# MangaMatrix: AI Story to Comic Engine

**Abhishek Kothari [1], Vignesh Ramaswamy Balasundaram [2], Rohan Ojha [3], Pranav Raghavendra Rao [4]**
**Northeastern University**
**{kothari.abhi, ramaswamybalasunda.v, ojha.r, raghavendrarao.p}@northeastern.edu**

**Abstract**

Creating visual narratives such as manga typically requires both storytelling and artistic expertise, making the medium largely inaccessible to non-artists. Our project, *MangaMatrix*, bridges this gap by introducing a generative AI system that transforms natural language prompts into fully illustrated manga panels. The pipeline combines a fine-tuned LLaMA 2 model for structured story generation with a Mistral model for panel-level elaboration. These textual outputs are then passed to OpenAI's DALL·E 3 to produce stylistically consistent illustrations. The final outputs are compiled into manga-style PDFs, providing an end-to-end solution for accessible and automated comic creation.

## 1. Introduction

Manga creation is a timeconsuming and complex endeavor that involves several creative actions. Illustrators and authors work collaboratively to design unique characters, develop engaging plots, and visually narrate stories through detailed illustrations. However, with artificial intelligence aiding most fields, there lies a promising opportunity to automate part of the process, making manga creation more efficient and accessible.

Large Language Models (LLMs) have achieved good results in generating humanlike text, and diffusion models and GANs have revolutionized image generation. Despite all these advances, the current landscape lacks an endtoend, consistent manga generation solution that utilizes both visual synthesis and text generation. This project, titled MangaMatrix, tries to bridge that gap by enabling users to give simple inputs, such as a oneline plot and brief character descriptions, and produce a complete, customized manga comprising a coherent storyline and anime style panel illustrations. With plot generation finetuning of LLMs and stateoftheart image generation models for illustration, MangaMatrix aims to make manga creation a unique and seamless experience for users.

## 2. Background

MangaMatrix is built on top of recent innovations in large language models (LLMs), instruction tuned transformers, and generative image models. These technologies enable the automated generation of rich, structured narratives and stylistically consistent visuals from natural language inputs — a capability particularly suited for the domain of manga storytelling.

At the core of our text generation module is the LLaMA 2 7B model, which we finetuned using a curated dataset of anime subtitle dialogues. This finetuning step allowed the model to learn the style, pacing, and tone commonly found in manga and anime including dramatic sentence structure, emotional dialogue, and visual scene hints. We applied parameter efficient finetuning (LoRA) to reduce computational overhead while maintaining generation quality. The output from this model is a 5 panel manga story summary that preserves narrative coherence and builds visual progression, just like traditional manga does across pages.

To further enrich the story, each panel summary is passed to an instruction tuned Mistral 7B model. This model is used to generate 3–4 highly detailed subpanel descriptions per panel. These descriptions are crafted to include not just content (e.g., a character entering a room) but also visual composition cues, such as camera angles, posture, background atmosphere, facial expressions, and mood making them suitable for feeding into text to image systems.

On the vision side, we use OpenAI's DALLE 3 API to generate manga style black and white illustrations based on these refined prompts. The prompts are engineered to align with manga art conventions with high contrast, dynamic posing, minimalistic backgrounds, and sequential frame flow. Since DALLE 3 supports detailed prompt interpretation, it serves as a strong backbone for stylistically coherent and high resolution image generation.

The system represents a multistage, multimodal architecture: natural language is processed and enriched by text based models, then translated into visual content via a powerful image model. The final outputs are compiled into a downloadable PDF with panel wise narration overlays, providing a complete manga experience. Our models are deployed on Hugging Face Spaces to enable scalable inference, while our frontend is hosted on Render, making the pipeline accessible through a lightweight web application.

Together, these components create an end to end AI storytelling platform that not only automates the comic generation process but also reflects a deep integration of state of

the art NLP and vision models in a real world creative application.

## 3. Related Work

Recent advancements in generative AI, particularly in natural language processing (NLP) and computer vision, have laid the groundwork for applications in creative domains such as visual storytelling and manga generation. Our project builds upon and extends multiple streams of research, including large language models, text to image synthesis, and multimodal storytelling systems.

