

アウトライン生成と評価・修正に基づく 二段階スライド生成フレームワーク

會田 光[†] 落合 桂一[†] 戸田 浩之[†]

[†] 横浜市立大学データサイエンス学部 〒236-0027 横浜市金沢区瀬戸 22-2

E-mail: †{d224001b,ochiai.kei.dk,toda.hir.xg}@yokohama-cu.ac.jp

あらまし プレゼンテーションは、ビジネスからアカデミックな場面に至るまで、情報伝達の主要な手段として広く用いられている。一方で、スライド作成には多大な時間と労力を要し、内容やレイアウトの品質が作成者の経験やスキルに依存しやすいという課題がある。一般に、スライド作成プロセスは、元文書から重要な情報を抽出・整理し、それらをどの順序・粒度・論理で提示するかを設計する「内容の整理・構造化」の段階と、構造化された内容をスライドとして視覚的に配置・表現する「視覚化・レイアウト生成」の段階から構成される。これらはいずれも一定の判断や試行錯誤を伴うが、特に視覚化・レイアウト生成の段階では、スライド全体の構成や各ページ間の関係性を考慮した設計が求められるため、品質を安定的に確保することが難しい。こうした背景のもと、近年では論文や報告書などの文書を入力として、スライド資料を自動生成する研究が進められている。しかし、既存研究の多くは、生成されたスライドのデザインやレイアウトの整合性を保つために、人手による修正や調整を前提としており、視覚化・レイアウト生成の品質を安定的に確保する点には依然として課題が残されている。本研究では、スライド作成の完全自動化に向けた第一歩として、内容の整理・構造化が既に与えられた状況において、視覚化・レイアウト生成がどの程度自動化可能であるかに着目する。提案手法では、スライドタイトルを起点として元文書内の関連箇所を探索し、その内容に基づいて各ページのスライドを生成する。さらに、ページ毎に含めるべき内容とそのレイアウトを示したデータを用いて大規模言語モデルによる反復的な評価と修正を行う二段階の生成フレームワークを採用する。これにより、参照スライドを直接的な雛形として用いることなく、過去資料やテンプレートに内在する構成・レイアウトの知見を活用しつつ、元文書との整合性を保ったスライド生成を実現することを目指す。

キーワード 大規模言語モデル (LLM)、スライド自動生成、レイアウト抽出、自動評価、自己修正・フィードバック学習

1 はじめに

現代のビジネスからアカデミックな場面に至るまで、プレゼンテーションは情報伝達の主要な手段として重要な役割を担っている。しかし、プレゼンテーション資料の作成には多大な時間と労力を要する。think-cell japan 株式会社の調査では、就業時間の約 5 分の 1 が資料作成に割かれているにもかかわらず、約 4 分の 3 の人が資料のわかりにくさや不備に起因する作業遅延や誤解を経験していることが示されている [1]。これらの問題は、単なる作業負荷の増大にとどまらず、組織全体の生産性や意思決定の質に悪影響を及ぼす要因となっている。

一般に、プレゼン準備におけるスライド作成プロセスは、

1. **スライドの元となる文書から重要な情報を抽出・整理する段階**
2. **抽出した内容をどの順序・粒度・論理で提示するかを設計する構成段階**
3. **構造化された内容をスライドとして視覚的に配置・表現する視覚化・レイアウト生成段階**

の三つの段階に分けて整理される [2]。以下では、(1) および (2) をまとめて「内容の整理・構造化」、(3) を「視覚化・レイ

アウト生成」と呼ぶ。これらはいずれも一定の判断や試行錯誤を伴い、特にスライド作成では、「何を・どの単位で・どの順番で・どのように示すか」といった複数の判断を統合的に行う必要がある。そのため、スライド作成には多くの時間を要し、経験やスキルへの依存が大きくなりやすい。

これらの課題を支援・効率化することを目的とした研究が数多く行われてきた。近年では、大規模言語モデル (Large Language Model, LLM) を用いて、スライドに記載するテキストを、スライドの元となる文書 (以下、元文書) から取捨選択・要約した上で自動生成する手法や、スライドの編集履歴やスライド作成時に参照すべき過去スライドを人手によって用意し、それらを基にスライド作成を支援する研究が進展している [3-5]。

例えば、Doc2PPT [3] では、主に「内容の整理・構造化」の段階に着目し、元文書からスライドに含める情報を抽出・要約した上で、スライド構成を自動的に生成する手法が提案されている。一方、PPTAgent [4] では、スライド作成プロセス全体の自動化が試みられている。この手法では、既存のスライドから、各ページに含める情報やその配置、提示順序といった構成およびレイアウトに関する特徴を参考にすることで、構成的・

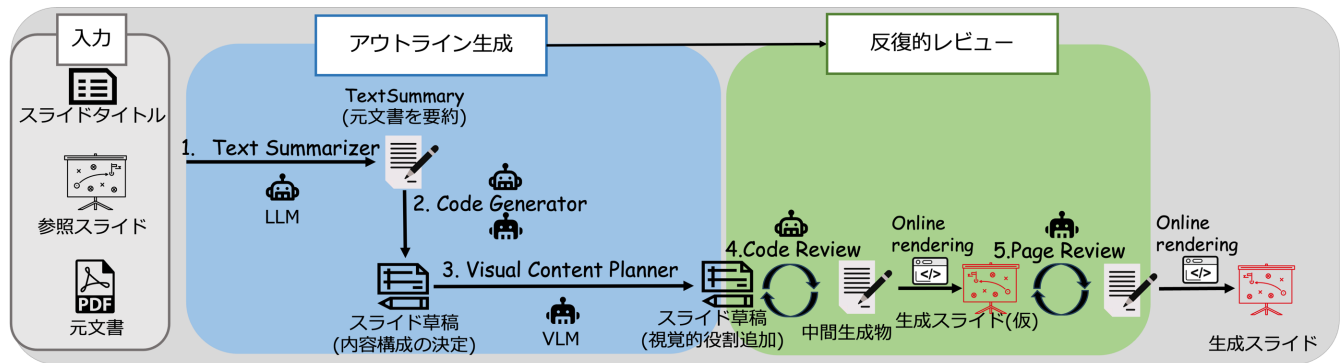


図 1: アウトライン生成 (左) と反復的レビュー (右) からなるスライド生成フレームワークの全体像

視覚的な一貫性の向上が報告されている [6]. 本研究では、このように構成やレイアウトの設計指針として参照されるスライドを「参照スライド」と呼ぶ。しかし、PPTAgent においても生成結果に対して人手による修正が必要となる場合が多く、特に視覚化・レイアウト生成の品質を安定的に確保する点には課題が残されている。

