



NUS
National University
of Singapore

UNDERGRADUATE RESEARCH OPPORTUNITIES PROGRAMME (UROP) PROJECT REPORT

Project Title: LG-PPT-Generation

Name: Li Borui

Department of Electrical & Computer Engineering, College of Design and
Engineering, National University of Singapore

Project link: <https://github.com/borui76105/LG-PPT-Generation>

Semester 2 / AY2023-2024

Table of Contents

Abstract	3
1. Introduction	3
2. Related Works	4
3. Methods	5
Ideation Process.....	5
Content Generation	6
Background Generation.....	6
Layout Generation	6
Integration with Langchain.....	6
Function Call Technique.....	7
PPT Generation Program	7
4. Experiments	7
Implementation Details	7
Background Generation.....	8
Visuals Description Generation.....	9
PPT Generation (fixed layout)	9
Cost Analysis:.....	10
Layout Generation	11
5. Findings and Discussions	12
1. Hallucination in Generation Models:	13
2. Characteristics of GPT-Generated Content:	13
3. Design of the Generation Pipeline:	15
4. Context-Aware PPT Generation	15
6. Future Work and Conclusion	16
Future Work.....	16
Conclusion	17
References	18
Appendices	20
1. All Slides from the Demonstration PPT.....	20
2. OpenAI's Function Call & Pydantic	22

Abstract

This project proposes a modest yet innovative automated system designed to assist in the creation of presentations. The method is grounded in current AI technology, accepting user inputs to direct the generation of textual and graphical content. Utilizing foundational AI models, the system processes semantic inputs, crafts corresponding textual narratives, and conceives relevant graphical elements. A layout generation module then pragmatically assembles these elements into a structured PowerPoint presentation. This initiative aims to serve as a supportive tool for users, potentially simplifying the presentation creation process while maintaining a focus on producing content that is both meaningful and visually engaging.

1. Introduction

Presentations are a cornerstone of professional and academic communication, serving as a bridge between knowledge and audience. Yet, the act of designing a presentation can be daunting and labour-intensive. Recognizing this, the present project introduces a method underpinned by artificial intelligence¹², aiming to modestly enhance the process of presentation creation.

At the heart of our approach is a user-centric system that welcomes input instructions to be interpreted by AI. Acknowledging the capabilities and limits of current technology, we employ large language models to analyze these instructions for semantic content, and foundation AI models, like Large Language Models (LLM)³⁴ and other Generative AI⁵⁶, to generate textual and visual elements relevant to the given topic.

The project does not claim to replace the nuanced creativity of human designers but strives to provide a supplementary tool. It seeks to reduce the burden of routine

¹ Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

² LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>

³ Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*. <https://arxiv.org/abs/2005.14165>

⁴ Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. <https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>

⁵ Harshvardhan, G., Gourisaria, M., Pandey, M., & Rautaray, S. (2020). A comprehensive survey and analysis of generative models in machine learning. *Comput. Sci. Rev.*, 38, 100285. <https://doi.org/10.1016/j.cosrev.2020.100285>.

⁶ Gozalo-Brizuela, R., & Garrido-Merch'an, E. (2023). A survey of Generative AI Applications. *ArXiv*, abs/2306.02781. <https://doi.org/10.48550/arXiv.2306.02781>.

aspects of presentation design, such as drafting initial content and visuals generation, by employing a layout generation module that thoughtfully places generated content into a PowerPoint format.

2. Related Works

In the domain of automated presentation generation, the ongoing exploration of integrating artificial intelligence to streamline the creation process is both a challenging and invigorating endeavour. A recent study that forms a point of reference for our work is the 2023 paper by Sebin Thomas et al., which discusses a method for generating PowerPoint presentations from textual reports⁷. Their system, though not directly instructive for our approach, offers an idea for generating relevant text for the presentation.

Our project's methodology also considers recent advances in automated visual-textual presentation layout. Notably, "PosterLayout: A New Benchmark and Approach for Content-aware Visual-Textual Presentation Layout" (Hsu et al., 2023)⁸ involves organizing layout elements to mimic human designer processes and utilizes a CNN-LSTM-based conditional GAN⁹. This research provides a reference for creating layouts that respect content relationships on a given canvas.

In crafting our method, we also acknowledge the foundational contributions to text generation techniques, such as GPT-3^{10,11} (Brown et al., 2020), and the strides made in AI-driven graphic generation, such as Stable Diffusion¹² (Rombach et al., 2022).

⁷ Thomas, S., et al. (2023). PPT Generation from Report. IJERA, 2023

⁸ Hsu, H., He, X., Peng, Y., Kong, H., & Zhang, Q. (2023). PosterLayout: A New Benchmark and Approach for Content-Aware Visual-Textual Presentation Layout. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 6018-6026. <https://doi.org/10.1109/CVPR52729.2023.00583>.