### 3.1 Large Language Models and Instruction Tuning

Our work builds upon transformer based language models like GPT 3 (Brown et al., 2020) and LLaMA 2 (Touvron et al., 2023), which have shown impressive capabilities in generating coherent and creative text. Prompt engineering is an important key to the successful fine tuning of large language models by the careful crafting of inputs to improve task specific performance, guide outputs, and enhance generalization to various applications and domains (Sahoo et al., 2024). The paper provides a comprehensive survey of prompt engineering techniques, detailing methods, tools, and challenges. The concept of instruction tuning was introduced in models like FLAN T5 (Chung et al., 2024) and later adopted by OpenAssistant and Mistral Instruct models, allowing models to better follow human written prompts. These principles guided our use of LLaMA 2 for manga story generation and Mistral 7B for elaborating individual panels into visually descriptive subcomponents.

### 3.2 Multimodal Generative Systems

Research in multimodal learning where language and vision models work in tandem has advanced significantly with projects such as CLIP (Radford et al., 2021), DALLE (Ramesh et al., 2021), and Imagen (Saharia et al., 2022). These models map language embeddings to visual outputs, enabling high fidelity image generation from natural language prompts. We specifically selected DALLE 3 due to its ability to preserve stylistic consistency and generate high resolution illustrations in a manga compatible aesthetic.

### 3.3 AI Generated Comics and Visual Storytelling

Prior work in AI generated comics includes StoryGAN (Li et al., 2019), which aimed to generate comic panels from story sequences using RNNs and GANs, and prose to panels (Sachdeva and Zisserman, 2025), which explored a novel framework that transformed comic panels into literary narratives using multimodal learning techniques. These systems often lacked detailed subpanel generation and high quality visual output, which we address through multi model chaining and prompt refinement. Other creative AI tools like NovelAI and Midjourney have explored narrative art, but they often lack panel level narrative structure or are not open source for integration.

### 3.4 Alternative Methods Considered

We evaluated using Stable Diffusion with ControlNet for greater local deployment control and sketch based image refinement. However, these methods required significantly more prompt engineering, hardware resources, and lacked out of the box stylistic consistency for manga panel generation. Furthermore, many open source image generation models struggle with maintaining character consistency and spatial layout across multiple frames, a critical requirement in manga storytelling.

We also considered autoregressive story generation using GPT 3.5 or GPT 4, but found that fine tuning a smaller, domain specific model like LLaMA 2 gave us greater control and better alignment with manga style narrative pacing.

### 3.5 Summary

Our project builds on the convergence of cutting edge LLMs and generative vision models, while addressing gaps in existing systems by integrating structure aware panel generation, stylistic control, and an accessible end to end storytelling pipeline. The combination of fine tuning, instruction following, and multimodal orchestration allowed us to create a system that not only generates coherent manga narratives but also produces professional quality visual outputs with minimal user input.

## 4. Project Description

### 4.1 Data Collection and Preprocessing

**4.1.1 Datasets Used.** For this project, we curated a diverse dataset by collecting open source transcripts of both mangas and movies from various publicly available websites. In total, we gathered around 50 detailed stories, including 30 full length manga scripts and 20 movie transcripts.

What made this dataset particularly valuable was the richness of the content and each story included dialogues tagged with character names, scene descriptions, and even contextual settings. This level of detail provided the language model with the narrative structure and context it needs to generate coherent, engaging, and visually descriptive stories.

**4.1.2 Preprocessing.** The LLaMA 2 model we fine tuned supports a maximum context window of 4096 tokens. We implemented a chunking strategy with 8000 character chunks ( 2000 tokens), overlapping by 1000 characters. Each chunk was prepended with metadata tags to indicate story ID and chunk number. This strategy helped maintain context and narrative flow across chunks.

### 4.2 Model Selection and Fine Tuning

We employed a two-model setup for textual generation. The first model, LLaMA 2 7B, was fine-tuned to generate structured five-panel manga stories from user-provided premises. The second model, Mistral v2 7B, was used to refine these outputs by elaborating each panel into multiple subpanels, enhancing narrative detail, visual cues, and emotional richness.