以上の研究動向を踏まえると、スライド作成の完全自動化に向けては、内容の整理・構造化と視覚化・レイアウト生成という性質の異なるプロセスを切り分けた上で、それぞれの実現可能性を段階的に検証することが重要であると考えられる。

そこで本研究では、スライド作成の完全自動化に向けた第一歩として、内容の整理・構造化が既に与えられた状況において、視覚化・レイアウト生成がどの程度自動化可能であるかを評価対象とする。具体的には、論文発表を対象とし、発表内容を構造化した各ページに対応するスライドタイトル列、元文書、および構成・デザインの参考とする参照スライドが入力として与えられる状況を前提とする。なお、スライドタイトルは参照スライドから取得するものではなく、人間が発表全体の構成を踏まえて事前に設計する入力情報とする。この設定により、構成設計に伴う不確実性を排除した上で、視覚化・レイアウト生成そのものの実現可能性と限界を独立に評価することが可能となる。

提案するフレームワークは、これらの入力情報を基に二段階の生成プロセスを採用する。第一段階では、タイトル列と元文書の対応関係を明示的に制約したプロンプト設計に基づきアウトラインを生成する。第二段階では、生成されたアウトラインに基づいて作成したスライド草稿に対し、評価および修正を反復的に行うことでスライド生成を行う (図 1)。本手法では、参照スライドを生成スライドの直接的な雛形として用いるのではなく、構成およびレイアウト設計に関する指針として限定的に活用することで、元文書に基づく内容生成との役割分離を図る。これにより、過去に作成された類似スライド資料やスライドテンプレートに内在する構成・レイアウトの知見を活用しつつ、元文書との整合性の高いスライド生成を実現することを目指す。

本研究の貢献は主に以下の三点である。第一に、元文書との整合性を維持しながら、スライド構成の検討を支援するフレームワークを整理・提示する点である。第二に、問題箇所限定

した修正を行う設計により専門的なスライド作成スキルを持たない利用者であっても、一定の品質のスライドを短時間で作成できるようになることである。第三に、組織内で過去に作成された類似資料やテンプレートの再利用を促進する点である。これらにより、本研究は、研究的意義と実務的価値の両面から、スライド作成支援に関する新たな知見を提供する。

2 既存研究

本章では、レイアウト抽出、スライド生成、生成されたスライドの自動評価手法、および自己改善手法に関する既存研究を概観し、本研究の位置づけを明確にする。

2.1 レイアウト抽出

近年の文書・スライドを対象としたレイアウト解析や視覚文書理解に関する研究では、視覚的な特徴を活用することで、文書やスライドの構造理解およびレイアウト抽出を行う手法が提案されている。LayoutLM [7,8] は、文書画像そのものを生成するのではなく、既存の文書画像から得られたテキストおよびその位置情報を入力として、文書内要素の構造的役割を推定することに主眼を置いた手法である。また、Donut [9] に代表される End-to-End 型手法では、文書画像を直接入力とし、OCR (画像中に含まれる文字を検出し、機械可読なテキストに変換する技術) を介さずに、文書構造を表す構造化されたテキストを生成する。この手法では、文書画像を視覚的特徴量に変換した上で、文書構造を表す記号や項目を順に予測することで、文書全体の構造を表現する。

しかし、これらの多くは論文や帳票などのレイアウトが比較的規則的である定型文書を主なレイアウト抽出タスクの対象として設計されている。そのため、テキストや図形が自由に配置され、ページごとに構造や役割が大きく異なるスライド特有のレイアウトを直接扱うことは難しい。スライド理解に特化した研究としては、視覚要素の近接性や配置関係を利用し、スライド内要素の入れ子状の構造を推定する手法が提案されている [10] が、主に単一ページ内の構造解析に焦点が当てられている。

こうした従来の静的な構造解析手法とは別に、近年では、LLM や視覚情報と言語情報を統合的に扱う視覚言語モデル (Vision-Language Model, VLM) をエージェントとして用い、