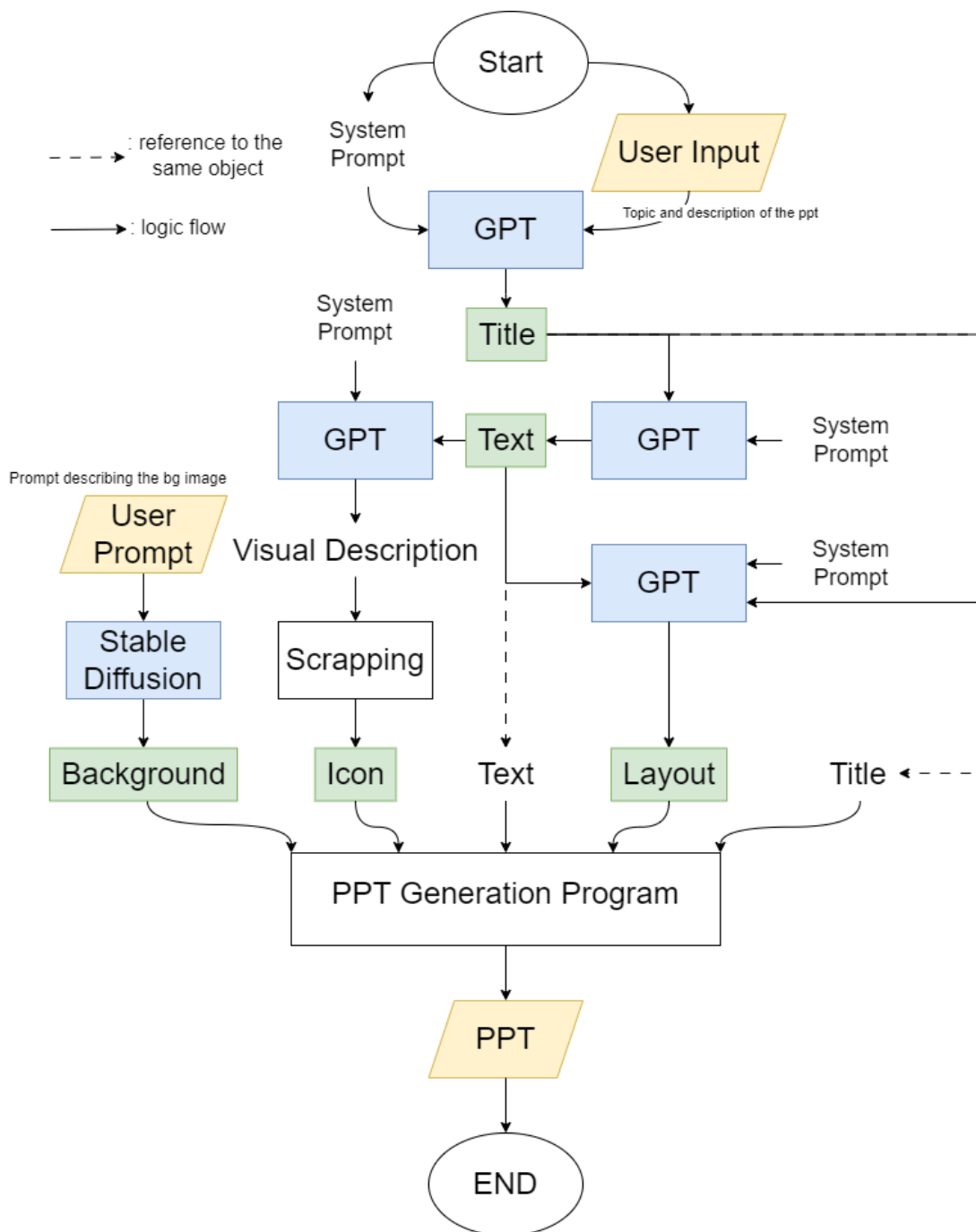
⁹ Goodfellow, I. J., et al. (2014). Generative adversarial networks. arXiv preprint arXiv:1406.2661.

¹⁰ Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165. <https://arxiv.org/abs/2005.14165>

¹¹ Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30, 681 - 694. <https://doi.org/10.1007/s11023-020-09548-1>.

¹² Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. arXiv preprint arXiv:2112.10752.

3. Methods



Ideation Process

Our automated presentation generation system begins with an ideation process that engages with user input to understand the context and content requirements for the presentation. This input includes key details such as the topic, purpose, target audience, and desired tone of the presentation. Using prompts generated by the system, the Generative Pre-trained Transformer (GPT) model processes these semantic inputs to create a conceptual framework for the content.

Content Generation

Initially, GPT is tasked with generating the title of each slide for the presentation based on the input. In the next step, the generated titles are used as input for downstream tasks. Consequently, it undertakes the generation of foundational textual elements in each slide, which is then utilized for the generation of possible visual descriptions that will guide the choice of images and icons. Finally, the descriptions of visuals are fed into a scrapping module, which sources and suggests appropriate visual elements.

