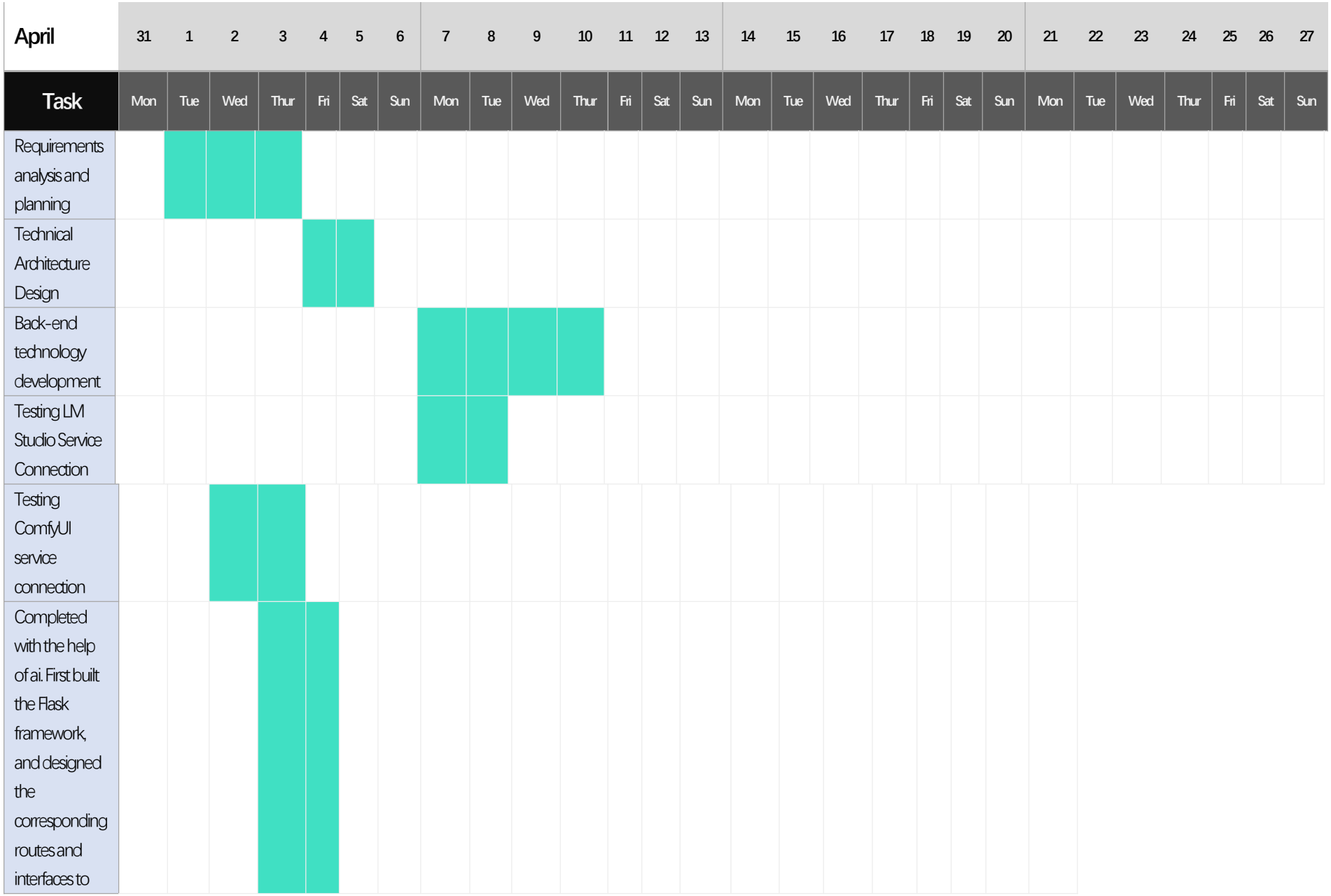