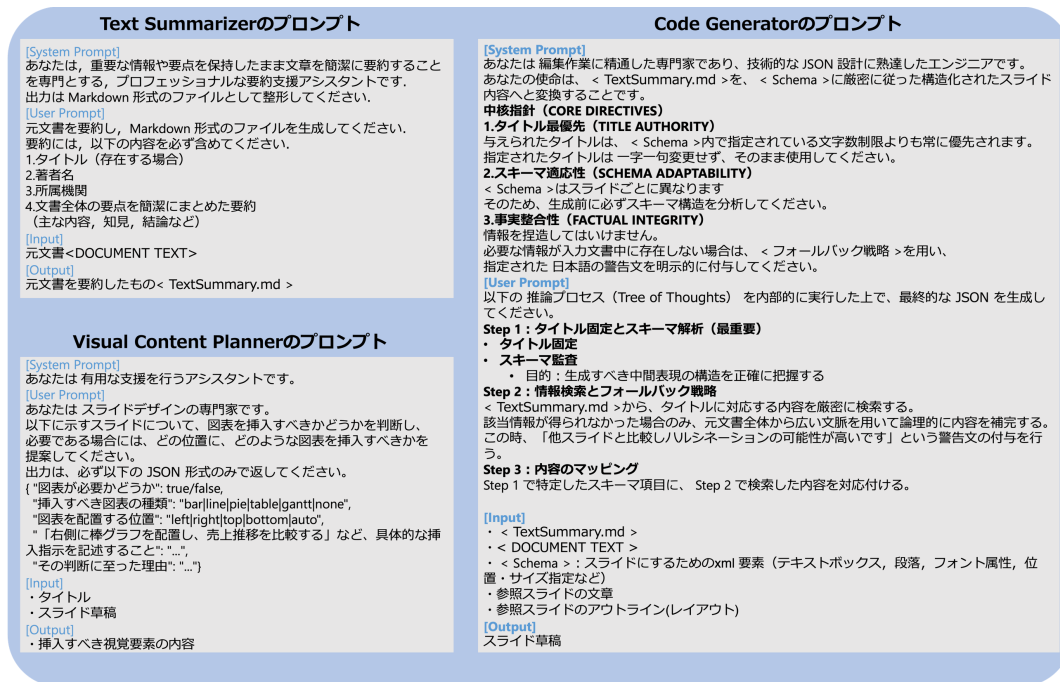


図 2: アウトライン生成段階で用いるプロンプト

スライドを段階的に観察・判断しながら構造を解析する Agentic Workflow が注目されている [11, 12].

2.2 スライド生成

近年では、LLM を用いてスライド構成を直接生成することで、スライド特有の簡潔な表現や視覚的可読性を考慮した手法が広く検討されている [13–15]. PPTAgent [4] は、複数のエージェントが役割分担しながら推論を行うマルチエージェント構成を採用し、構成検討、情報収集、内容生成を段階的に実行することで、論理一貫性の高いスライド生成を実現している。また、同手法では、参照スライドやテンプレートを参照することで、スライド構成や表現形式の一貫性を保つ工夫がなされている。しかし、これらの手法は、構成生成を含みつつも主に内容生成に焦点を当てており、参照スライドに基づく構成の妥当性や、生成されたスライドが視覚的制約に適合しているかを厳密に保証する点には課題が残されている。

2.3 生成されたスライドの自動評価手法

生成されたスライドの品質を客観的に測定するための自動評価手法は、内容の適合性を評価する手法と、視覚的可読性やレイアウトを評価する手法がある。内容の評価においては、自然言語処理 (Natural Language Processing, NLP) 分野で広く用いられてきた ROUGE [16] や METEOR [17] などの語彙的類似度指標が利用されてきた。これらの指標は、生成文と参照文の間で共通する単語や表現の割合を計算することで、内容の一致度を定量化するものであり、文章生成タスクにおいて一定の有効性を示している。しかし、スライドでは文章をそのまま記述するのではなく、要点を簡潔な箇条書きとしてまとめることが多く、元文書と語彙的に一致しない表現が用いられることがある。そのため、これらの指標では、スライド特有の要約表

現や論理構成の妥当性、分かりやすさといった品質を十分に評価できないという課題がある。

デザインやレイアウトの評価に関しては、要素間の整列状態、余白の占有率、文字密度、フォントサイズの可読性などを幾何学的・統計的特徴として定義し、定量的にスコア化する研究が進んできた。例えば、SlideAudit [18] は、スライドの視覚的特徴に基づいてデザイン品質を自動評価するための指標体系とデータセットを提示しており、レイアウト破綻や視認性低下を検知するための基盤的なフレームワークを提供している。

さらに、近年では LLM や VLM を評価者として活用する「LLM-as-a-judge」の概念が普及している。PPTAgent [4] や PreGenie [5] では、生成されたスライドに対して LLM や VLM が論理的一貫性や視覚的な不整合を自然言語でフィードバックする手法が採用されている。これにより、従来の固定的な指標では困難であった、文脈に応じた柔軟な品質評価が可能となった。しかし、これらの多くは画像やテキストに対する抽象的な評価に留まっており、構造的な整合性を厳密かつ高速に判定する手法についてはさらなる検討の余地がある。

2.4 生成されたスライドに対する自己改善手法

生成されたスライドを評価し、段階的に品質を向上させる自己改善や自己反省のフレームワークは、LLM の推論能力を引き出す手法として注目されている。Self-Refine [19] は、生成・フィードバック・修正のサイクルを反復的に実行することで、生成するスライドの品質を向上させられることを示した。

一方で、既存の自己改善手法の多くは、自然言語による抽象的な指示や生成後の視覚的確認を基に修正しており、テンプレートに基づく配置制約や文字数制限といった構造的ルールを明示的に扱う点には課題がある。そのため、スライド特有の制約を考慮した体系的な修正手法の確立が求められている。

3 提案手法

本章では、本研究で提案する反復的レビューに基づくスライド生成フレームワークについて説明する。提案手法は、人間がスライドを作成する際に、まず全体構成を検討し、その後に内容や表現を調整するという作業工程を模倣することで、利用目的に合ったスライドを生成することを目的として設計されている。具体的には、図 1 に示すように、入力情報に基づきスライド全体の構成を決定する「アウトライン生成」を行い、そのアウトラインに従ってスライドを生成する。その後、生成されたスライドに対して、レビューと自己修正を反復的に行う「反復的レビュー」を適用し、内容や表現を段階的に洗練させることで、最終的なスライドを得る。段階的なスライド生成により、内容の正確性と視覚的な一貫性の両立を図る。

