi-agent framework that addresses the challenge of automated academic poster generation. Our approach is uniquely guided by core poster design and aesthetic principles, which embeds them into a specialized agent workflow that mirrors the process of professional designers. To systematically assess visual quality, we also introduce a VLM-based rubric that measures layout design, readability, and aesthetic coherence. Experimental and Qualitative results demonstrate that **PosterGen** improves generation quality, and produce posters that are visually compelling and require minimal manual refinement. By automating the most challenging design aspects of the poster generation, our method makes the poster a more accessible medium for the one-on-one scholarly dialogue that is crucial for scientific communication.

References

- Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- Seth Aycock, David Stap, Di Wu, Christof Monz, and Khalil Sima'an. Can llms really learn to translate a low-resource language from one grammar book? *arXiv preprint arXiv:2409.19151*, 2024.
- Xiaohe Bo, Zeyu Zhang, Quanyu Dai, Xueyang Feng, Lei Wang, Rui Li, Xu Chen, and Ji-Rong Wen. Reflective multi-agent collaboration based on large language models. *Advances in Neural Information Processing Systems*, 37: 138595–138631, 2024.
- Juntai Cao, Xiang Zhang, Raymond Li, Chuyuan Li, Chenyu You, Shafiq Joty, and Giuseppe Carenini. Multi2: Multi-agent test-time scalable framework for multi-document processing. *arXiv preprint arXiv:2502.20592*, 2025a.
- Tri Cao, Chengyu Huang, Yuexin Li, Wang Huilin, Amy He, Nay Oo, and Bryan Hooi. Phishagent: a robust multimodal agent for phishing webpage detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025b.
- Haoyu Chen, Xiaojie Xu, Wenbo Li, Jingjing Ren, Tian Ye, Songhua Liu, Ying-Cong Chen, Lei Zhu, and Xinchao Wang. Posta: A go-to framework for customized artistic poster generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 28694–28704, 2025a.
- SiXiang Chen, Jianyu Lai, Jialin Gao, Tian Ye, Haoyu Chen, Hengyu Shi, Shitong Shao, Yunlong Lin, Song Fei, Zhaohu Xing, et al. Postercraft: Rethinking high-quality aesthetic poster generation in a unified framework. *arXiv preprint arXiv:2506.10741*, 2025b.
- Zen Faulkes. *Better posters: plan, design and present an academic poster*. Pelagic Publishing Ltd, 2021.
- Tsu-Jui Fu, William Yang Wang, Daniel McDuff, and Yale Song. Doc2ppt: Automatic presentation slides generation from scientific documents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.
- Yifan Gao, Zihang Lin, Chuanbin Liu, Min Zhou, Tiezheng Ge, Bo Zheng, and Hongtao Xie. Postermaker: Towards high-quality product poster generation with accurate text rendering. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 8083–8093, 2025.
- Mingyang Gong, Jing Fan, Guohui Lin, Bing Su, Zihan Su, and Xiang Zhang. Multiprocessor scheduling with testing: improved online algorithms and numerical experiments. *Journal of Scheduling*, pages 1–15, 2025.
- Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*, 2024.
- Yue Hu and Xiaojun Wan. Ppsgen: Learning-based presentation slides generation for academic papers. *IEEE transactions on knowledge and data engineering*, 27(4):1085–1097, 2014.

- Zhi Jin, Sheng Xu, Xiang Zhang, Tianze Ling, Nanqing Dong, Wanli Ouyang, Zhiqiang Gao, Cheng Chang, and Siqi Sun. Contranovo: a contrastive learning approach to enhance de novo peptide sequencing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 144–152, 2024.
- Kyudan Jung, Hojun Cho, Jooyeol Yun, Soyoung Yang, Jaehyeok Jang, and Jaegul Choo. Talk to your slides: Language-driven agents for efficient slide editing. *arXiv preprint arXiv:2505.11604*, 2025.
- Keshav Kumar and Ravindranath Chowdary. Slidespawn: An automatic slides generation system for research publications. *arXiv preprint arXiv:2411.17719*, 2024.
- Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbulin, and Bernard Ghanem. Camel: Communicative agents for "mind" exploration of large language model society. *Advances in Neural Information Processing Systems*, 36: 51991–52008, 2023.
- Aiwei Liu, Sheng Guan, Yiming Liu, Leyi Pan, Yifei Zhang, Liancheng Fang, Lijie Wen, Philip S Yu, and Xuming Hu. Can watermarked llms be identified by users via crafted prompts? *arXiv preprint arXiv:2410.03168*, 2024.
- Puyuan Liu, Xiang Zhang, and Lili Mou. A character-level length-control algorithm for non-autoregressive sentence summarization. *Advances in Neural Information Processing Systems*, 35:29101–29112, 2022.
- Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, et al. Agentbench: Evaluating llms as agents. *arXiv preprint arXiv:2308.03688*, 2023.
- Zhu Liu, Zijun Wang, Jinyuan Liu, Fanqi Meng, Long Ma, and Risheng Liu. Deal: Data-efficient adversarial learning for high-quality infrared imaging. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 28198–28207, 2025a.
- Zirui Liu, Jiatong Li, Yan Zhuang, Qi Liu, Shuanghong Shen, Jie Ouyang, Mingyue Cheng, and Shijin Wang. am-elo: A stable framework for arena-based llm evaluation. *arXiv preprint arXiv:2505.03475*, 2025b.
- Ishani Mondal, S Shwetha, Anandhavelu Natarajan, Aparna Garimella, Sambaran Bandyopadhyay, and Jordan Boyd-Graber. Presentations by the humans and for the humans: Harnessing llms for generating persona-aware slides from documents. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2664–2684, 2024.
- Seungeun Oh, Jihong Park, Sihun Baek, Hyelin Nam, Praneeth Vepakomma, Ramesh Raskar, Mehdi Bennis, and Seong-Lyun Kim. Differentially private cutmix for split learning with vision transformer. *arXiv preprint arXiv:2210.15986*, 2022.
- Randy Olson. *Narrative is everything: The ABT framework and narrative evolution*. Prairie Starfish Productions, 2019.
- Wei Pang, Kevin Qinghong Lin, Xiangru Jian, Xi He, and Philip Torr. Paper2poster: Towards multimodal poster automation from scientific papers. *arXiv preprint arXiv:2505.21497*, 2025.
- Vik Paruchuri. Marker: Convert pdf to markdown and json quickly with high accuracy, 2025.
- Yu-Ting Qiang, Yan-Wei Fu, Xiao Yu, Yan-Wen Guo, Zhi-Hua Zhou, and Leonid Sigal. Learning to generate posters of scientific papers by probabilistic graphical models. *Journal of Computer Science and Technology*, 34(1):155–169, 2019.
- Yuting Qiang, Yanwei Fu, Yanwen Guo, Zhi-Hua Zhou, and Leonid Sigal. Learning to generate posters of scientific papers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016.
- Md Ashiqur Rahman and Raymond A Yeh. Truly scale-equivariant deep nets with fourier layers. *Advances in Neural Information Processing Systems*, 36:6092–6104, 2023.
- Anindya Sarkar, Alex DiChristofano, Sanmay Das, Patrick J Fowler, Nathan Jacobs, and Yevgeniy Vorobeychik. Active geospatial search for efficient tenant eviction outreach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.
- Rohit Saxena, Pasquale Minervini, and Frank Keller. Postersum: A multimodal benchmark for scientific poster summarization. *arXiv preprint arXiv:2502.17540*, 2025.
- Jaejung Seol, Seojun Kim, and Jaejun Yoo. Posterllama: Bridging design ability of language model to content-aware layout generation. In *European Conference on Computer Vision*, pages 451–468. Springer, 2024.
- Jingwei Shi, Zeyu Zhang, Biao Wu, Yanjie Liang, Meng Fang, Ling Chen, and Yang Zhao. Presentagent: Multimodal agent for presentation video generation. *arXiv preprint arXiv:2507.04036*, 2025.
- M Sravanthi, C Ravindranath Chowdary, and P Sreenivasa Kumar. Slidesgen: Automatic generation of presentation slides for a technical paper using summarization. In *FLAIRS*, 2009.