**4.2.1 LLaMA 2 7B.** This serves as the first-stage model in our text generation pipeline, responsible for producing structured story summaries and panel descriptions. We fine-tuned the `meta-llama/Llama-2-7b-hf` model on our curated manga-style dataset. The fine-tuned version, named `Anime-Gen-Llama-2-7B`, is a causal language model comprising 7 billion parameters.

Fine-tuning was performed using the Parameter-Efficient Fine-Tuning (PEFT) framework and the `bitsandbytes` library to support 8-bit optimization. We employed Low-Rank Adaptation (LoRA) to reduce computational overhead while preserving generation quality.

The LoRA configuration was as follows: rank $r = 8$, scaling factor $\alpha = 32$ (yielding an effective scale of $\alpha/r = 4$), and a dropout rate of 0.05. Adaptation was limited to the `q_proj` component of the attention mechanism, allowing us to improve token focus without overfitting. Other components such as `k_proj` and `v_proj` remained frozen. The training objective was set to causal language modeling (`CAUSAL_LM`), enabling proper left-to-right prediction and attention masking.

**4.2.2 Prompt Strategy.** We employed structured prompts derived from real anime story arcs. Each prompt included a short narrative followed by exactly five panel descriptions written in manga format. The model learned to mimic this structure and generate new five-panel storyboards from user-specified premises.

Table 1: LLaMA 2 7B Generation Configuration

| Parameter | Value |
|---|---|
| Temperature | 0.8 |
| Top-p | 0.95 |
| Repetition Penalty | 1.1 |
| Do Sample | True |
| EOS Token ID | None |

**4.2.3 Mistral v2 7B.** This model serves as the second stage in our generation pipeline. It takes the output of the LLaMA 2 model—high-level panel summaries—and elaborates on them to produce multiple subpanels per panel. This process enriches visual direction, emotional nuance, and cinematic structure.

The Mistral model introduces enhancements in four key areas: it enables dynamic camera angles and views, improves the pacing and continuity of visual storytelling, adds emotional depth through parallel reactions and micro-expressions, and supports split perspectives, such as simultaneous or contrasting viewpoints between characters.

**4.2.4 Prompt Template.** Each panel summary is passed to the Mistral model using structured prompt templates. These prompts instruct the model to expand each single panel description into 3–4 visually descriptive subpanels. The prompts emphasize camera positioning, character posture, facial expressions, and atmosphere, details essential for high-quality image generation in the next stage of the pipeline.

Table 2: Mistral v2 7B Generation Configuration

| Parameter | Value |
|---|---|
| Temperature | 0.85 |
| Top-p | 0.95 |
| Repetition Penalty | 1.2 |
| Do Sample | False |
| EOS Token ID | From tokenizer |

**4.2.5 Impact of Repetition Penalty.** A noteworthy configuration in our setup is the use of a repetition penalty of 1.2, which is relatively high compared to common defaults. This hyperparameter plays a crucial role in discouraging the model from reusing similar phrasing and structural patterns across subpanel descriptions. As a result, the model is encouraged to produce subpanels that feature varied camera angles, diverse narrative expressions, and more imaginative prompt structures. This enhanced diversity directly benefits the subsequent image generation stage, resulting in richer and more visually engaging manga illustrations that help sustain reader interest through stylistic variation.

This diversity directly benefits the image generation stage, resulting in richer and more engaging manga illustrations that maintain reader interest through visual variety.

**4.2.6 Pipeline Synergy.** The two-stage generation pipeline beginning with high-level story and panel generation using LLaMA 2, followed by subpanel elaboration through Mistral 7B forms a powerful and controllable system for manga comic creation. The application of LoRA for parameter-efficient fine-tuning, in conjunction with carefully calibrated generation configurations, enables the system to consistently produce narratively coherent and visually compelling outputs that align with the stylistic expectations of manga storytelling. This architecture results in coherent, structured manga-style storytelling and enables the generation of visually dynamic, emotionally expressive panel compositions that align with the tone and pacing of traditional manga.

**4.2.7 Image Generation.** For the visual rendering component of our pipeline, we primarily employed OpenAI's DALL·E 3, while also experimenting with Stable Diffusion models fine-tuned on anime datasets.

DALL·E 3 was selected as the primary image generation model due to its flexibility, high-quality outputs, and seamless integration through OpenAI's API. Its ability to interpret detailed textual prompts allowed us to preserve narrative fidelity and stylistic coherence in the generated illustrations.