3.1 アウトライン生成

まず、図 1 左部の青色領域に示す「アウトライン生成」を行い、スライドタイトル列、参照スライドおよび元文書を基に、スライドの各ページの内容構成と視覚的役割を決定する。この段階では、スライドの内容となるテキストの生成、視覚要素となる図表挿入の設計、スライドのレイアウト生成といった役割を分担した複数の生成モジュールを用いることで、スライド全体の初期構成を段階的に生成する。具体的には、元文書の要約を行う Text Summarizer、スライド草稿の生成を行う Code Generator、および挿入すべき視覚要素の内容を文章として生成する Visual Content Planner に対して、スライドタイトルを起点とした内容抽出および視覚要素設計を行うための専用プロンプトを設計している。これにより、各ページにおけるテキスト内容と視覚的役割を段階的かつ明示的に決定することが可能となる。

3.1.1 Text Summarizer

Text Summarizer の役割は、入力するスライドタイトルに基づいて元文書中の該当箇所を抽出し、スライド用に要点を整理・要約することである。図 2 左上に示すプロンプトでは、必ず含める要素を明示的に指定することで、スライド作成に必要な重要情報が欠落しないよう制約を与えている。これにより、これにより、元文書の文脈を保ったまま、スライドの目的に沿った要約結果 (TextSummary) を出力する。

3.1.2 Code Generator

Code Generator の役割は、TextSummary を基に、PowerPoint 形式のスライド草稿の生成を行うことである。図 2 右に示すプロンプトでは、まず TextSummary の結果からスライドタイトルに対応する内容を厳密に検索し、該当する情報が存在する場合はその範囲に基づいて内容を構成する。該当情報が得られなかった場合に限り、スライドタイトルの文脈を拡張した上で探索を再実行することでハルシネーションを抑制している。次に、参照スライドから抽出したスライド構造を利用し、各要素に対応するスライド領域に割り当てることで、PowerPoint 形式のスライド草稿 (レビュー前のスライド) を生成する。

3.1.3 Visual Content Planner

Visual Content Planner は、Code Generator によって決定されたスライド草稿の内容を入力として、該当ページにおいて視覚的な補助として画像や図表を用いることが望ましいかを判定し、必要に応じて挿入すべき視覚要素の内容を文章として生成する。図 2 左下に示すプロンプトでは、挿入すべき図表の種類や具体例、およびその判断に至った理由を自然言語で明示的に記述するよう制約を与えている。これにより、後続のレビュー工程において、テキスト内容と視覚要素の対応関係を扱うことが可能となる。

3.2 反復的レビュー

次に、生成されたスライド草稿に対して、図 1 右部の緑色領域に示す反復的レビューを行い、内容および視覚的観点から品質評価を実施する。本研究では、スライド品質に関わる問題が、論理構成や情報内容に起因する内容面の問題と、文字量やレイアウトといった視覚的可読性に起因する視覚面の問題とで性質が異なる点に着目し、Code Review と Page Review の二種類のレビューを行う設計としている。

- **Code Review** : スライド草稿のテキストを対象とし、既存テンプレート由来の装飾要素 (箇条書きや段落番号など) と生成要素との不整合、論理構成の不整合、冗長表現、情報過多などを検出する。
- **Page Review** : 生成されたスライドを対象とし、文字量超過、視覚的可読性の低下といった視覚的問題を検出する。

Code Review (左) と Page Review (右) の役割分担を図 3 に示す。また、反復的レビューにおける評価対象の違いを図 4 に示す。レビューの結果、問題が検出されたページに対してのみ再生成を行うことで、不要な再生成を抑制し、品質改善を行いつつ生成時間の増大を防ぐ。

これにより、従来の一括生成型手法では困難であった、文書内容に基づくスライド構成の妥当性と、各ページの視覚的制約を踏まえた品質制御を両立した生成手法を実現する。さらに、問題箇所に限定した修正を行う設計により、生成品質を維持しつつ生成時間の増加を抑制できる点に本手法の特徴がある。

4 実験設定と評価指標

本章では、提案手法の有効性を検証するため、実験設定および評価指標について述べる。本研究では、生成結果の品質および実行効率を評価対象とし、人間による主観評価に基づく定性評価、LLM と人間による定量評価、ならびにスライド生成に要する時間の比較を評価項目として設定する。なお、スライド生成および評価に用いる LLM はすべて GPT-4o (version: 2024-11-20) に統一し、モデル差による影響を排除した条件下で評価を行う。

4.1 評価用資料

今回は提案手法の有効性を検証するため、実際のスライド作成プロセスの一例として、研究発表用の元文書および参照用スライドを用いた。元文書としては、研究室所属メンバーが作成