- Tao Sun, Enhao Pan, Zhengkai Yang, Kaixin Sui, Jiajun Shi, Xianfu Cheng, Tongliang Li, Wenhao Huang, Ge Zhang, Jian Yang, et al. P2p: Automated paper-to-poster generation and fine-grained benchmark. *arXiv preprint arXiv:2505.17104*, 2025.
- Hao Wang, Shohei Tanaka, and Yoshitaka Ushiku. Scipostlayout: A dataset for layout analysis and layout generation of scientific posters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8136–8141, 2024.
- Jiaqi Wei, Hao Zhou, Xiang Zhang, Di Zhang, Zijie Qiu, Wei Wei, Jinzhe Li, Wanli Ouyang, and Siqi Sun. Alignrag: An adaptable framework for resolving misalignments in retrieval-aware reasoning of rag. *arXiv e-prints*, pages arXiv–2504, 2025.
- Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, et al. Autogen: Enabling next-gen llm applications via multi-agent conversations. In *First Conference on Language Modeling*, 2024.
- Sheng Xu and Xiaojun Wan. Posterbot: A system for generating posters of scientific papers with neural models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.
- Heng Yang and Marco Pavone. Object pose estimation with statistical guarantees: Conformal keypoint detection and geometric uncertainty propagation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8947–8958, 2023.
- Kaichun Yao, Lan Zeng, Chuan Qin, Hengshu Zhu, Yanjun Wu, and Libo Zhang. SciPG: A new benchmark and approach for layout-aware scientific poster generation, 2025. <https://openreview.net/forum?id=zmHqlXGTT1>.
- Yuwei Yin, Jean Kaddour, Xiang Zhang, Yixin Nie, Zhenguang Liu, Lingpeng Kong, and Qi Liu. Ttida: Controllable generative data augmentation via text-to-text and text-to-image models. *arXiv preprint arXiv:2304.08821*, 2023.
- Hui Zhang, Dexiang Hong, Maoke Yang, Yutao Chen, Zhao Zhang, Jie Shao, Xinglong Wu, Zuxuan Wu, and Yu-Gang Jiang. Creatidesign: A unified multi-conditional diffusion transformer for creative graphic design. *arXiv preprint arXiv:2505.19114*, 2025a.
- Xiang Zhang, Muhammad Abdul-Mageed, and Laks VS Lakshmanan. Autoregressive+ chain of thought= recurrent: Recurrence’s role in language models’ computability and a revisit of recurrent transformer. *arXiv preprint arXiv:2409.09239*, 2024a.
- Xiang Zhang, Senyu Li, Ning Shi, Bradley Hauer, Zijun Wu, Grzegorz Kondrak, Muhammad Abdul-Mageed, and Laks VS Lakshmanan. Cross-modal consistency in multimodal large language models. *arXiv preprint arXiv:2411.09273*, 2024b.
- Xiang Zhang, Juntai Cao, Jiaqi Wei, Yiwei Xu, and Chenyu You. Tokenization constraints in llms: A study of symbolic and arithmetic reasoning limits. *arXiv preprint arXiv:2505.14178*, 2025b.
- Xiang Zhang, Juntai Cao, Jiaqi Wei, Chenyu You, and Dujian Ding. Why prompt design matters and works: A complexity analysis of prompt search space in llms. *arXiv preprint arXiv:2503.10084*, 2025c.
- Yizi Zhang, Yanchen Wang, Mehdi Azabou, Alexandre Andre, Zixuan Wang, Hanrui Lyu, The International Brain Laboratory, Eva Dyer, Liam Paninski, and Cole Hurwitz. Neural encoding and decoding at scale. *arXiv preprint arXiv:2504.08201*, 2025d.
- Hao Zheng, Xinyan Guan, Hao Kong, Jia Zheng, Weixiang Zhou, Hongyu Lin, Yaojie Lu, Ben He, Xianpei Han, and Le Sun. Pptagent: Generating and evaluating presentations beyond text-to-slides. *arXiv preprint arXiv:2501.03936*, 2025.

Appendix

Table of Contents

– Extended Related Work	13
– Implementation Details	13
• Experimental Environment	13
• Configuration Parameters	13
– Prompts	14
• Baseline Prompt	14
• PosterGen Prompts	15
• VLM-as-Judge Evaluation Prompt Template	17
– Additional Qualitative Results	18
– Limitations	21
– Broader Impacts	21

A Extended Related Work

Slide Generation. A similar task to poster generation is the automatic generation of presentation slides from documents (Mondal et al., 2024; Zheng et al., 2025; Jung et al., 2025; Fu et al., 2022; Hu and Wan, 2014; Zhang et al., 2025b; Kumar and Chowdary, 2024; Sravanthi et al., 2009; Shi et al., 2025). Some works develop agents for general purposes, such as efficient slide editing (Jung et al., 2025) or narrated presentation videos (Shi et al., 2025). Other methods, such as PPTAgent (Zheng et al., 2025), focus on holistically improving the content, design, and coherence of the slides. Among these, several works specifically aim to generate slides for academic presentations. Early approaches like PPSGen (Hu and Wan, 2014), SlidesGen (Sravanthi et al., 2009), and SlideSpawn (Kumar and Chowdary, 2024; Gong et al., 2025) utilized summarization and information extraction techniques to generate draft slides. More recent approaches utilize end-to-end systems (Fu et al., 2022) or persona-aware models (Mondal et al., 2024) to generate more tailored slides. However, scientific poster design is far more challenging than slide generation, as slides can distribute content across multiple pages and work together with a presenter’s oral explanation to convey the full message, while a poster need to contain all necessary information from a paper onto a single page, and be visually appealing to attract the attention and initiate a dialogue (Faulkes, 2021) with the authors.