Stable Diffusion was evaluated as a secondary option for potential future integration. Its open-source nature and support for advanced capabilities such as ControlNet and style transfer make it a promising candidate for offline deployment or greater customization.

To improve output quality and character consistency across panels, we applied textual inversion techniques to retain character traits and used ControlNet to enforce pose stability in generated images. The combination of these methods allowed the system to produce visually compelling

manga-style panels that align with the descriptive fidelity of the subpanel prompts.

**4.2.8 Prompt Engineering.** Prompt engineering is the process of carefully crafting input queries to guide large language models (LLMs) toward producing desired outputs. Since LLMs operate by identifying and extrapolating from learned patterns, the design of the prompt significantly influences the relevance, structure, and quality of the generated content. This becomes especially critical in generative tasks such as storytelling, where coherence, pacing, and narrative style must be explicitly modeled.

In this project, we applied few-shot prompt engineering and instruction tuning in a two-stage pipeline involving LLaMA 2 and Mistral 7B. Each model was strategically guided to fulfill a distinct role in the manga generation process.

In the first stage, the LoRA fine-tuned LLaMA 2 model was prompted to generate a complete five-panel manga story from a user-provided premise. Prompts included example story arcs, each divided into five well-structured panels that mimicked the format and tone of anime narratives. This approach enabled the model to learn consistent story pacing, character introduction, and visual progression.

Once the five high-level panels were generated, the second stage involved passing each panel summary into the Mistral 7B model. Here, task-specific prompts were used to expand each panel into 3–4 subpanels. These instruction-based prompts emphasized visual elements such as camera angles, atmosphere, character posture, and emotional cues. Mistral's output served as detailed visual blueprints for image generation, effectively bridging text and visual storytelling.

This two-stage prompt engineering approach ensured narrative consistency, thematic structure, and creative richness. LLaMA 2 functioned as the high-level story generator, while Mistral 7B served as the visual elaborator—each driven by tailored prompts designed to optimize their individual capacities. Without this controlled prompting strategy, the models would have produced fragmented or stylistically inconsistent outputs. The success of the MangaMatrix system lies in this careful orchestration of prompt design, which enables a simple narrative seed to blossom into a coherent, manga-style story with strong visual and narrative alignment.

**4.2.9 Methodology.** MangaMatrix is designed as an end-to-end AI-powered storytelling pipeline that transforms a user-defined premise into a fully illustrated manga-style comic, which is ultimately compiled and served as a downloadable PDF. The system consists of five modular stages, each contributing to the overall generative process.

**4.2.9.1 User Prompt (Frontend Input).** The process begins when a user enters a custom story premise via our web-based interface. This frontend, built using Flask and styled with a manga-inspired dark theme, is hosted on Render for seamless access across devices. Once submitted, the prompt is forwarded to the backend for processing.

**4.2.9.2 Panel Generation (LLM Stage 1).** The story prompt is routed to a fine-tuned instance of the LLaMA 2 7B model, hosted on a Hugging Face Inference Endpoint running on an A10 GPU. This model, trained specifically on manga-style narratives, generates a five-panel story summary. Each panel in the output corresponds to a distinct moment in the story arc, laying the foundation for downstream elaboration and visualization.

**4.2.9.3 Subpanel Elaboration (LLM Stage 2).** Each panel summary is then passed to the Mistral 7B model, also hosted via Hugging Face Inference API. This second-stage model elaborates on the content by producing 3–4 subpanels per panel, incorporating detailed visual cues such as camera angles, emotional expressions, and environmental context. These refined textual outputs serve as rich prompts for image synthesis.

**4.2.9.4 Image Generation (Visual Synthesis).** The subpanel descriptions are fed into OpenAI's DALL·E 3 API, which generates high-resolution illustrations in a manga-compatible aesthetic. The system supports both color and black-and-white generation modes. Each final image represents a full manga panel composed from the visual themes described in the corresponding subpanels.

**4.2.9.5 PDF Compilation & Delivery.** Once all images are generated, they are assembled into a multi-page PDF using the Python libraries PIL and FPDF. This PDF maintains the logical panel order, embeds optional narration overlays, and includes a cover page based on the first panel. The final PDF is stored in a static directory and is both downloadable and viewable directly within the user interface.