Background Generation

A key component is the image (background) generation process, which employs the Stable Diffusion technology to create custom background images for the presentations. Specifically, our system utilizes SDXL-Turbo, a distilled version of the robust SDXL 1.0, designed for real-time synthesis of high-quality images. This innovative approach is grounded in Adversarial Diffusion Distillation (ADD), a technique outlined in our technical report, which enables the rapid production of images in as few as 1 to 4 steps while maintaining high fidelity.

SDXL-Turbo leverages score distillation, where large-scale foundational image diffusion models provide a teacher signal, guiding the generation process. In tandem with this, an adversarial loss component ensures that even when operating in a low-step regime, the visual quality of the generated backgrounds is not compromised. This methodology allows our system to efficiently produce visually compelling and contextually relevant backgrounds that enhance the overall aesthetic of the generated presentations, contributing to the creation of slides that are both informative and visually striking.

Layout Generation

The GPT model also delineates textual content for the slides and proposes an initial layout structure. This involves breaking down complex text into discrete, digestible pieces suitable for slide-based presentation and creating bounding boxes for text, images, and icons within the slide canvas. These components are defined by size and content, and are informed by the generated title and visual descriptions to ensure relevance and coherence.

Integration with Langchain

To facilitate the management of AI-generated content, Langchain¹³ is employed. This tool streamlines the dialogue between the user and GPT, transforming the outputs into structured data formats such as JSON or dictionaries. This is essential for translating GPT's output into actionable data that the PPT generation program can interpret and manipulate, moving beyond plain text to a structured arrangement that aligns with the proposed layout.

Function Call Technique

An integral part of our method is the function call technique, utilizing OpenAI's API to extract salient information from the input text. This information is systematically converted into structured data which can be directly used as input for subsequent functions within the system. This method significantly enhances the efficiency of generating presentation content that is both accurate and contextually relevant.

PPT Generation Program

The final stage in our methodology involves the PPT generation program, which takes the structured data from previous steps and organizes it into a PowerPoint (.pptx) file. This automated assembly is informed by the system prompts and outputs generated through GPT, ensuring that each slide is arranged with a balance of text and visuals that align with the presentation's overall narrative.

Each step of the method can be iteratively refined, drawing from continuous human feedback. The outcome is an end-to-end system that not only generates a presentation but does so with consideration for the narrative flow, aesthetic arrangement, and communicative effectiveness, culminating in the creation of a PowerPoint presentation that meets the specified user requirements.

4. Experiments

Implementation Details

User input:

“Vacation in Japan”

Notes:

This demo will focus on a travel advertisement presentation.

Sample workflow:

```
topic='vacation in japan.'  
get_ppt(name=name,topic=topic,slides=5)
```

¹³ Topsakal, O., & Akinci, T. (2023). Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast. International Conference on Applied Engineering and Natural Sciences. <https://doi.org/10.59287/icaens.1127>.

```
ppt=Prs.PPT(name)
for slide in ppt.prs.slides:
    print(slide.shapes[1].text)

ppt.add_img(slide,img='../output/bg_transparent.png',pos=[0,0],width=ppt.prs.slide_width,height
=ppt.prs.slide_height)... (code)
```

Output

```
Tokens Used: 111
    Prompt Tokens: 81
    Completion Tokens: 30
Successful Requests: 1
Total Cost (USD): $0.00018150000000000002... (costs)

Japan is a country in East Asia known for its... (contents)
```

Background Generation

Image prompt:

8k, photorealistic, beautiful, ultra high-res, beautiful scenery in Japan, mountains

Model:

SDXL-Turbo is a distilled version of SDXL 1.0, trained for real-time synthesis.

Illustration:



Notes:

This prompt is manually created by humans. However, depending on the expectation of quality and complexity of the image to be generated, such human prompting could be largely replaced by LLM, given the core idea of the presentation.

Visuals Description Generation

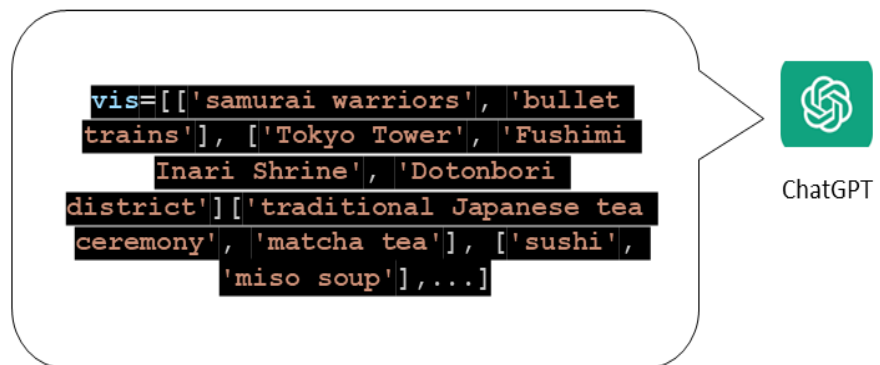
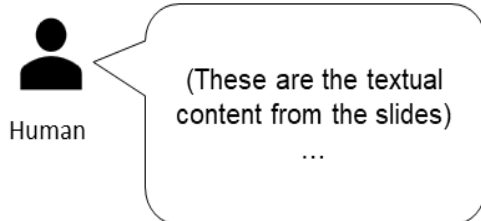
Input:

Text content for each slide.