B Implementation Details

B.1 Experimental Environment

The framework is compatible with both Linux and macOS operating systems and is implemented in Python 3.11. While a GPU is not strictly required, it is strongly recommended to accelerate document parsing and optical character recognition (OCR) tasks handled by the `marker-pdf` tool. The framework relies on several key libraries; the most critical dependencies for reproducing our work include:

- `python-pptx >= 1.0.2` for PPT generation.
- `langchain >= 0.3.0` and `langgraph >= 0.2.45` for building the multi-agent workflow.
- `marker-pdf == 1.7.5` for PDF-to-Markdown conversion and parsing.
- `Pillow >= 10.0.0` for image manipulation.

B.2 Configuration Parameters

To ensure the reproducibility of our implementation, we list the most critical parameters that directly influence the behavior of the agents and the quality of generated posters in Table 4. Aesthetic parameters related to

Category	Parameter	Value
LLM/VLM Configuration	Model	gpt-4.1-2025-04-14
	↳ Alternatives	gpt-4o-2024-08-06
		gpt-4.1-mini-2025-04-14
		claude-sonnet-4-20250514
Content Constraints	Temperature	0.7
	Num. of Sections	[5, 8]
	Num. of Visual Assets	[4, 6]
	Max Words per Section	1000
Text Height Estimation	Height Precision (ϵ)	0.001 inches
	Newline Offset Ratio	1.0

Table 4 Key hyperparameters for PosterGen Implementations. The default LLM/VLM model is gpt-4.1-2025-04-14, with others as alternatives. The “Content Constraints” are controlled via instructions within the agent prompts.

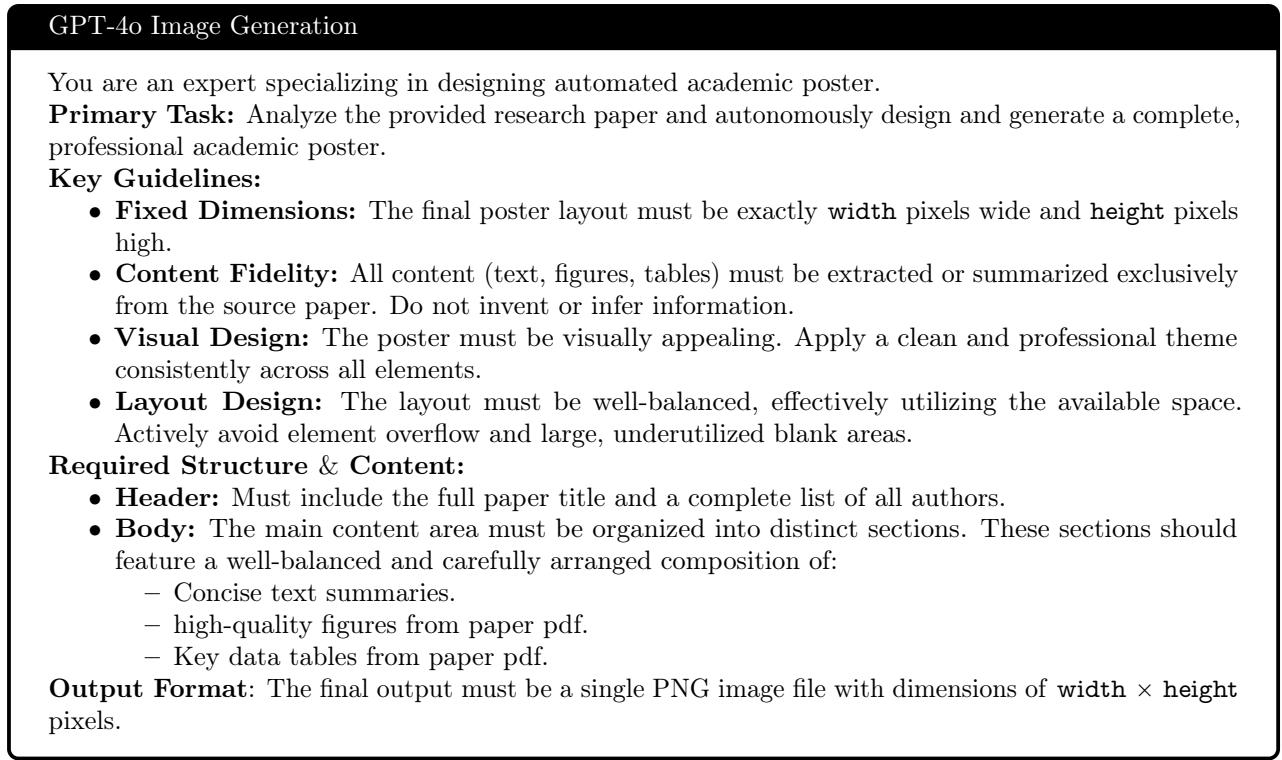


Figure 7 Prompt for GPT-4o Image Generation.

layout, color and typography (font sizes, margins, color values) are defined in the configuration files within our source code but omitted here for clarity.

C Prompts

C.1 Baseline Prompt

We present the prompt of GPT-4o Image Generation via ChatGPT web interface (as illustrated in Figure 7), which is alongside the input paper file.

Parser Agent (1): Title and Authors Extraction

You are an expert academic paper parser. Your task is to extract the title and authors from the provided academic paper text.

Please extract:

1. **Title:** The main title of the paper
2. **Authors:** All author names using initials (no affiliations, emails, or other metadata)

Strict Formatting Requirements:

- **Title:** Use proper title case where each word has only the first letter capitalized, EXCEPT for established acronyms, technical terms, or proper nouns that are conventionally written in all uppercase letters (such as abbreviations for organizations, technologies, or methodologies). Keep such terms in their original case. Example: "A Study of Machine Learning Methods" but preserve acronyms like "Using LLM for Data Analysis" or "CNN Architecture Design".
- **Authors:** Use initials for authors' names. Convert full names to initials format, preserving middle initials when present. Examples: "Kevin W. Jones" → "K.W. Jones", "Yann LeCun" → "Y. LeCun", "Mary Smith Johnson" → "M.S. Johnson". Separate multiple authors with ", ". Remove all affiliations, emails, institutions, departments, addresses, and other metadata.

Input text: {{ markdown document }}

Required JSON structure:

```
1 {
2   "title": "Title With Proper Case Formatting",
3   "authors": "F. Author, S. Author, T. Author"
4 }
```

Figure 8 Prompt for the Parser Agent to extract the title and authors from the source paper.