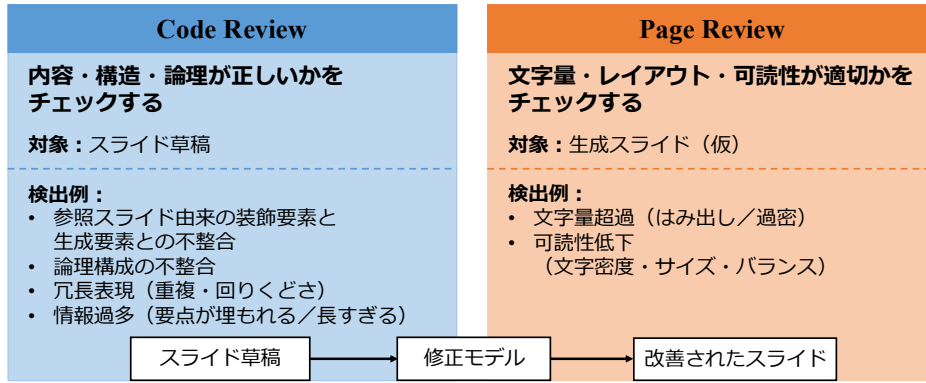


図 3: Code Review（左）と Page Review（右）の役割分担

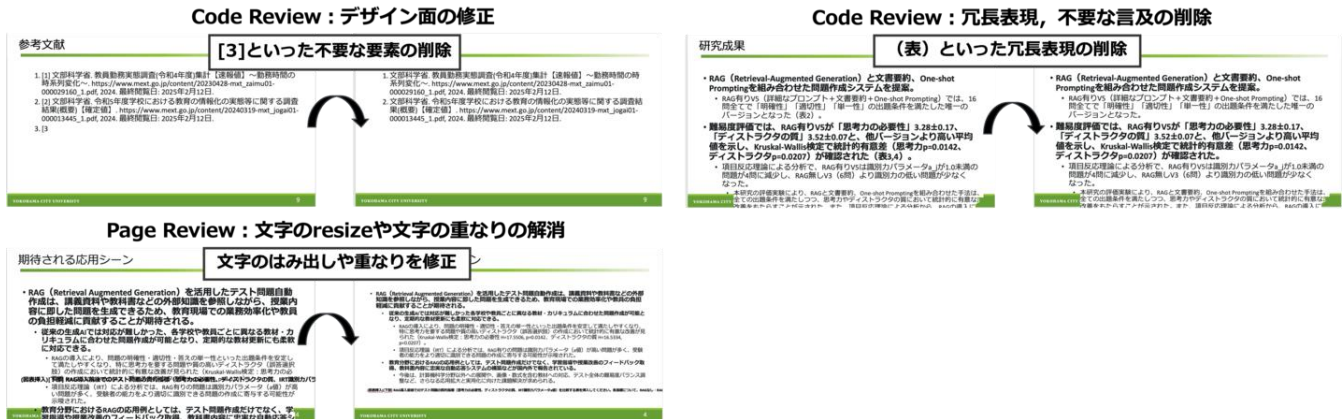


図 4: Code Review と Page Review における評価対象の例

した学会での論文を 6 件使用した。これらの文書は、A4 で 8 ページ程度書かれた学会論文になっており、研究背景、既存研究、提案手法、実験結果といった、一般的な学会での予稿論文に共通する構成要素を備えている。そのため、構成や表現のばらつきが比較的小さい。

参照スライドとしては、NTT グループが主催する研究開発成果の展示イベントである NTT R&D FORUM 2025 の公式 Web ページに掲載されている展示一覧¹を基に作成したスライド資料を 1 件用いた。当該スライド資料は、背景、研究目的、技術的特徴、想定応用といった要素が、スライドのタイトルによって明確に整理され、企業として統一された構成・表現方針に基づいて作成された資料である。本研究では、参照スライド資料を生成結果の正解（Ground Truth）としては扱わず、各ページのタイトル例と構成・情報配置の設計方針を与えるための参考資料として用いる。

また、スライド生成時には、参照スライドから取得した「利用したいタイトルが記載されたページ番号」に加え、生成対象として明示的に指定したスライドタイトル列を入力として与えた。本実験では、スライドタイトル列として [研究課題, 研究内容 1, 研究内容 2, 研究成果, 参考文献] を用いており、これらのタイトルを起点として、元文書から対応する内容を抽出・構成することでスライド生成を行っている。

¹ : <https://www.rd.ntt/forum/2025/exhibitions/>

これらの元文書および参照スライドを用いてスライド生成を行い、生成されたスライドに対して内容および視覚的観点からの評価を実施した。

4.2 評価方法

本研究では、生成されたスライドの品質を定量的に評価するため、LLM を評価者として用いた自動評価手法と人間による主観評価を採用する。スライド生成タスクにおける品質は、情報の正確性や語彙の一致度のみによって決まるものではなく、情報量の適切さ、視覚的可読性、ページ間の一貫性といった複合的な要因に依存する。このような特性を踏まえ、本研究では、人間による主観評価とこれに近い観点からスライド品質を多面的に捉える手法として、LLM を用いた自動評価を導入する（図 5）。評価は、ページ単位での品質と、スライド全体としての品質の両面から行う。

各ページ単位の評価指標

- **Page Design（ページデザイン）**：レイアウトのバランス、文字量や余白の適切さ、視線誘導の自然さ、および全体的な視覚的可読性を評価する。
- **Text Coherence（テキスト整合性）**：簡条書き間の論理的な一貫性、冗長表現や矛盾の有無、日本語としての自然さを評価する。
- **Text - Image Relevance（テキスト-画像関連）**：Visual Content Planner を用いて挿入した図表の内容をテキ

[System Prompt]

あなたは、優れた「スライド記述者」です。渡されたスライド画像を分析し、以下の項目について詳細な「記述テキスト」を出力してください。客観的な事実のみを記述してください。

1. テキスト内容：タイトル、箇条書き、本文のすべての文字情報
2. デザイン構造：レイアウト配置、余白、フォント、配色
3. 図表・画像：何が描かれているか、(図表挿入)の指示がある場合の位置
4. 論理構造：タイトルと本文の整合性。

出力はプレーンテキストで行ってください。

[Reasoning Prompt]

あなたは冷徹な審査員です。「記述テキスト」を読み、厳格にスコアリングしてください。

< 採点ルール >

1. 差をつける：必ず点数に差をつける(同点回避)。
2. 忖度なし：悪いものは1~3点をつける。

< 評価項目 (1-10点) >

1. Page Design：レイアウト、余白、美しさ。
2. Text Coherence：日本語の自然さ、論理。
3. Text-Image Relevance：図表指示の適切さ(画像なしは0点)。
4. Structural Appropriateness：タイトルと本文の整合性。
5. Page Consistency：用語や表記の統一、レイアウトやトーンの一貫性、およびスライド全体としてのストーリーの流れ。
6. Content Coverage / Faithfulness (元文書との整合性)：元文書(PDF)に記載された重要な情報が適切に反映されており、情報の欠落や不正確な内容が含まれていないか。

[Input]

生成されたスライド

[Output]