Output:

Visual descriptions

Example:



PPT Generation (fixed layout)

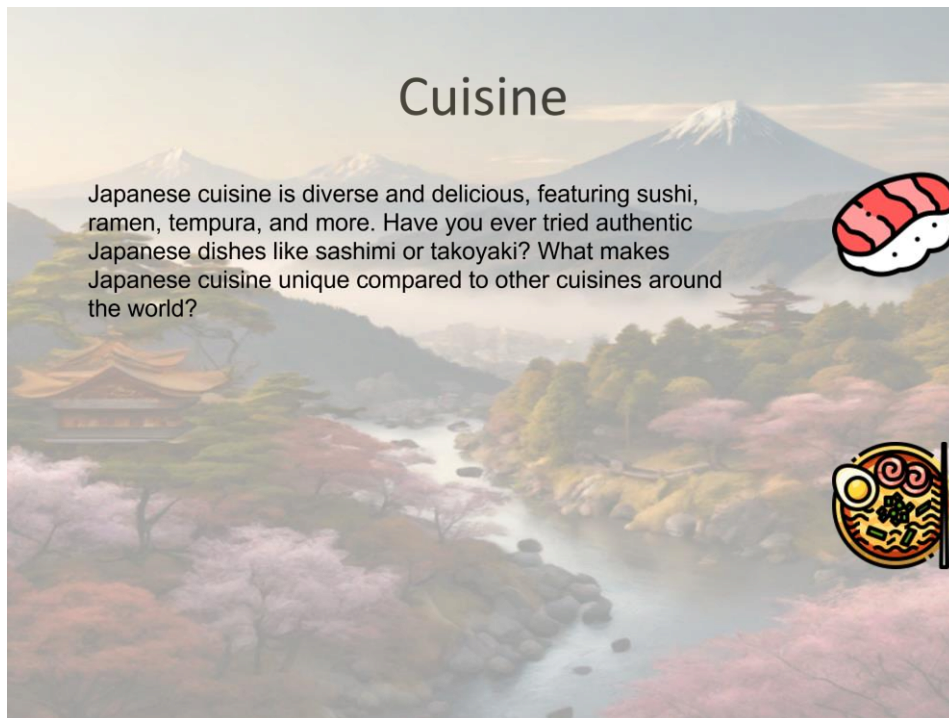
Input:

Titles, Texts, Visuals, Background Image(s)

Output:

A '.pptx' file

Illustration:



Cost Analysis:

The creation of a PowerPoint presentation consisting of five slides, each containing a similar amount of text as demonstrated in the initial example, will incur costs.


```
Tokens Used: 111
  Prompt Tokens: 81
  Completion Tokens: 30
Successful Requests: 1
Total Cost (USD): $0.00018150000000000002
Tokens Used: 1485
  Prompt Tokens: 1211
  Completion Tokens: 274
Successful Requests: 5
Total Cost (USD): $0.0023645
Tokens Used: 903
  Prompt Tokens: 849
  Completion Tokens: 54
Successful Requests: 5
Total Cost (USD): $0.00138150000000000001
```

The development process, including the sourcing of visuals and content generation as previously described, involves multiple stages of Langchain LLM interactions. Consequently, expenses will be accrued incrementally at various stages throughout the project. As such, the total cost is **\$0.0039275** for this demonstrated presentation.

Plans & Pricing

Monthly

Yearly

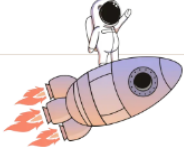


Free
\$0 /user per month

What you get:

- 400 AI credits at signup
- Unlimited users & gammas
- PDF export (Gamma branded)
- PPT export (Gamma branded)
- 30-day change history
- Basic analytics

Get started

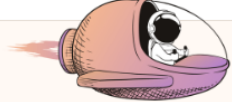


Plus
\$8 /user per month

Plan includes:

- Unlimited AI creation - up to 15 cards at a time
- Remove "Made with Gamma" badge
- PDF export
- PPT export
- Unlimited change history
- Unlimited folders

Get started



Pro
\$15 /user per month

Plan includes:

- Unlimited AI creation - up to 30 cards at a time
- Advanced AI models
- Priority support
- Remove "Made with Gamma" badge
- Custom fonts
- Unlimited change history
- Detailed analytics