Parser Agent (2): Narrative (ABT) Extraction

Extract ABT narrative structure optimized for poster presentation:

Input: {{ Academic paper markdown text }}

Output: JSON with poster-ready ABT structure

Guidelines:

- Each section (and/but/therefore) should be 1-2 concise sentences
- Focus on visual impact and poster audience understanding
- Emphasize key contributions and results
- Avoid technical jargon where possible

Required JSON structure:

```
1 {
2   "and": "Current knowledge and established facts (background context)",
3   "but": "Specific problem, gap, or challenge identified",
4   "therefore": "Your solution, contribution, and key findings",
5   "poster_hook": "One compelling sentence that grabs attention",
6   "key_impact": "Why this research matters (practical implications)"
7 }
```

Paper content: markdowndocument

Figure 9 Prompt for the Parser Agent to extract ABT-structured narratives.

C.2 PosterGen Prompts

We present the detailed prompts design (Zhang et al., 2025c) used in our PosterGen multi-agent workflow as follows.

Parser Agent. This includes: (1) Title and Authors Extraction (Figure 8); (2) Narrative (ABT) Extraction (Figure 9); (3) Visual Asset Classification (Figure 10); and (4) Structured Section Extraction (Figure 11).

Parser Agent (3): Classify Visual Assets

Classify visual assets by column-aware poster placement and research role:

Available visuals with captions: visuals list

Classify each visual into exactly one category based on column-specific poster design:

Column-Aware Categories:

1. **key visual:** Most important method visual representing core research innovation (max 1, middle column)
2. **problem illustration:** Visuals showing research problem, challenges, or motivation (left column introduction)
3. **method workflow:** Method architecture, system diagrams, algorithmic workflows (middle column method)
4. **main results:** Primary experimental results, performance tables, key findings (right column)
5. **comparative results:** Baseline comparisons, ablation studies, validation charts (right column)
6. **supporting:** Background concepts, supplementary analysis, minor details (flexible placement)

Classification Guidelines:

- Problem Context: Figures showing “what’s wrong” or “why this matters” → problem illustration
- Method Core: Most important technical diagram → key visual
- Method Details: Architecture/workflow diagrams → method workflow
- Primary Evidence: Main performance results → main results
- Validation Evidence: Comparisons with baselines → comparative results
- Background/Supplementary: Minor or supporting content → supporting

Consider:

- Visual content and research narrative role
- Optimal column placement for logical flow
- Visual impact and audience comprehension

Required JSON output:

```
1 {
2   "key_visual": "visual_id or null",
3   "problem_illustration": ["visual_id1", ...],
4   "method_workflow": ["visual_id1", ...],
5   "main_results": ["visual_id1", ...],
6   "comparative_results": ["visual_id1", ...],
7   "supporting": ["visual_id1", ...]
8 }
```

Ensure every visual id appears exactly once across all categories.

Figure 10 Prompt for the Parser Agent to classify extracted visual assets.

Curator Agent. The Curator Agent generates an effective storyboard through strategic content planning and by applying visual height constraints. Due to space limitations, we split the prompt into three parts: (1) Input and Design Patterns (Figure 12), (2) visual asset selection and content organization (Figure 13), and (3) output format (Figure 14). We also omit less important parts and replace them with ellipsis mark (...). The full prompt is available in the source code.

Layout Balancer. This is a sub-agent of Layout Agent designed to improve column utilization and prevent overflows. Its prompt is detailed in Figure 15 and Figure 16.

Color Agent. We present only the prompt for extracting the theme color from an affiliation logo using a VLM (as shown in Figure 17); the fallback method, which uses a key visual asset, is omitted for clarity.

Font Agent. The font agent calls LLM once to extract and classify different keywords, which is illustrated in Figure 18.

Parser Agent (4): Structured Section Extraction

Extract structured sections from academic paper text for poster creation.

Paper Text: {{ raw text }}

Task: Extract all major sections from the paper and organize them with their content. Focus on sections that would be relevant for an academic poster.

Section Extraction Guidelines:

1. **Identify Major Sections:**

- Introduction/Background
- Related Work (if substantial)
- Methodology/Approach
- Experiments/Results
- Discussion/Analysis

2. **Content Processing:**

- Extract the main content for each section
- Keep section content under **1000** words
- Preserve key technical details, formulas, and findings
- Maintain important bullet points and lists
- Remove excessive citations and references

3. **Section Classification:**

- **foundation:** Introduction, background, motivation, problem statement
- **method:** Methodology, approach, algorithm, system design
- **evaluation:** Experiments, results, analysis, validation

Required JSON structure:

```
1  {
2    "paper_sections": [
3      {
4        "section_name": "Introduction",
5        "section_type": "foundation",
6        "content": "Main content of the section (max 1000 words)",
7        "key_points": [ ... ],
8        "importance": "high | medium | low",
9        "contains_figures": ["figure_1", "figure_2"],
10       "contains_tables": ["table_1"]
11     }
12   ],
13   "paper_structure": {
14     "total_sections": 5,
15     "foundation_sections": 2,
16     "method_sections": 2,
17     "evaluation_sections": 1
18   }
19 }
```

Critical Requirements:

- Extract ALL major sections (don't skip any)
- Keep each section under 1000 words
- Preserve technical accuracy
- Identify which figures/tables belong to each section
- Classify section importance for poster layout

Generate structured sections that provide comprehensive paper coverage for poster creation.

Figure 11 Prompt for the Parser Agent to extract structured sections.

C.3 VLM-as-Judge Evaluation Prompt Template

We present the prompt template used for our VLM-as-Judge evaluation, as shown in Figure 19. This standardized template is applied to every evaluation focus area and uses a 5-point scale. To counteract the tendency of Vision-Language Models (VLMs) to provide overly generous scores, our prompt design incorporates a targeted

Curator Agent: Spatial Story Board Generation (Part 1 of 3)

You are an Expert Academic Poster Designer specializing in visual-dense poster layouts with strategic spatial organization.

Mission: Transform research papers into spatially-organized poster sections that maximize visual asset utilization while following human design patterns. Prioritize visual impact over text density.

Input:

- Paper Structured Sections: $\{\{ \text{structured sections} \}\}$
- Enhanced ABT Narrative: $\{\{ \text{narrative content} \}\}$
- Classified Visuals: $\{\{ \text{classified visuals} \}\}$
- Available Images: $\{\{ \text{available images} \}\}$
- Available Tables: $\{\{ \text{available tables} \}\}$
- Visual Heights Information: $\{\{ \text{visual heights info} \}\}$
- Available Height Per Column: $\{\{ \text{available height per column} \}\}$

Human Poster Design Patterns:

Based on analysis of successful academic posters:

1. Left Column Strategy - Foundation & Context:

- Introduction/Background/Motivation (priority placement)
- Problem definition and challenges
- Related work and background context
- Method overview or workflow diagrams

2. Middle Column Strategy - Core Technical Content:

- Primary methodology (highest priority content)
- Technical details and algorithms
- Theoretical analysis and key innovations
- System architecture diagrams

3. Right Column Strategy - Experiments & Results:

- Experimental results (tables and performance charts)
- Key findings and validation data
- Performance comparisons and analysis

Figure 12 Part 1 of the Curator Agent prompt, focusing on the mission, inputs, and high-level human design patterns.

few-shot example strategy. For high scores (4 and 5), we provide positive examples of desired qualities, while for low-to-mid scores (1, 2, and 3), we provide negative examples of common flaws. This approach is designed to calibrate the VLM’s judgment and yield more accurate, evidence-based scoring.