各評価項目の点数

図 5: LLM による定量評価のプロンプト

ストで記述したものが、本文の内容を適切に補完・説明しているかを評価する。

- **Structural Appropriateness (構造妥当性)**：スライドタイトルに対して内容が適切であるか、および「1 ページ 1 メッセージ」の原則が保たれているかを評価する。

スライド全体の評価指標

- **Page Consistency (ページ間の一貫性)**：用語や表記の統一、レイアウトやトーンの一貫性、およびスライド全体としてのストーリーの流れを評価する。
- **Content Coverage / Faithfulness (元文書との整合性)**：元文書に記載された重要な情報が適切に反映されており、情報の欠落や不正確な内容が含まれていないかを評価する。

LLM と人間による評価は、各指標について段階的なスコアとして出力され、各スコアは 1 から 10 の範囲で算出され、値が大きいほど品質が高いことを示す。これらのスコアを集約することで、スライドの総合的な品質を定量的に算出する。このように、視覚的側面、内容の構造、および文書忠実性を統合的に評価することで、スライド生成手法の性能をより包括的に分析することを可能にする。なお、比較にあたっては、LLM と人間による定量評価に加え、生成されたスライドを確認する定性的な分析も行う。

4.3 比較対象

本研究では、提案手法の有効性を検証するため、既存のスライド自動生成手法である PPTAgent [4] を主要な比較手法として用いる。

また、提案手法において新たに設計したプロンプトの効果を明確化するため、PreGenie において提示されている各エージェントのプロンプトをそのまま適用した手法を Prompt-based

Method として実装し、比較対象に含める。ここで用いる PreGenie のプロンプトは、Text Summarizer, Code Generator, Visual Content Planner および Code Review に対応するプロンプト一式である。

5 結果

本章では、スライド作成プロセスのうち、内容の整理・構造化が事前に与えられた状況を前提とし、視覚化・レイアウト生成の段階に着目した評価結果を示す。

5.1 定性評価 (視覚的な評価)

本節では、生成結果を実際に比較することで、各手法が持つ特徴的な傾向や課題を定性的に分析する。提案手法、Prompt-based Method および PPTAgent による定性的比較結果を図 6 に示す。

Prompt-based Method では、赤枠で示すように、参照スライドに含まれる内容をそのまま引き継いでしまう問題が確認された。これは、参照スライドの構造や記述内容を十分に抽象化せずに生成処理が行われることに起因すると考えられる。また、青枠で示すように、スライドタイトルに対応する内容が元文書中に明示的に存在しない場合、該当セクションが空白のまま出力される傾向が見られた。このことから、Prompt-based Method は、元文書や参照スライドへの依存度が高く、柔軟な内容補完や構成調整が困難であることが示唆される。

一方、PPTAgent では、緑枠で示すように、参照スライドに由来する装飾要素と生成された内容との間で不整合が生じるケースが確認された。具体的には、箇条書きが二重構造となるなど、参照スライド側の表現と生成要素が競合する現象が見られる。また、全体を通して記述内容が抽象的に留まり、元文書に含まれる具体的な情報や技術的詳細が十分に反映されていない傾向があった。

これに対し、提案手法では、参照スライドを活用しつつもスライドの内容が反映されることを防ぎ、スライドタイトルと本文との対応関係を保った構成が一貫して生成されていることが確認された。さらに、テンプレート由来の装飾要素と生成内容との整合性も保たれており、箇条書き構造や情報粒度の面でも、人手で作成されたスライドに近い自然な表現が得られている。

以上の結果から、提案手法は、既存手法が抱える参照スライドへの依存や構造的な不整合といった課題を軽減し、内容の具体性と視覚的一貫性の両立を実現していることが、定性的に確認された。

5.2 定量評価 (LLM と人間による主観評価)

図 7 に、提案手法と PPTAgent による定量評価の結果を示す。図の上側は人間による主観評価の結果、下側は GPT-4o による自動評価の結果を表している。

人間による主観評価では、Page Design を除くすべての評価項目において、提案手法が PPTAgent よりも良いスコアを示した。Page Design のスコアが低いのは、対象となる情報量ができる限り削減しなかったため、文字量が増え、視覚的にわかりづ

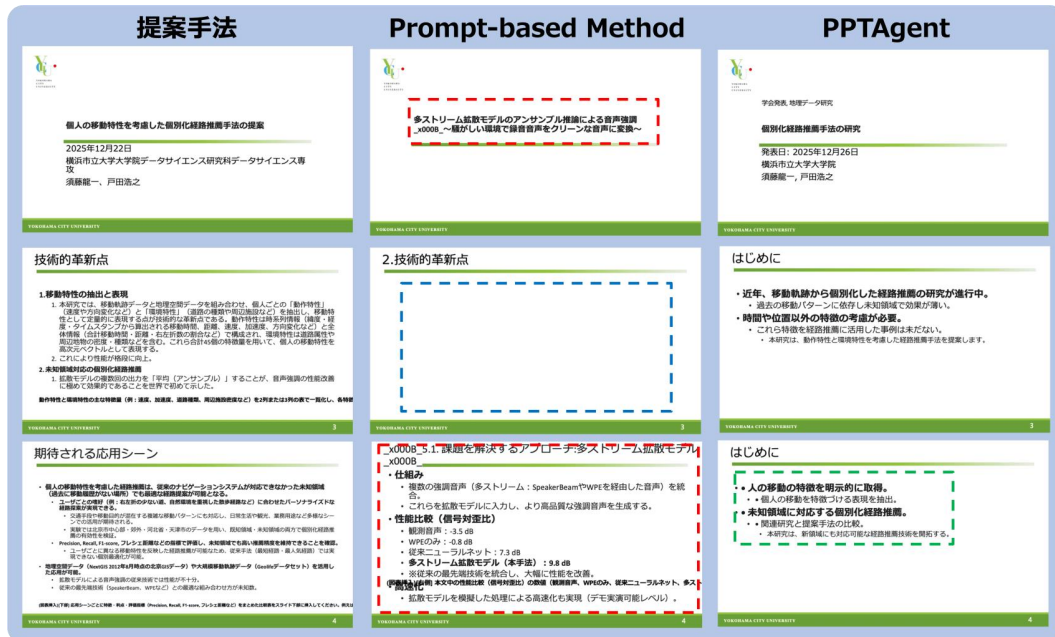


図 6: 提案手法, Prompt-based Method, PPTAgent による定性的比較結果レイアウトおよび内容上の誤りについては, 色付きの枠で強調して示している. 参照スライドに含まれる内容をそのまま引き継ぐ (赤枠), 該当セクションが空白のまま出力される (青枠), 参照スライドに由来する装飾要素と生成された内容との間で生じる不整合 (緑枠)