This modular, API-first architecture enables independent scaling and enhancement of each pipeline component, offering a smooth and accessible storytelling experience for end users while leveraging the latest in NLP and generative vision technologies.

# 5. Empirical Results

## 5.1 Achievements.

The MangaMatrix system successfully demonstrated its ability to generate coherent five-panel manga stories with engaging plotlines and logical narrative flow. The accompanying illustrations produced by the DALL·E 3 model were stylistically consistent, visually compelling, and reflective of typical manga aesthetics. In addition, the system maintained basic character consistency across panels, particularly in terms of visual traits such as hairstyles and attire. Perhaps most notably, the entire pipeline operates in near real-time, delivering complete manga outputs from brief user-provided premises through a single interactive web interface.

## 5.2 Evaluation Metrics.

To evaluate the quality and convergence of our text generation model, we monitored the training loss of the fine-tuned LLaMA 2 7B model across 100 steps. As expected in effective fine-tuning setups, the training loss showed a monotonically decreasing trend, indicating progressive learning without overfitting.

During the early phase of training (steps 10 to 30), we observed a rapid drop in loss, suggesting that the model quickly adapted to the stylistic and narrative properties of

the anime and manga dataset. In the middle phase (steps 40 to 70), the loss reduction continued at a slower but steady rate, implying deeper structural learning of pacing and formatting. In the final phase (steps 80 to 100), the loss curve plateaued while continuing to decline gradually, indicating model convergence and stabilization. These results confirm that the fine-tuning procedure was effective in aligning the model with the domain-specific storytelling tasks at the core of MangaMatrix.

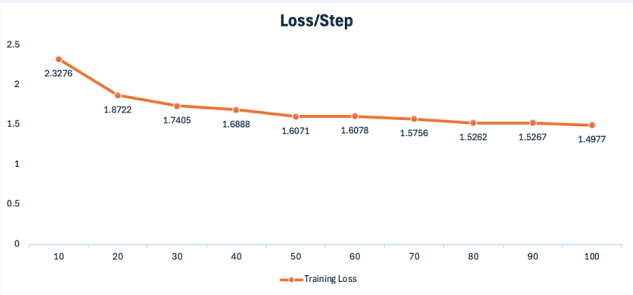


Figure 1: Training loss curve of LLaMA 2 model fine-tuning.

### 5.3 Model Output Example (Pipeline Demonstration)

To better illustrate the interplay between our two LLMs, we present an end-to-end example from a sample user prompt. The table below shows the high-level panel generated by our fine-tuned LLaMA 2 model, followed by subpanels generated by Mistral 7B.

**Premise**: A shy high school girl stumbles into a cursed book and awakens in a realm where every world ever imagined exists. Tasked with restoring balance between collapsing stories, she gains the ability to "rewrite" fate using genre rules. But each change alters her memories—and threatens to erase the real world entirely.

**Panel (from LLaMA 2)**: Before she can leave, she is knocked unconscious and wakes up in a strange land filled with monsters.

- **Subpanel 1**: Medium close-up of Emi's face, her eyes wide with wonder as she flips open the cover.
- **Subpanel 2**: Long shot of the dimly lit room, cobwebs hanging from shelves, books strewn about.
- **Subpanel 3**: Bird's eye view looking down onto Emi, who has stepped into the book.
- **Subpanel 4**: Split diagonal panels — one showing Emi's shocked expression as she lands in the new world; another, a panoramic view of the land unfolding before her.