D Additional Qualitative Results

In this section, we present the qualitative results for the remaining eight papers. These papers were selected from top-tier AI conferences from the last three years, including ICLR, ICML, AAAI, CVPR, and NeurIPS.

The results further highlight the limitations of end-to-end GPT-4o Image Generation. Beyond content fidelity issues like gibberish text, this method struggles with poster boundary problems. For example, in Figures 22, 23, 24, 25, and 27, the generated posters exhibit substantial horizontal blank space, which indicates poor canvas utilization. Moreover, several examples (Figures 21, 23, 25, and 27) suffer from vertical truncation, where content is abruptly cut off. These fundamental layout control issues confirm that text-to-image models are not yet capable of handling the complex constraints of the paper-to-poster task.

Compared to direct image generation, PosterAgent (Pang et al., 2025) shows significant improvements in content fidelity and basic layout structure. However, it demonstrates several limitations in adhering to design principles. (1) Sections are often poorly aligned, as seen in Figures 20 and 22, which disrupts the reader’s logical flow. (2) Several posters contain large areas of wasted white space within their central sections, as shown in (Figures 20, 25, and 27). (3) Its color scheme is monotonous, which consistently uses the same blue background with white

Oversized Visual Exclusion:

- **Exclusion Rule:** Any visual with height percentage > 50% in visual heights info MUST BE EXCLUDED from poster
- **Reasoning:** Even with 80% shrinking, these visuals would still exceed 40% column height
- **Smart Substitution:** ...
- **FALLBACK RULE:** If only ONE oversized visual (> 50%) is selected, allow it to proceed. For multiple oversized visuals, only select the one with SMALLEST height percentage.

Visual Asset Strategic Selection Process

1. Key Visual Mandatory Placement:

- Identify the “key visual” from classified visuals. This is the MOST important visual
- Place key visual in middle column, top priority section
- This anchors the entire poster layout around the core research contribution

2. Column-Based Visual Distribution:

- **Column 1 (Left) - Foundation & Context:**
 - **MINIMUM:** 1 visual asset required
 - **Purpose:** Express core research problem or contradiction visually
 - **Selection Priority:** Choose visuals that illustrate problem context, background concepts, or prior work limitations
 - **Maximum:** 2 visual assets
- **Column 2 (Middle) - Methodology:**
 - **MANDATORY:** Contains key visual from classified visuals
 - **Additional:** May include 1 supporting method diagram
 - **Maximum:** 2 visual assets
- **Column 3 (Right) - Results & Impact:**
 - **STRICT MAXIMUM:** 2 visual assets ONLY
 - **Selection Criteria:** Choose the 2 most critical visuals that directly validate main claims
 - **Priority Order:** ...

3. Visual Distribution Enforcement: ...

4. Column Space Optimization Strategy: ...

Core Task: Create 5-8 poster sections with BOTH content organization AND strategic spatial placement to achieve perfect space utilization across all three columns. DO NOT create any conclusion, takeaway, future work, or impact sections. Focus ONLY on problem, method, and results/experiments.

Content Organization Guidelines:

1. Section Requirements:

- **Section titles:** Maximum 4 words (e.g., “Our Method”, “Key Results”)
- **Text content:** 2-3 concise entries using different rich hierarchical formatting (see examples below) based on section contents
- **Visual integration:** Each visual assigned to exactly ONE section
- **Complete content:** No ellipsis (...), write full bullet points

2. Rich Text Formatting Options:

- **A) Nested Bullet Structure:**

```
1  "* Primary concept or finding",
2   "  - Supporting detail or sub-point",
3     "    - Additional supporting evidence"
```
- Other formats like Bold Headers and Ordered Lists are also available.

Figure 13 Part 2 of the Curator Agent prompt, specifying the detailed rules for visual asset selection, content organization, and other planning requirements.

text for all section titles, thus lacking visual appeal. (4) All body text is presented with the same font styling and size. This leads to a dull visual hierarchy and potential reader fatigue.

In contrast to these baselines, our proposed PosterGen framework demonstrates superior performance by

Curator Agent: Spatial Story Board Generation (Part 3 of 3)

Output Format:

```

1   {
2     "spatial_content_plan": {
3       "poster_strategy": {
4         "narrative_flow": "How the story progresses across columns",
5         "space_utilization_approach": "Strategy for filling all three columns",
6         "column_balance_rationale": "Why content is distributed this way"
7       },
8       "sections": [
9         {
10          "section_id": "unique_identifier",
11          "section_title": "Max 4 Words",
12          "column_assignment": "left|middle|right",
13          "vertical_priority": "top|middle|bottom",
14          "importance_level": 1,
15          "content_type": "foundation|method|results",
16          "expected_content_density": "high|medium|low",
17          "text_content": [
18            "* **Key Innovation:** Core contribution with bold emphasis",
19            " - Supporting technical detail",
20            "* **Impact:** Quantifiable result or benefit"
21          ],
22          "visual_assets": [
23            {
24              "visual_id": "figure_1",
25              "visual_purpose": "How this supports the section",
26              "placement_rationale": "Why this visual belongs in this spatial location"
27            }
28          ],
29          "spatial_rationale": "Why this section belongs in this column/position"
30        }
31      ]
32    },
33    "column_distribution": {
34      "left_column": {
35        "focus": "Foundation and context",
36        "assigned_sections": ["section_id_1", "section_id_2"],
37        "content_strategy": "Build problem understanding and motivation"
38      },
39      "middle_column": {
40        "focus": "Core methodology",
41        "assigned_sections": ["section_id_3", "section_id_4"],
42        "content_strategy": "Present technical innovation and approach"
43      },
44      "right_column": {
45        "focus": "Results and impact",
46        "assigned_sections": ["section_id_5", "section_id_6"],
47        "content_strategy": "Demonstrate effectiveness and validations"
48      }
49    }
50  }

```

Figure 14 Part 3 of the Curator Agent prompt, defining the exact JSON output format and data structure required from the agent.

integrating design and aesthetic principles directly into the agent workflow. As shown across the qualitative examples, our method yields improvements in color harmony, typographic hierarchy, logical reading flow, and overall spatial efficiency, resulting in more professional and presentation-ready posters.

Balancer Agent: Balance Column Space Utilization (Part 1 of 2)

You are an expert academic poster layout optimization specialist. Your goal is to achieve optimal three-column space utilization through conservative within-column content adjustments only.