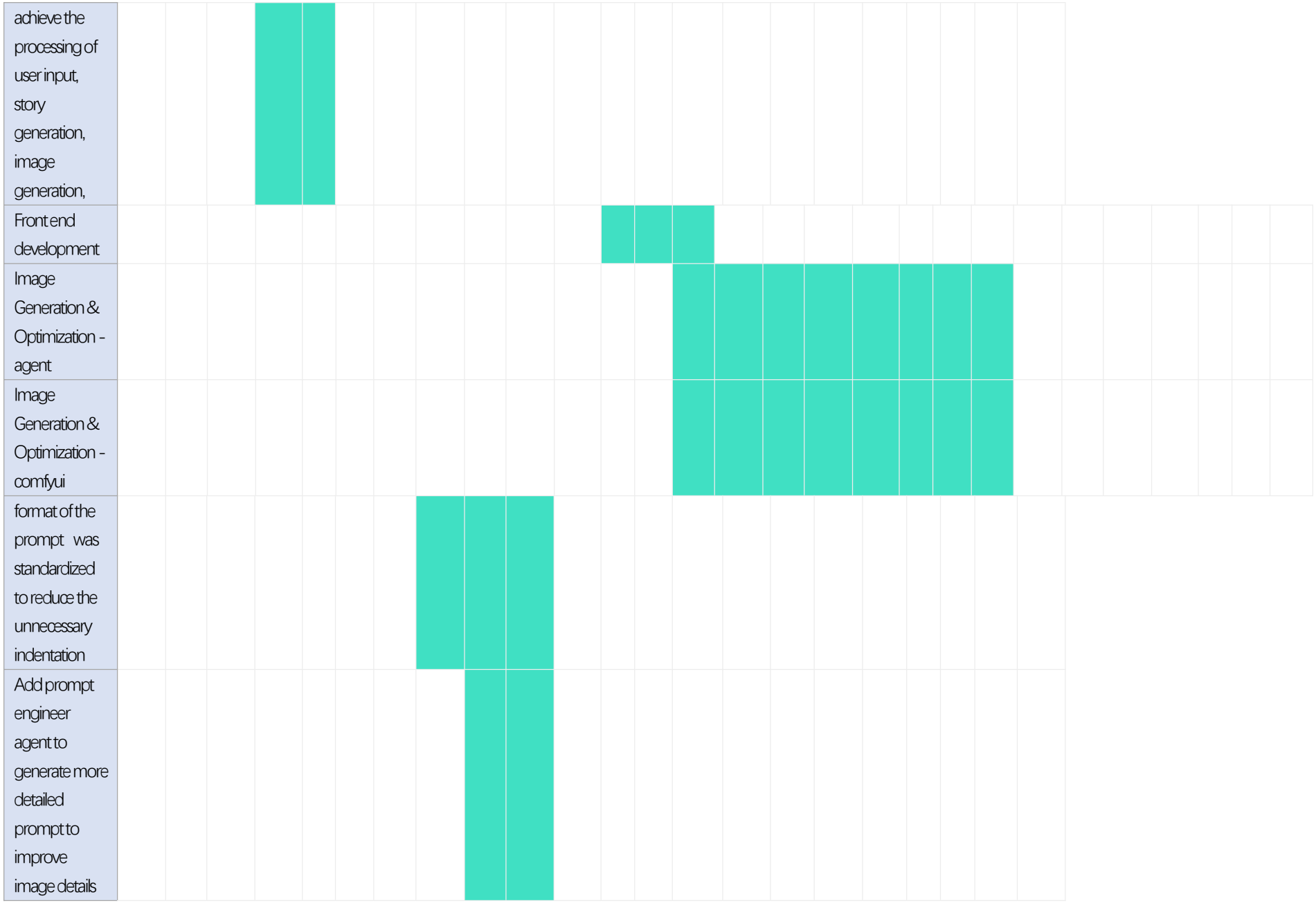