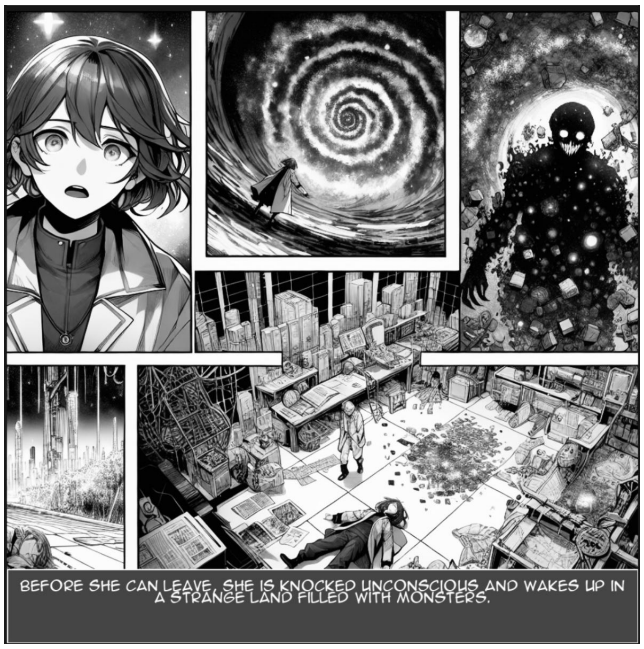


Figure 2: Panel Image

This example demonstrates how our fine-tuned LLaMA-2-7B model establishes the story framework, while Mistral-v2-7B enhances the visual narrative structure, enabling detailed, panel-ready prompts. Finally, the image was generated using OpenAI's DALL·E model, a diffusion-based neural network that synthesizes images from natural language prompts.

### 5.4 Challenges Faced

Despite the system's overall success, several challenges emerged during development and deployment. One of the key difficulties was ensuring that character voices remained distinct throughout the narrative, particularly when managing evolving emotions or interactions. Maintaining dialogue consistency required careful prompt design and sometimes post-processing to align with character profiles.

Another challenge stemmed from the training data itself. Our dataset, based heavily on subtitle transcripts, introduced noise and variability in formatting and linguistic style. This occasionally impacted the coherence of generated outputs, especially in edge cases involving nuanced emotional tone or narrative transitions.

On the systems side, inference latency proved to be a bottleneck. Because the generation pipeline involved sequential API calls to large models (LLaMA 2 followed by Mistral 7B, then DALL·E 3), real-time responsiveness was occasionally hindered, especially under limited GPU availability or high user load.

Lastly, we observed occasional hallucinations or visual inconsistencies in the image generation phase. While prompt engineering helped reduce such issues, some images produced by DALL·E 3 would diverge from character or scene continuity, particularly when fine-grained visual memory was required across panels.

# 6. Broader Implications

## 6.1 Reflection

Our project, MangaMatrix, reflects the growing potential of generative AI to democratize creativity by making traditionally skill intensive processes, like manga illustration and storytelling, accessible to anyone with a simple prompt. This project showcases how large language models and image generation technologies can collaborate in a structured pipeline to produce end to end artistic outputs, previously achievable only through expert level human effort.

## 6.2 Impact and Implications

The most immediate impact of MangaMatrix is its accessibility. By lowering the barrier to entry for storytelling, the system empowers aspiring writers, students, and hobbyists, particularly those without artistic backgrounds, to visually express narratives. Educators and content creators can also leverage the platform to rapidly prototype stories, presentations, or educational comics, saving time while engaging audiences through visuals.

Furthermore, this project highlights the potential of multimodal AI systems in creative industries. As the demand for personalized, on demand visual content grows, such pipelines could support applications ranging from comic generation and digital marketing to AI assisted filmmaking and game design.

At a technical level, MangaMatrix also demonstrates how model chaining and prompt engineering can extend the capabilities of generative models beyond standalone outputs, enabling structured, coherent, and narratively consistent results.

## 6.3 Societal and Ethical Considerations

Despite its promise, MangaMatrix raises important questions. First, the use of copyrighted anime subtitles for fine-tuning must be handled with care, particularly regarding dataset licensing and attribution. Ensuring that the source data respects creators' rights is essential for ethical deployment.

Second, as with any generative system, there is potential for misuse, including the creation of offensive, misleading, or plagiarized content. Implementing content safety filters and user guidelines is critical to prevent harm and misuse, especially as such tools become publicly available.

Finally, as AI generated content continues to blur the lines between human and machine authorship, we must consider how credit, originality, and artistic integrity are preserved. Questions about whether AI generated manga should be considered "authored," and how artists can maintain creative control are ongoing discussions within the creative AI community.