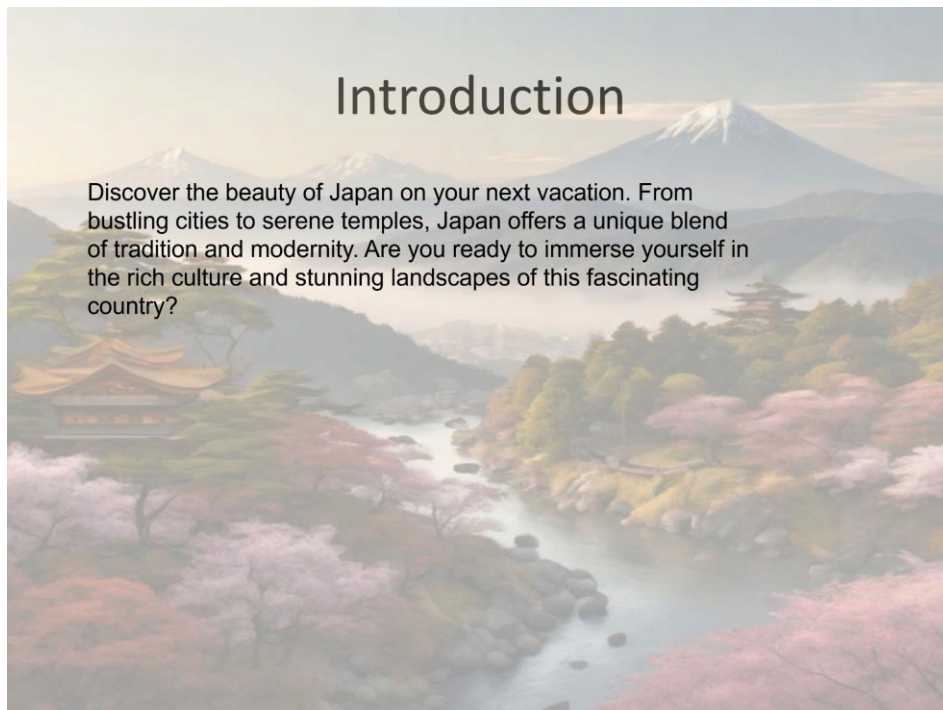
Get started

As compared to online PPT generation tools today (<https://gamma.app/pricing>, Gamma, April 2024), payments are usually in the form of subscriptions. In comparison, a pay-per-use model provides greater flexibility than subscription services, allowing the freedom to switch between products without being tied down by a prior payment.

Layout Generation

In the process of generating layouts, it is essential to not only adjust the positioning of various elements within a PowerPoint presentation but also to modify the sizes of certain elements to improve aesthetic appeal. This approach ensures a more visually engaging and effectively organized presentation.

Default Layout:



Generated Layout:



5. Findings and Discussions

In the development and testing of our automated presentation creation system, several key findings have emerged regarding the performance and reliability of the AI models, particularly the Generative Pre-trained Transformer (GPT) used in the generation of textual and graphical content. This section discusses the outcomes,

challenges, and insights gained from the application of these AI technologies in the project.

1. Hallucination in Generation Models:

Our experiments highlighted a recurring issue of 'hallucination,' where the AI, primarily GPT, generated plausible but factually incorrect or irrelevant content. This phenomenon was particularly noticeable when the model was prompted with complex or lengthy contexts. Such hallucinations not only undermine the credibility of the generated content but also risk misguiding users. Moreover, this can lead to deviations in the content that stray from the user's original direction, posing challenges in maintaining the intended narrative flow and relevancy of the presentation.

2. Characteristics of GPT-Generated Content:

The content generated by GPT models exhibited certain commonalities, such as a coherent structure and deficiency in the diversity of output.

Part of the prompt:

...the result is or has: point form, succinct text, essential examples. Keep it simple: Use layman's terms to explain complex concepts. Engage your audience: Pose questions or present hypothetical scenarios. Ensure a word limit of...

Results:

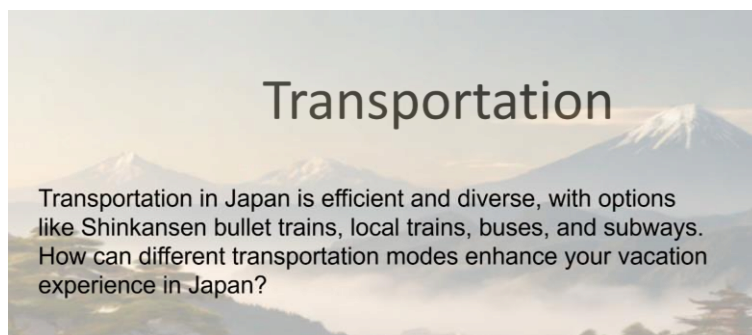


Fig 5.1

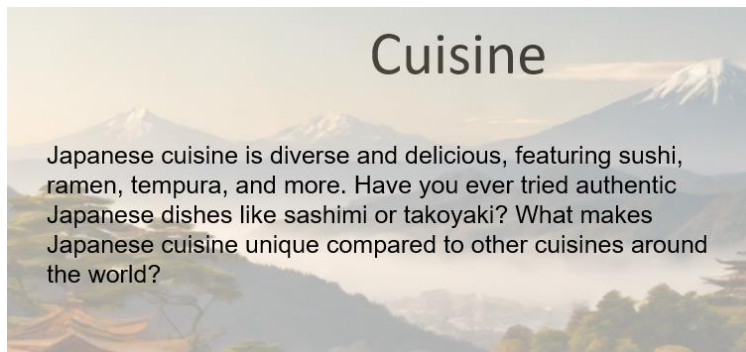


Fig 5.2



Fig 5.3

The sample slides present a discernible pattern in structure—predominantly initiating with a declarative statement followed by interrogatives—reflecting a rigid **adherence** to paragraph construction rather than **content richness**. In striving to maintain brevity, the system, exemplified by ChatGPT, occasionally sacrifices vital information.