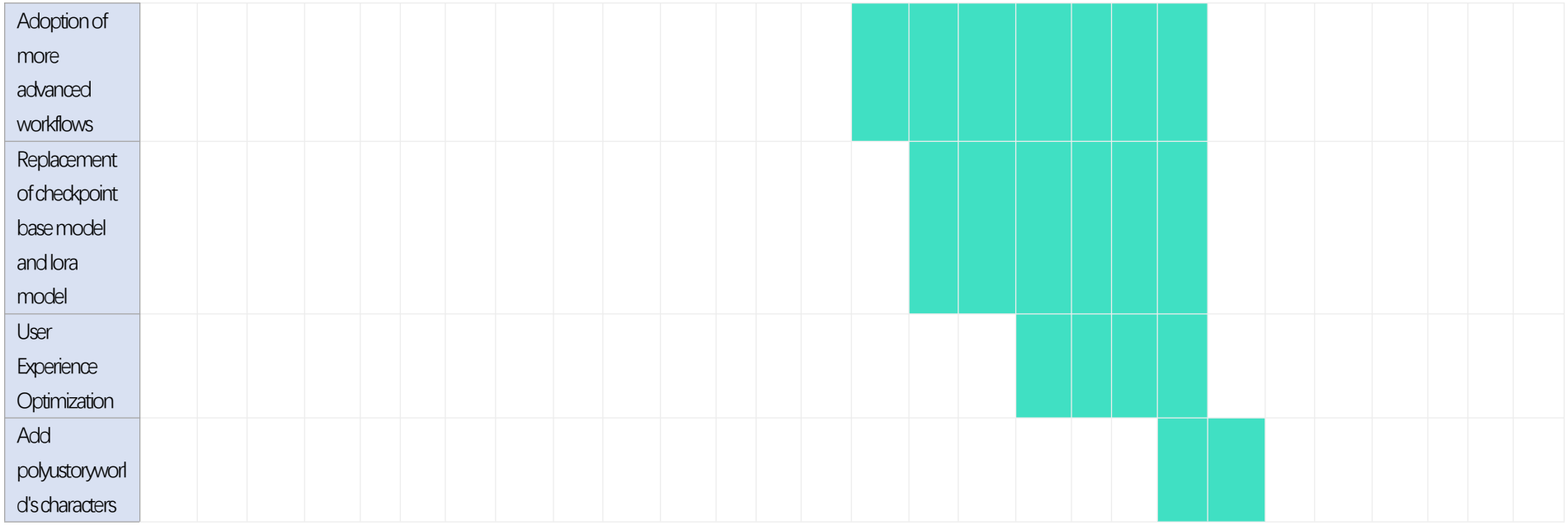