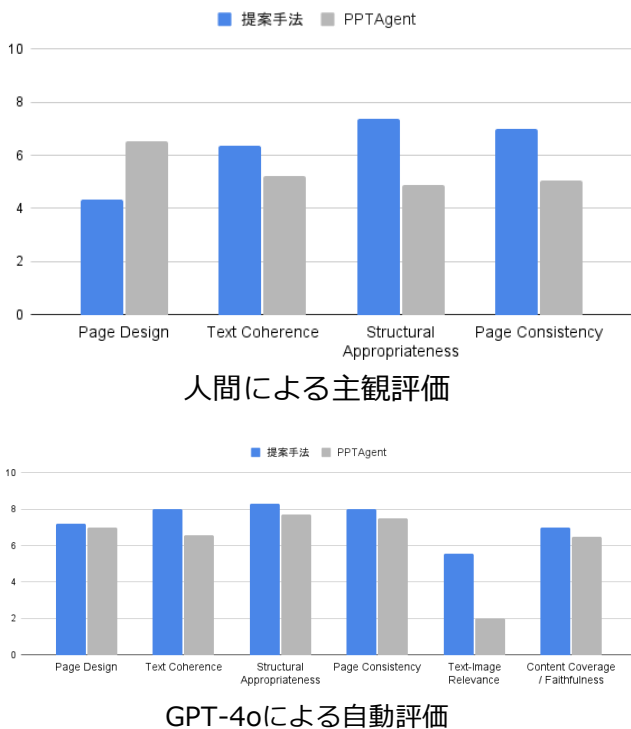


図 7: 提案手法と PPTAgent による定量評価の結果: 上: 人間による主観評価, 下: GPT-4o を用いた自動評価

らなくなってしまっていたためであると考えられる. よって, 文書要約技術等を組み合わせ, 重要な部分だけを端的に示すことで改善が見込まれる. 一方, GPT-4o による自動評価では, すべての評価項目において提案手法の方が高い評価結果となった. このことから, 提案手法は, 人間評価および GPT-4o による自

表 1: 提案手法および PPTAgent によるスライド生成に要した時間の比較

手法	スライド生成に要した時間 [分]
提案手法	8.7
PPTAgent	43.2

動評価の双方において, 総合的に概ね高い品質を持つスライドを生成できていることが確認された. 特に, Text Coherence, Structural Appropriateness, および Page Consistency に着目すると, 人間評価および GPT-4o による自動評価の双方において, 提案手法が PPTAgent より明確に良い結果が得られた. これらの結果から, 提案手法では, スライドタイトルに対する内容の適切性が高く保たれていることに加え, 用語や表記の統一, レイアウトやトーンの一貫性, およびスライド全体としてのストーリーの流れが, より良好に構成されていることが定量的に確認できる.

以上より, スライドタイトルが事前に与えられるという本実験設定のもとでは, 提案手法は, 構造的整合性やスライド全体の一貫性といった観点において, 既存手法である PPTAgent よりも優れたスライド生成を実現していることが, 人間評価および GPT-4o による自動評価の双方から示された.

5.3 スライド生成に要する時間の比較

提案手法および PPTAgent によるスライド生成に要した時間を比較した結果を表 1に示す. 本実験における計算機環境としては, CPU に Apple M2 Pro, 16GB のメモリを搭載した端末を使用した. OS には macOS 26.2 を用い, ネットワー

ク環境には家庭用無線 LAN を使用している。この結果から、提案手法では平均 8.7 分でスライド生成が完了したのに対し、PPTAgent では平均 43.2 分を要することが確認された。

この結果から、提案手法は、一定の品質を備えたスライド構成を効率的に生成できるだけでなく、実運用を想定した場合においても、生成時間の観点から高い実用性を有することが示された。

6 考察

定量評価の結果から、提案手法は Code Generator におけるプロンプト設計の影響により、スライドタイトルとの内容一貫性を重視したスライド生成を行う傾向を持つことが示唆された。この傾向は、GPT-4o による自動評価においては一貫した高評価として現れたが、人間による主観評価では、Page Design のスコアが相対的に低い場合が確認された。この結果から、情報密度の高さに対する評価が、自動評価と人間評価とで異なる可能性が示唆される。

さらに、GPT-4o による自動評価では、手法間の差が明確に表れにくい傾向が確認された。このことから、現在の評価プロンプトでは、人間による主観評価で重視される観点を十分に反映できていない可能性があり、評価基準をより具体化することで、人間評価に近い自動評価が実現できると考えられる。

また、スライド生成に要した時間の比較結果から、提案手法は PPTAgent と比べて、大幅に短い時間でスライド生成を完了できることが確認された。これは、参照スライドを用いたアウトライン生成によってスライド草稿の段階でスライド全体の構成を明確化し、生成後の評価・修正を問題箇所限定して行う設計を採用していることに起因すると考えられる。