This oversight is evident in Figure 5.1, where the content would benefit from **practical** advice on local public transportation nuances, an aspect omitted.

Similarly, in Figure 5.2, the mere cataloguing of local cuisine names falls short of being **informative** for an audience that might lack prior knowledge of the dishes.

Figure 5.3 highlights another gap; ChatGPT does not specify methods or platforms for researching attractions and accommodations, which are **details** of potential significance to the audience.

Overall, there's an observed imbalance where the system sometimes prioritizes **adherence** to instructions over the judicious evaluation of **content** significance. For instance, essential contextual information may be undervalued in comparison to the use of rhetorical devices, which can be perceived as formulaic. This suggests an area for improvement where the model needs to develop a more nuanced understanding of content prioritization against rigidly following instructions.

3. Design of the Generation Pipeline:

The design of our generation pipeline was intentional, aiming to balance **autonomy** and **user control** as we have noticed that many generation methods and technologies are immature as of date. By allowing semantic inputs and providing options for manual intervention, the system seeks to leverage the strengths of AI while mitigating its limitations. The **modular** design ensures that each component can be updated independently as better techniques become available¹⁴ or as user requirements evolve.

4. Context-Aware PPT Generation

In our approach to automated PowerPoint generation, we initially generate the title for each slide, which serves as a guide for the Large Language Model (LLM) to produce the main text content based on both the slide titles and the original user input. While this methodology emphasizes **minimizing computational costs** and **maximizing efficiency and stability**, it presents a notable **limitation** in terms of content **coherence and quality**. Specifically, because the LLM is not provided with information on the content of previous slides, there is a lack of continuity and reference among slides, potentially resulting in a disjointed presentation.

This approach aligns with practical constraints, especially in scenarios where computational resources are limited or when quick generation is prioritized. However, the quality of the generated content is crucial, particularly in professional or educational settings where the coherence of information can significantly impact the audience's understanding and engagement.

Insight and Recommendation:

To enhance the coherence and overall quality of the presentations, integrating a **context management system** into the generation process could be beneficial. Such a system would maintain a dynamic context model that includes information from all previously generated slides. By doing this, the LLM can refer back to earlier content, ensuring that each new slide is contextually aligned with the overall narrative. This could involve a more sophisticated algorithm that tracks key themes and topics discussed in the presentation (possibly in latent representations) and uses this information to guide the generation of subsequent slides.

Additionally, employing a **review or refinement phase** where the LLM revisits the entire slide deck once all individual slides have been initially generated could further

¹⁴ Cheng, Y., Chen, J., Huang, Q., Xing, Z., Xu, X., & Lu, Q. (2023). Prompt Sapper: A LLM-Empowered Production Tool for Building AI Chains. ACM Transactions on Software Engineering and Methodology. <https://doi.org/10.1145/3638247>.

improve coherence. During this phase, the model could make adjustments based on a holistic view of the presentation, enhancing transitions and thematic connections between slides.

Implementing these enhancements would likely increase computational demands but could significantly elevate the quality of the final product. Balancing efficiency with content quality is critical, and incremental improvements in contextual awareness could provide a substantial benefit without a proportional increase in resource consumption.

6. Future Work and Conclusion

Future Work

Looking forward, the landscape of automated presentation generation presents several avenues for enhancement and exploration:

- **Benchmarking for Evaluation:** As of now, there is no established benchmark for the evaluation of automated PPT generation. The introduction of evaluation models, possibly utilizing the CLIP model, could set new standards, helping to objectively assess and improve the quality of AI-generated presentations.
- **Insights from Layout Generation Research:** The findings from PosterLayout-CVPR2023 suggest training a designated layout generation module would be greatly beneficial. By integrating these insights, we can enhance the system's understanding of spatial dynamics, leading to more sophisticated and visually coherent layouts.
- **Utilization of Automatic Chain of Thought (CoT):** There is potential in employing an automatic CoT¹⁵ to facilitate a more advanced narrative structure within presentations. This could lead to a more logical and intuitive flow of information, closely mimicking human cognitive processes in organizing content.
- **Advancements in Multi-Modal Language Models:** The incorporation of multi-modal language models^{16,17} could further enhance the quality of both layout and content. Such models would allow for a more nuanced

¹⁵ Zhang, Z., Zhang, A., Li, M., & Smola, A. (2022). Automatic Chain of Thought Prompting in Large Language Models. *ArXiv*, abs/2210.03493. <https://doi.org/10.48550/arXiv.2210.03493>.