April
31 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Task
Mon Tue Wed Thur Fri Sat Sun Mon Tue Wed Thur Fri Sat Sun Mon Tue Wed Thur Fri Sat Sun Mon Tue Wed Thur Fri Sat Sun
Requirements analysis and planning
Technical Architecture Design
Back-end technology development
Testing LM Studio Service Connection
Testing ComfyUI service connection
Completed with the help of ai. First built the Flask framework, and designed the corresponding routes and interfaces to


achieve the processing of user input, story generation, image generation,
Front end development
Image Generation & Optimization - agent
Image Generation & Optimization - comfyui
format of the prompt was standardized to reduce the unnecessary indentation
Add prompt engineer agent to generate more detailed prompt to improve image details


Adoption of more advanced workflows
Replacement of checkpoint base model and lora model
User Experience Optimization
Add polyustoryworld's characters

I.Development framework

1.Requirements analysis and planning

The core function of the product: to allow the user to enter information about a character, automatically generate a picture book story around that character, and ultimately generate a PDF document.

2.Technical Architecture Design

Refer to crewai Library, crewAI-examples/write_a_book_with_flows/src/write_a_book_with_flows at main - crewAIInc/crewAI-examples
Identify the AGENT who created the main lead > character creator > story creation > art designer > web integration

3.Back-end technology development

· Completed with the help of ai. First built the Flask framework, and designed the corresponding routes and interfaces to achieve the processing of user input, story generation, image generation, PDF generation and other functions.
· Testing ComfyUI service connection (port 8188)
· Testing LM Studio Service Connection (port 1234)
· Test Flask application startup

4.Front end development

Use streamlit to display user input and output

5.Image Generation & Optimization - agent

· In response to the inconsistency of the picture book characters, it was first found that there were a lot of blank lines in the cue words passed from the character designers to the art designers, which led to the fact that ai didn't read the characteristics of the characters, so the format of the prompt was standardized to reduce the unnecessary indentation.
· Add prompt engineer agent to generate more detailed prompt to improve image details

6.Image Generation & Optimization - comfyui

· Replacement of checkpoint base model and lora model
· Adoption of more advanced workflows

7.User Experience Optimization

Since waiting for the lm studio and sd service takes a long time, add the ability to give users a clear picture of the generation process through real-time feedback and progress displays.

8. Add polyustoryworld's characters

II. Reflection and Suggestions for Improvement

During the project development process, we identified two major issues that directly affected the quality of the final product and the user experience. First, the images generated by ComfyUI lacked coherence. Although each scene's image was generated based on the same character description, inconsistencies frequently occurred in the protagonist's appearance, clothing, and posture across different scenes due to the inherent randomness of Stable Diffusion and the limitations of prompt expression. This inconsistency severely undermined the coherence and immersion of the story. To address this, we adopted a method where the character designer passed the character prompt to the art designer, and we forced the addition of the character designer's prompt to the prompt of each scene image. We also standardized the prompt generation format to prevent unnecessary indentation that could cause the AI to fail to read the complete prompt. Additionally, to improve the quality of ComfyUI image generation, we added a prompt engineer and specified guidelines for generating scene prompts to create more detailed prompts for higher-quality images. We also optimized the original simple ComfyUI workflow by incorporating a Lora model to maintain a consistent style in the generated images.

The second issue was related to user experience. Gio informed us that calling the API to generate content took a significant amount of time, and initially, we did not provide progress feedback on the user interface. This could lead to users worrying whether the program had crashed while using our product. Therefore, we should synchronize the backend debug information on the interface to allow users to clearly understand what the program is currently doing and alleviate their anxiety while waiting.

Furthermore, we can add functionality for users to customize story length, number of scenes, and other parameters, enabling them to adjust the complexity of the story according to their needs. Additionally, we can adopt a multi-stage generation and stitching strategy, first generating a story outline and then refining the content of each scene, to enhance the richness and completeness of the story. Overall, although the project has achieved a complete workflow from character information input to PDF picture book generation, there is still considerable room for improvement in image coherence and story richness. By introducing reference image mechanisms, optimizing prompts, and enriching generation parameters, we can further enhance the quality of the product and user experience. These improvements will not only solve the existing problems but also bring more possibilities and development space for the product.

III. Future Development and Improvement Directions

Based on the current functionality, we can further enhance and optimize the project from multiple dimensions to improve the overall quality and user experience of the product. First, in terms of technical architecture, we can consider introducing more advanced AI models and algorithms. For example, we can try using the latest multimodal large language models, such as GPT-4V or Gemini.

These models can better understand and generate text and image content, thereby improving the quality of the story and images. Additionally, we can consider incorporating knowledge graph technology by constructing a knowledge base related to stories, making the generated stories richer and more in-depth.

In terms of functional expansion, we can add more personalized options and interactive features. For example, we can allow users to choose different story styles and themes, such as fairy tales, science fiction, adventure, etc., to meet the diverse needs of users. In terms of interaction, we can add a real-time preview function, enabling users to adjust and modify content at any time during the generation process, thereby enhancing the user experience. Additionally, we can increase content diversity by introducing more story templates and scene types, making the generated content richer and more interesting.

Finally, in terms of user experience, we can optimize in multiple ways. For example, we can optimize the interface design to make it more intuitive and user-friendly, thereby improving the user's operational experience.