Current Column Status:

- **Column 1 (Left):** {left utilization} utilization - {left status}
- **Column 2 (Middle):** {middle utilization} utilization - {middle status}
- **Column 3 (Right):** {right utilization} utilization - {right status}
- **Available Height per Column:** {available height} inches

Target Utilization: 85-95% for each column

Core Optimization Principle: Prioritize content reduction over content expansion. Better to have 80% utilization than risk overflow beyond available space.

Column Content Rules:

1. Left Column: Foundation & Context
 - **Purpose:** Introduction, background, prior work, problem setup, supporting context
 - **Content Types:** Motivation, challenges, related work, problem definitions, supporting materials
 - **Reading Role:** Sets up the research problem and provides necessary background
2. Middle Column: Core Methodology
 - **Purpose:** Method details, algorithms, implementation, technical innovation
 - **Content Types:** Core methods, algorithms, technical approach, key innovations
 - **Reading Role:** Presents the technical contribution and methodology
 - **CRITICAL:** Contains key visual (importance level=1). NEVER remove method sections
3. Right Column: Results & Impact
 - **Purpose:** Experiments, evaluation, findings, conclusions, future work
 - **Content Types:** Experimental results, performance analysis, conclusions, future directions
 - **Reading Role:** Demonstrates validation and impact of the proposed method

Figure 15 Part 1 of the Balancer sub-agent prompt, outlining its role, the current column status, and the fundamental content rules for each column.

E Limitations

Although **PosterGen** demonstrates a significant advancement in aesthetic-aware poster generation, we identify several limitations that present opportunities for future work.

First, while the Marker tool used by the Parser Agent is generally effective, it occasionally encounters extraction errors. These can include the omission of title or author text, or the generation of slightly flawed visual assets from the source PDF. Future improvements would benefit from a more robust PDF conversion tool specialized for the complex structure of academic documents.

Second, the Curator Agent is currently restricted to using only the visual assets that exist within the source paper. A promising direction is to empower the agent to generate visuals based on the paper's content, e.g., creating a flowchart to more clearly illustrate the background in the introduction.

F Broader Impacts

In this paper, we are the first to explore the integration of design and aesthetic principles into a multi-agent framework for academic poster generation, and propose a novel system named **PosterGen**. By mirroring the specialized workflow of professional designers, **PosterGen** achieves content fidelity that rivals the state-of-the-art while significantly enhancing the final poster's visual design and aesthetic quality. Though it may not fully eliminate the need for human fine-tuning, its core contribution lies in systematically embedding design principles into multi-agent design, a step often overlooked even in manual creation. The posters generated by our framework can thus serve as a firm reference for authors, which can greatly streamline their design process.

Within-Column Optimization Strategies:

1. Strategy A: Conservative Text Content Adjustment (for 80-100% utilization)
 - **When to use:** Column utilization is close to optimal range (80-100%)
 - **Actions allowed:**
 - MINIMAL text expansion: Add only 1-2 short phrases to underutilized columns (75-85%)
 - Aggressive text reduction: Significantly shorten content in overflow columns (>95%)
 - CONSERVATIVE APPROACH: Prefer slight underutilization over any risk of overflow
 - **Text Length Limits:**
 - Maximum per bullet: 25 words (count carefully)
 - Maximum sub-bullets: 2 per main bullet
 - Expansion limit: Add maximum 10-15 words total per section
 - Reduction target: Remove 30-50% of content from overflow sections
2. Strategy B: Section Management (for <80% or >100% utilization)
 - **When to use:** Column has severe underutilization (<80%) or overflow (>100%)
 - **Actions allowed:**
 - Add sections from structured sections: Use additional content from paper sections that fit the column's purpose
 - Remove less important sections: Remove sections with importance level = 3 or lower importance
 - **Section Removal Priority:**
 - NEVER remove: Method sections with key visual (importance level = 1)
 - NEVER remove: Core experimental results or main findings
 - Remove first: Supporting context, minor experiments, supplementary details (importance level = 3)
 - Remove second: Secondary analysis, additional background (importance level = 2)

Strict Constraints:

1. **NO CROSS-COLUMN MOVES:** Never change column assignment for any existing section
2. **PRESERVE READING FLOW:** Maintain left→middle→right logical progression
3. **SECTION ID PRESERVATION:** Never change section id, section title, visual assets, or other identifying fields
4. **IMPORTANCE RESPECT:** Never remove critical sections (importance level=1 or core results)
5. **TARGET UTILIZATION:** Achieve 85-95% utilization for each column

Input: {{structured sections}, {current story board}, {column analysis}}

Output Format:

Output the complete optimized story board JSON. Each section's 'text content' must be an array of complete strings only:

```
1 "text_content": [  
2     "* **Point Title:** Complete description text here",  
3     " - Supporting detail in complete sentences",  
4     "* **Another Point:** Full explanation without truncation"  
5 ]
```

Preserve all original structure and field names. Only modify content within string values.

Figure 16 Part 2 of the Balancer sub-agent prompt, detailing the specific optimization strategies, strict constraints, and the required input/output format.

We hope our research will advance this emerging yet meaningful field of automated scientific communication. We believe our work can substantially relieve researchers of the time and effort typically required for poster creation, allowing them to focus more on the one-on-one scholarly dialogue that poster sessions are meant to facilitate.

Color Agent: Theme Color Extraction

Extract a sophisticated theme color from an affiliation logo that will work well as a poster accent color.

Core Task: Analyze the provided affiliation logo and identify the most prominent, meaningful color that can serve as a poster theme color. This color should be:

- Representative of the organization's visual identity
- Suitable for poster design applications (text highlights, accents)
- Professional and readable when used on white backgrounds
- Harmonious for academic poster contexts

Color Extraction Guidelines:

1. Primary Color Identification:

- Look for the main brand color of the organization
- Ignore pure white, black, and very light grays (background/outline colors)
- Focus on colored elements that define the logo's visual identity
- Consider text colors, graphic elements, symbols, and emblematic elements

2. Color Suitability Assessment:

- Too Bright: If the main color is very bright/saturated (e.g., neon yellow #FFFF00), generate a more subdued version
- Appropriate Saturation: Aim for colors that are vibrant but professional
- Readability: Ensure the color provides sufficient contrast on white backgrounds for text

3. Color Adjustment Rules:

- If original color is too bright (lightness > 85% or saturation > 90%), reduce brightness by 15-25%
- If original color is too dark (lightness > 25%), lighten slightly for better visibility
- Maintain the color's hue character while optimizing for poster applications

Output Requirements: Return ONLY a JSON object with the following structure:

```
1  {
2      "extracted_color": "#1E3A8A",
3      "color_name": "Professional Navy Blue",
4      "adjustment_made": "reduced_brightness | lightened | none",
5      "original_color": "#0000FF",
6      "suitability_score": 8.5,
7      "reasoning": "Extracted the primary blue from the university emblem. Reduced
       brightness from bright blue to professional navy to ensure readability and
       sophisticated appearance on white backgrounds.",
8      "usage_notes": "Excellent for text highlights, section headers, and accent
       elements. Provides strong contrast while maintaining professional appearance."
9  }
```

Scoring Criteria (1-10 scale):

- **Contrast/Readability:** How well it works on white background
- **Professional Appearance:** Appropriate for academic/research contexts
- **Brand Representation:** How well it represents the organization
- **Poster Suitability:** Effectiveness for highlights and accents

Figure 17 Prompt for Color Agent to extract theme color from affiliation logo.