¹⁶ Jangra, A., Jatowt, A., Saha, S., & Hasanuzzaman, M. (2021). A Survey on Multi-modal Summarization. *ACM Computing Surveys*, 55, 1 - 36. <https://doi.org/10.1145/3584700>.

¹⁷ Zhang, Z., Zhang, A., Li, M., Zhao, H., Karypis, G., & Smola, A. (2023). Multimodal Chain-of-Thought Reasoning in Language Models. *ArXiv*, abs/2302.00923. <https://doi.org/10.48550/arXiv.2302.00923>.

understanding of the interplay between text and visual elements, pushing the boundaries of automated design.

- **Structured Data Integration:** There is an opportunity for future research to refine the system's capability to integrate structured data inputs. Researchers could aim to develop mechanisms whereby a presentation generation system can interpret data sets, extract salient points, and seamlessly weave this analysis into the narrative of a presentation. Advancements in this area would allow for presentations that not only display data but also offer contextual insights, making complex information more accessible and engaging.
- **Advances in Human-Agent-Computer Interaction:** The subtleties of human-agent-computer interaction¹⁸¹⁹ present another avenue for research. The next generation of systems could be designed to more intuitively understand the input and feedback from human users. In this interactive loop, the agent would take on a more proactive role, researching and developing frameworks on its own based on the user's initial insights, then refining its output through iterative feedback until the system can confidently execute the creation process. This approach would streamline the collaboration between human intelligence and computational efficiency, fostering an environment where the machine's role transitions from a passive tool to an active participant in the creative process.

Conclusion

The goal of this project was to streamline the creation of presentations through a system that harnesses the power of artificial intelligence. We developed a method that intuitively integrates user input to automate the generation of both textual and graphical content, ending with structured PowerPoint presentations. Our system successfully leverages the complexity of AI-driven content creation with the simplicity of user interfaces, yielding a tool that is as innovative as it is user-friendly. The resultant presentations bear the hallmarks of effective communication: clarity, engagement, and aesthetic appeal.

In conclusion, while our system represents a step forward in presentation automation, the field remains ripe with opportunities for innovation. Through continuous research and development, we aim to expand the capabilities of AI in presentation design, making it an indispensable tool for effective communication in both professional and educational settings.

¹⁸ Shneiderman, B., Plaisant, C., Cohen, M., et al. (2016). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (6th ed.). Pearson.

¹⁹ Norman, D. A. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.