7 まとめ

本研究では、文書からスライドを自動生成する課題に対して、参照スライドを活用したアウトライン生成と、スライド生成後の評価・修正を組み合わせた二段階のスライド生成フレームワークを提案した。提案手法は、参照スライドの構造や記述内容を十分に抽象化せずに生成処理が行われるという既存手法の課題を抑制し、スライドタイトルと本文の対応関係や、スライド全体としての内容およびレイアウトの一貫性を重視した点に特徴がある。

定性評価、定量評価およびスライド生成に要した時間の比較の結果、提案手法はスライドタイトル列が入力として与えられる条件下において PPTAgent と比較し、スライド生成に要する時間を大幅に短縮しつつ、スライドタイトルに対する内容の適切性や、スライド構造およびスライド全体としてのストーリーの流れがより良好に維持されていることが示された。

一方で、定量評価の結果から、提案手法はスライドタイトルとの内容一貫性を重視する傾向を持ち、スライド内の情報密度が高くなる場合があることも明らかとなった。このことから、スライド内の情報量を適切に制御し、重要度に応じた内容選択や要約を行うことが今後の課題である。

謝辞

本研究の一部は JSPS 科研費 JP24K03046 の助成を受けたものです。

文献

- [1] PR TIMES. think-cell japan 株式会社, “日本人ビジネスパーソンの残業は『負の資料修正スパイラル』がつくっていた!? think-cell japan が 2024 年の働き方を考える残業実態調査を実施. <https://prtimes.jp/main/html/rd/p/000000012.000103466.html>, May 22, 2025.
- [2] Johns Hopkins University. Presentation skills workbook. <https://carey.jhu.edu/sites/default/files/2023-02/presentation-skills-workbook.pdf>, 2023.
- [3] Daniel McDuff Tsu-Jui Fu, William Yang Wang and Yale Song. Doc2ppt: automatic presentation slides generation from scientific documents. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(1):634–642, Jun. 2022.
- [4] Hao Kong Wenkai Zhang Jia Zheng Weixiang Zhou Hongyu Lin Yaojie Lu Xianpei Han Hao Zheng, Xinyan Guan and Le Sun. Pptagent: Generating and evaluating presentations beyond text-to-slides. *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 14413–14429, Nov. 2025.
- [5] Sirui Chen Haoyu Chen Fan Zhang Xiaojie Xu, Xinli Xu and Ying-Cong Chen. Pregenie: An agentic framework for high-quality visual presentation generation. *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 3045–3063, Nov. 2025.
- [6] Jordan Turner. Beautiful.ai, “what are the principles of visual hierarchy and how do they apply to presentations?”. <https://www.beautiful.ai/blog/what-are-the-principles-of-visual-hierarchy-and-how-do-they-apply-to-presentations>, 2023.
- [7] Lei Cui Shaohan Huang Furu Wei Yiheng Xu, Minghao Li and Ming Zhou. Layoutlm: Pre-training of text and layout for document image understanding. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1192–1200, 2020.
- [8] Lei Cui Yutong Lu Yupan Huang, Tengchao Lv and Furu Wei. Layoutlmv3: Pre-training for document ai with unified text and image masking. *Proceedings of the 30th ACM International Conference on Multimedia*, pages 4083–4091, 2022.
- [9] Moonbin Yim Jeongyeon Nam Jinyoung Park Jinyeong Yim Wonseok Hwang Sangdoon Yun Dongyoon Han Geewook Kim, Teakgyu Hong and Seunghyun Park. Ocr-free document understanding transformer. *Proceedings of the 17th European Conference on Computer Vision*, pages 498–517, 2022.
- [10] Danqing Huang Haidong Zhang Danqing Shi, Weiwei Cui and Nan Cao. Reverse-engineering information presentations: Recovering hierarchical grouping from layouts of visual elements. *Visual Intelligence*, 1:9, 2023.
- [11] Dian Yu Nan Du Izhak Shafran Karthik Narasimhan Shunyu Yao, Jeffrey Zhao and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2023.
- [12] Jianfeng Wang Kevin Lin Ehsan Azarnasab Faisal Ahmed Zicheng Liu Ce Liu Michael Zeng Zhengyuan Yang, Linjie Li and Lijuan Wang. Mm-react: Prompting chatgpt for multimodal reasoning and action. *arXiv preprint arXiv:2303.11381*, 2023.
- [13] Keshav Kumar and Ravindranath Chowdary. Slidespaw:

- An automatic slides generation system for research publications. *arXiv preprint arXiv:2411.17719*, 2024.
- [14] Yiwei Xu Siqu Sun Xin Liang, Xiang Zhang and Chenyu You. Slidegen: Collaborative multimodal agents for scientific slide generation. *arXiv preprint arXiv:2512.04529*, 2025.
 - [15] Xuhui Zhou Yi-Hao Peng Sanjay Subramanian Qinyue Tan Maarten Sap Alane Suhr Daniel Fried Graham Neubig Jiaxin Ge, Zora Zhiruo Wang and Trevor Darrell. Autopresent: Designing structured visuals from scratch. *Proceedings of the IEEE Visualization Conference*, 2025.
 - [16] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. *Proceedings of the Workshop on Text Summarization Branches Out*, pages 74–81, 2004.
 - [17] Satyanjeev Banerjee and Alon Lavie. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, 2005.
 - [18] Mingyuan Zhong Zhuohao Jerry Zhang, Ruiqi Chen and Jacob O. Wobbrock. Slideaudit: A dataset and taxonomy for automated evaluation of presentation slides. *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 1–23, 2025.
 - [19] Prakhar Gupta Skyler Hallinan-Luyu Gao Sarah Wiegrefe Uri Alon Nouha Dziri Shrimai Prabhumoye Yiming Yang Shashank Gupta Bodhisattwa Prasad Majumder Katherine Hermann Sean Welleck Amir Yazdanbakhsh Aman Madaan, Niket Tandon and Peter Clark. Self-refine: Iterative refinement with self-feedback. *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 36:46534–46594, 2023.