Font Agent: Keyword Extraction

Analyze poster content and identify keywords for strategic visual highlighting using three distinct formatting styles.

Input Data:

- Enhanced Narrative: enhanced narrative
- Curated Content: curated content

Core Task: For each section, identify keywords and assign them to specific highlighting styles based on their semantic importance and role in the research narrative.

Highlighting Style Categories:

1. BOLD + CONTRAST COLOR:

- **Purpose:** Core method/methodology names that represent the paper's unique contribution
- **Criteria:** Novel algorithms, architectures, or techniques introduced by this work; the main methodological innovation that defines the paper; must be unique to this research (not generic terms)
- **Limit:** Maximum 2 per section, prefer 1 if it captures the main contribution

2. BOLD:

- **Purpose:** Important quantitative results and core technical terms within each section
- **Criteria:** Performance metrics and numerical results (e.g., “**95% accuracy**”, “**5.2× speedup**”); key technical concepts central to understanding the section; architecture names, dataset names, established method names; word-level emphasis, not entire phrases
- **Limit:** Maximum 3 per section

3. ITALIC:

- **Purpose:** Defining terms, single-word emphasis, and foreign terminology
- **Criteria:** Technical terms being defined or introduced for the first time; single-word emphasis (e.g., “This was the *only* experiment”); foreign words, Latin terms, or specialized vocabulary; word-level application only, never entire sentences
- **Limit:** Maximum 2 per section

Output Format:

```
1  {
2      "section_keywords": {
3          "motivation": {
4              "bold_contrast": ["DP-CutMixSL"],
5              "bold": ["Vision Transformers", "privacy leakage"],
6              "italic": ["federated"]
7          },
8          "method": {
9              "bold_contrast": ["CutMix", "differential privacy"],
10             "bold": ["95% accuracy", "ResNet-50"],
11             "italic": ["only"]
12         },
13         "results": {
14             "bold_contrast": ["TransformerNet"],
15             "bold": ["top-1 accuracy", "5.2x speedup", "CIFAR-10"],
16             "italic": ["a priori"]
17         }
18     },
19     "formatting_summary": {
20         "total_bold_contrast": 4,
21         "total_bold": 7,
22         "total_italic": 3,
23         ...
24     }
25 }
```

Return a JSON object with the exact schema above that maximizes research impact through strategic visual emphasis.

Figure 18 Prompt for Font Agent to extract different types of keywords.

VLM-as-Judge Evaluation Template

You are an expert academic poster design critic. Your evaluation must be strict, detailed, and evidence-based. For each dimension, assess the poster against the specific examples provided in the 5-point scale. **High scores require adherence to professional design principles (positive examples). Low scores should be assigned when common design failures (negative examples) are present.** A poster that is merely functional but exhibits poor design choices must be scored significantly lower on design metrics than one that is both functional and visually excellent.

Please carefully examine the provided poster image and evaluate it across the following metrics. Provide a detailed explanation and score on a 5-point scale.

FOCUS AREA: < Focus Area >

Metric : < Metric Description >

- **Score 5 (Excellent):**
 - Descriptor: ...
 - Positive Examples: ...
- **Score 4 (Good):**
 - Descriptor: ...
 - Positive Examples: ...
- **Score 3 (Acceptable):**
 - Descriptor: ...
 - Negative Examples: ...
- **Score 2 (Poor):**
 - Descriptor: ...
 - Negative Examples: ...
- **Score 1 (Failed):**
 - Descriptor: ...
 - Negative Examples: ...

Output Format: Please provide your evaluation as a JSON array with exactly 1 object for the metric:

```

1 [ 
2   { 
3     "metric": "Core Graphic Principles",
4     "explanation": "Detailed analysis of the application of repetition, alignment, contrast, and proximity principles ...",
5     "score": 4
6   } 
7 ]

```

Evaluation Instructions: ...

Please evaluate the poster step by step according to these criteria.

Figure 19 Prompt Template for VLM-as-Judge Evaluation.

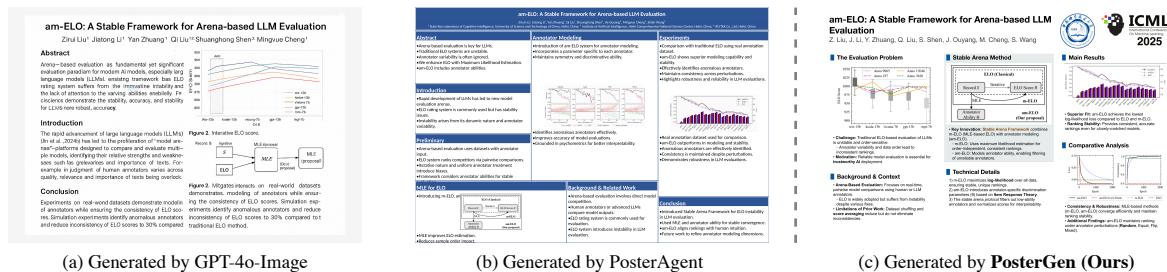


Figure 20 Qualitative results of Liu et al. (2025b).

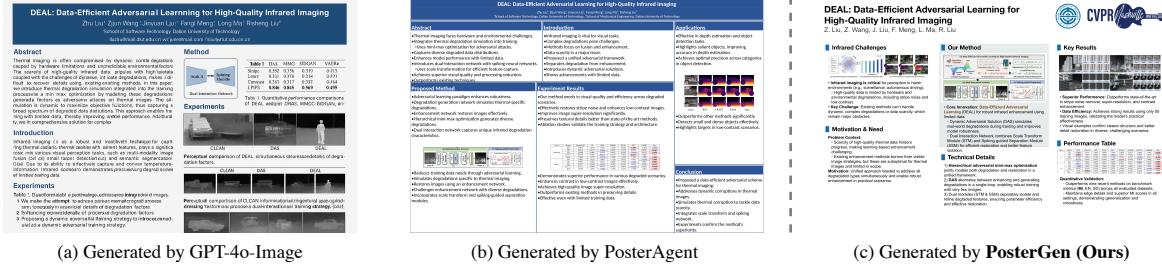


Figure 21 Qualitative results of Liu et al. (2025a).