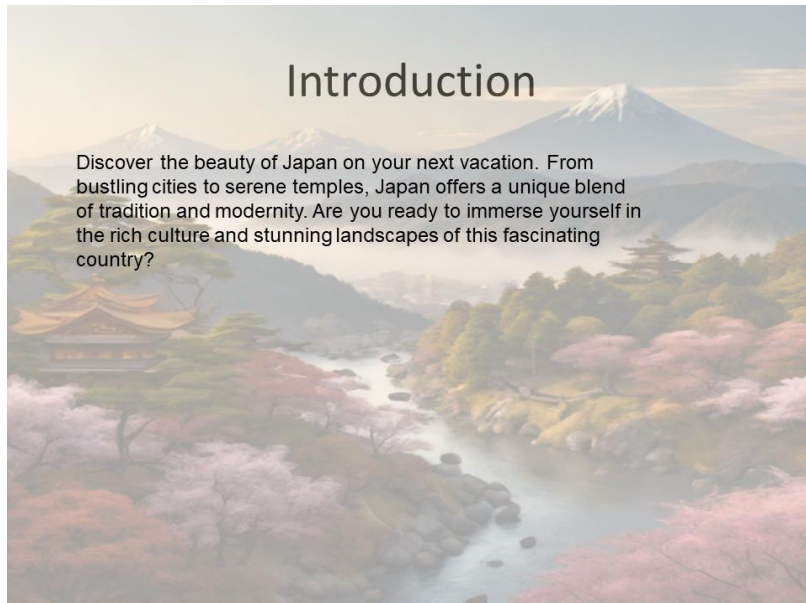
References

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
2. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
3. Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
<https://arxiv.org/abs/2005.14165>
4. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
<https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845a-a-Abstract.html>
5. Harshvardhan, G., Gourisaria, M., Pandey, M., & Rautaray, S. (2020). A comprehensive survey and analysis of generative models in machine learning. *Comput. Sci. Rev.*, 38, 100285.
<https://doi.org/10.1016/j.cosrev.2020.100285>.
6. Gozalo-Brizuela, R., & Garrido-Merch'an, E. (2023). A survey of Generative AI Applications. *ArXiv*, abs/2306.02781.
<https://doi.org/10.48550/arXiv.2306.02781>.
7. Thomas, S., et al. (2023). PPT Generation from Report. *IJERA*, 2023
8. Hsu, H., He, X., Peng, Y., Kong, H., & Zhang, Q. (2023). PosterLayout: A New Benchmark and Approach for Content-Aware Visual-Textual Presentation Layout. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 6018-6026.
<https://doi.org/10.1109/CVPR52729.2023.00583>.
9. Goodfellow, I. J., et al. (2014). Generative adversarial networks. *arXiv preprint arXiv:1406.2661*.
10. Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
<https://arxiv.org/abs/2005.14165>
11. Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30, 681 - 694.
<https://doi.org/10.1007/s11023-020-09548-1>.
12. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *arXiv preprint arXiv:2112.10752*.
13. Topsakal, O., & Akinci, T. (2023). Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast. *International Conference on Applied Engineering and Natural Sciences*.
<https://doi.org/10.59287/icaens.1127>.
14. Cheng, Y., Chen, J., Huang, Q., Xing, Z., Xu, X., & Lu, Q. (2023). Prompt Sapper: A LLM-Empowered Production Tool for Building AI Chains. *ACM Transactions on Software Engineering and Methodology*.
<https://doi.org/10.1145/3638247>.

15. Zhang, Z., Zhang, A., Li, M., & Smola, A. (2022). Automatic Chain of Thought Prompting in Large Language Models. *ArXiv*, abs/2210.03493.
<https://doi.org/10.48550/arXiv.2210.03493>.
16. Jangra, A., Jatowt, A., Saha, S., & Hasanuzzaman, M. (2021). A Survey on Multi-modal Summarization. *ACM Computing Surveys*, 55, 1 - 36.
<https://doi.org/10.1145/3584700>.
17. Zhang, Z., Zhang, A., Li, M., Zhao, H., Karypis, G., & Smola, A. (2023). Multimodal Chain-of-Thought Reasoning in Language Models. *ArXiv*, abs/2302.00923. <https://doi.org/10.48550/arXiv.2302.00923>.
18. Shneiderman, B., Plaisant, C., Cohen, M., et al. (2016). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (6th ed.). Pearson.
19. Norman, D. A. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.

Appendices

1. All Slides from the Demonstration PPT



Cultural Experiences

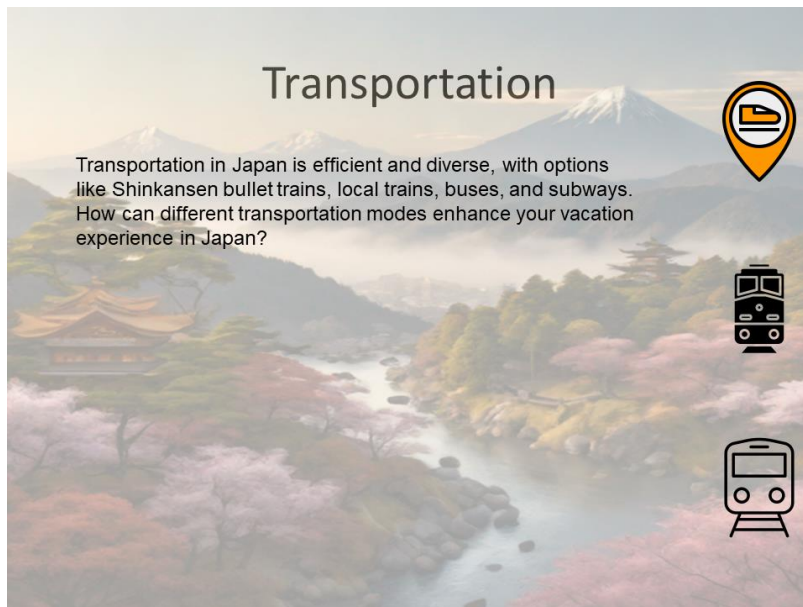
Experience traditional tea ceremonies in Kyoto with a tea master guiding you through the rituals. Learn about the history and significance of this ancient practice. How does the serene atmosphere enhance the tea-drinking experience?



Cuisine

Japanese cuisine is diverse and delicious, featuring sushi, ramen, tempura, and more. Have you ever tried authentic Japanese dishes like sashimi or takoyaki? What makes Japanese cuisine unique compared to other cuisines around the world?





2. OpenAI's Function Call & Pydantic

Work Flow:

```
# Import libraries
from pydantic import BaseModel, Field
from langchain_openai import ChatOpenAI

# Define framework
class Visual(BaseModel):
    visual: List[str]=Field(description="Visual/image for each slide is crucial and imperative. Follow the step: 1.Find the core contents of the slide 2.Pick only 1 or 2 most important parts that will need an image/icon for illustration. The output is the obj or its description, less than 3 instances, not a link")
```

```
# Function call
chat(txt,chain=create_chain(functions_config={ 'functions': Visual,'function_call': { "name":
'Visual' }},temperature=0,))
```

Details of the implementation can be found in the GitHub repository.