PPTAgent: Generating and Evaluating Presentations Beyond Text-to-Slides

Hao Zheng[1,2,*], **Xinyan Guan**[1,2,*], **Hao Kong**[3], **Jia Zheng**[1], **Hongyu Lin**[1]
Yaojie Lu[1], **Ben He**[1,2], **Xianpei Han**[1], **Le Sun**[1]
[1]Chinese Information Processing Laboratory, Institute of Software, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences
[3]Shanghai Jiexin Technology
{zhenghao2022,guanxinyan2022,zhengjia,hongyu,luyaojie}@iscas.ac.cn
{xianpei,sunle}@iscas.ac.cn haokong@knowuheart.com

Abstract

Automatically generating presentations from documents is a challenging task that requires balancing content quality, visual design, and structural coherence. Existing methods primarily focus on improving and evaluating the content quality in isolation, often overlooking visual design and structural coherence, which limits their practical applicability. To address these limitations, we propose *PPTAgent*, which comprehensively improves presentation generation through a two-stage, edit-based approach inspired by human workflows. *PPTAgent* first analyzes reference presentations to understand their structural patterns and content schemas, then drafts outlines and generates slides through code actions to ensure consistency and alignment. To comprehensively evaluate the quality of generated presentations, we further introduce *PPTEval*, an evaluation framework that assesses presentations across three dimensions: Content, Design, and Coherence. Experiments show that *PPTAgent* significantly outperforms traditional automatic presentation generation methods across all three dimensions. The code and data are available at https://github.com/icip-cas/PPTAgent.

1 Introduction

Presentations are a widely used medium for information delivery, valued for their visual effectiveness in engaging and communicating with audiences. However, creating high-quality presentations requires a captivating storyline, visually appealing layouts, and rich, impactful content (Fu et al., 2022). Consequently, creating well-rounded presentations requires advanced presentation skills and significant effort. Given the inherent complexity of presentation creation, there is growing interest in automating the presentation generation process (Mondal et al., 2024; Maheshwari et al., 2024) by leveraging the generalization capabilities of large language models (LLM).

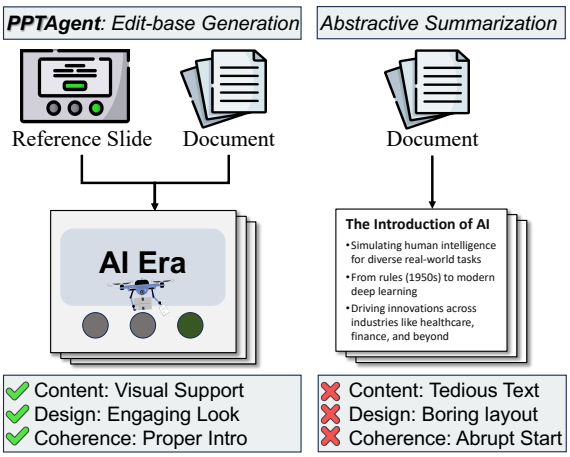

Figure 1: Comparison between our *PPTAgent* approach (left) and the conventional abstractive summarization method (right). Our method, which begins by editing a reference slide, aligns more closely with the human presentation creation process.

Existing approaches often adopt an end-to-end text-generation paradigm, focusing solely on textual content while neglecting layout design and presentation structures, making them impractical for real-world applications. For example, as shown in Figure 1, prior studies (Mondal et al., 2024; Sefid et al., 2021) treat presentation generation as an abstractive summarization task, focus primarily on textual content while overlooking the interactive nature of presentations. This results in simplistic and visually uninspiring outputs that fail to engage audiences.

However, automatically creating visually rich and structurally clear presentations remains challenging due to the complexity of data formats and the lack of effective evaluation frameworks. First, most presentations are saved in PowerPoint's XML format, which is inherently tedious and redundant (Gryk, 2022). This complex format poses signifi-

[*] These authors contributed equally