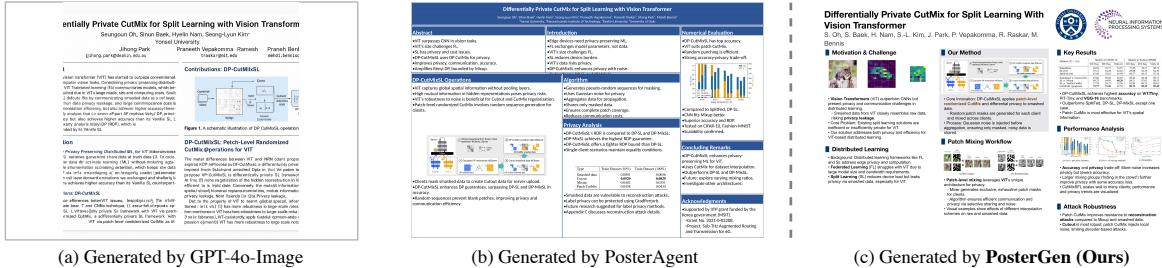


Figure 22 Qualitative results of Oh et al. (2022).

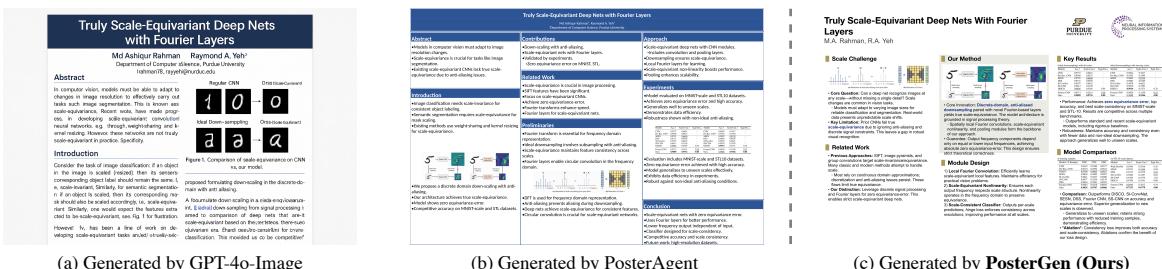


Figure 23 Qualitative results of Rahman and Yeh (2023).

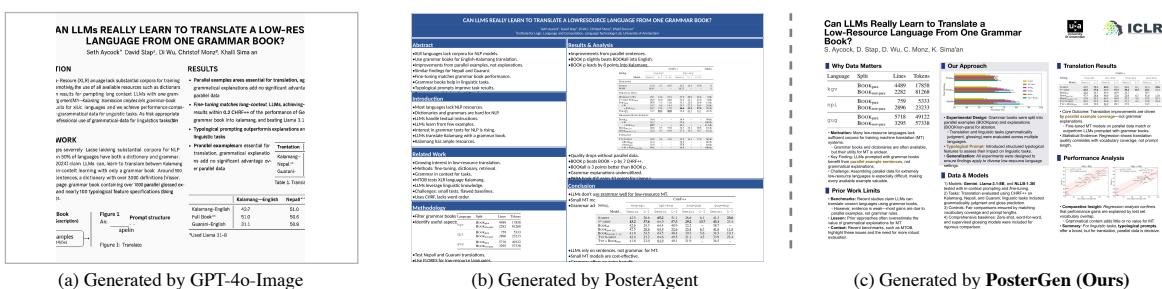


Figure 24 Qualitative results of Ayccock et al. (2024).

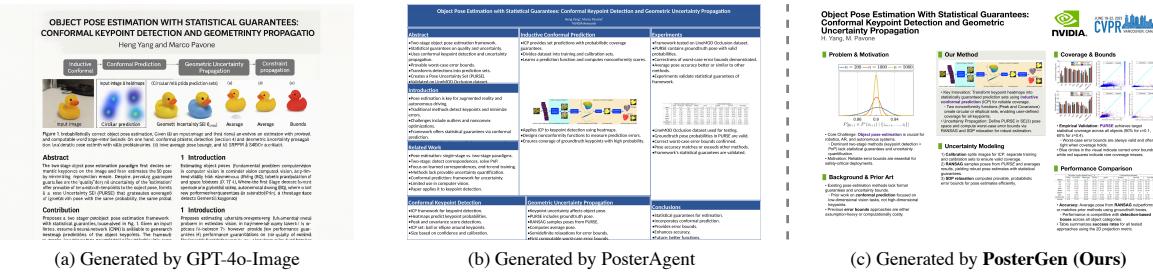


Figure 25 Qualitative results of Yang and Pavone (2023).

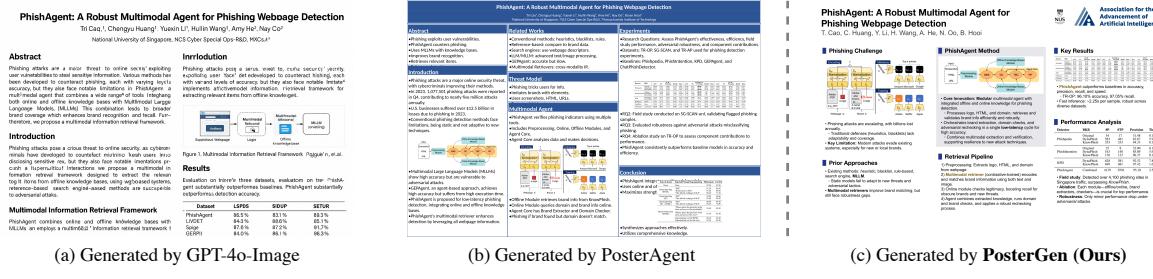


Figure 26 Qualitative results of Cao et al. (2025b).

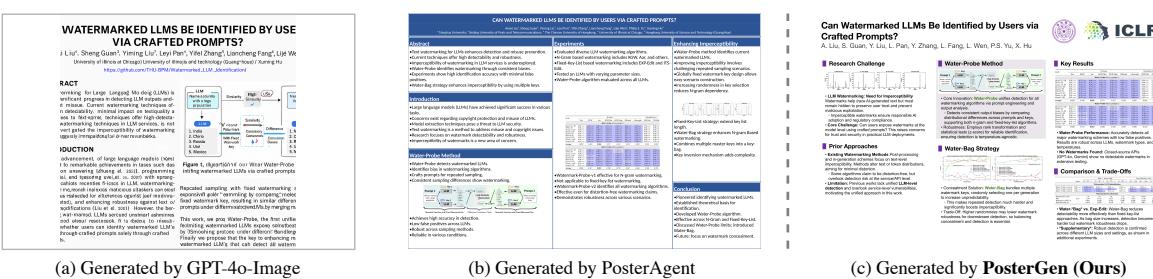


Figure 27 Qualitative results of Liu et al. (2024).