In summary, MangaMatrix demonstrates the exciting possibilities of generative AI in visual storytelling while underscoring the responsibility that comes with such innovation. As we continue to explore AI's creative potential, it is essential to balance progress with transparency, fairness, and respect for human creativity.

# 7. Conclusion and Future Work

MangaMatrix allowed us to explore the full lifecycle of a generative AI application, from dataset collection and model fine tuning to deployment and interactive user experience. Through this project, we successfully built a multimodal pipeline that transforms a simple prompt into a fully illustrated manga, using a combination of fine tuned LLaMA 2, an instruction tuned Mistral model, and OpenAI's DALL E 3. The outcome demonstrates the power of chaining large language models and image generation systems in a way that is both creative and technically robust.

## 7.1 Key Takeaways

Throughout the project, we learned how to fine tune transformer based language models using domain specific data and engineer multi step prompts to structure model output for creative generation. We also gained hands on experience in designing and deploying an end to end AI pipeline that integrates both text and image generation. Along the way, we tackled real world challenges such as inference latency, API limitations, and ensuring model consistency. This project also deepened our understanding of working with instruction tuned models, optimizing prompts for visual quality, and thinking critically about user experience when deploying generative systems.

## 7.2 If We Had More Time

Given more time and resources, we would focus on expanding the system's capabilities in several ways. One enhancement would be adding a dialogue system that automatically generates speech bubbles and overlays them into manga panels, thereby enriching the storytelling experience. We would also aim to improve character consistency using tools like ControlNet or image guided generation to maintain visual continuity across panels. Additionally, allowing users to customize the panel count, select between art styles such as shoujo or shounen, or insert custom dialogue would create a more personalized experience. Another avenue of improvement would be fine tuning an open source vision model to deploy a lightweight, offline alternative to DALL E, improving accessibility and control.

## 7.3 Advice for Future DS 5983 Students

For future students taking DS 5983, we recommend starting with a strong but flexible idea. You will learn a great deal as the project evolves, so it is important to leave room for iteration. Modularizing your pipeline early on is crucial, as it makes testing, improvement, and scaling easier without risking the stability of the overall system. Prompt engineering should not be underestimated; it is not merely about clever phrasing but about precisely guiding models to achieve desired outcomes. Finally, active collaboration is key. Each team member's strengths, whether in model building, visualization, or deployment, can deeply influence the quality and success of the final product.

In conclusion, MangaMatrix was a highly rewarding project that challenged us to combine creativity, technical skill, and design thinking. It reinforced our belief in the

power of generative AI to unlock new forms of storytelling, and it leaves us excited to explore even more ambitious multimodal applications in the future.

## 8. Project Links

- **GitHub Repository**: https://github.com/vigneshrb250/MangaMatrix-AI-Story-to-Comic-Engine
- **Dataset (Hugging Face)**: https://huggingface.co/datasets/vignesh0007/anime-data
- **Fine Tuned Model**: https://huggingface.co/vignesh0007/Anime-Gen-Llama-2-7B

## 9. Acknowledgements

We are grateful to Professor Roi for his constant guidance, insightful feedback, and encouragement throughout the project. We also thank Northeastern University for providing access to critical infrastructure such as the Discovery research cluster, which supported our fine-tuning and large-scale inference processes.

We would like to acknowledge the use of generative AI tools such as OpenAI's ChatGPT and code assistants for supporting debugging, prompt optimization, and narrative design guidance during the development of MangaMatrix. These tools were instrumental in iterating on model prompts, improving subpanel clarity, and refining system logic under time constraints.

## References

[1] Brown, T. B., et al. (2020). Language models are few-shot learners. *NeurIPS*.

[2] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). LLaMA 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

[3] Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927*.

[4] Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., ... & Wei, J. (2024). Scaling instruction-finetuned language models. *Journal of Machine Learning Research*, 25(70), 1–53.

[5] Agarwal, S., Krueger, G., Clark, J., Radford, A., Kim, J. W., & Brundage, M. (2021). Evaluating CLIP: Towards characterization of broader capabilities and downstream implications. *arXiv preprint arXiv:2108.02818*.

[6] Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., ... & Chen, M. (2021). Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*.

[7] Chen, W., Hu, H., Saharia, C., & Cohen, W. W. (2022). Re-imagen: Retrieval-augmented text-to-image generator. *arXiv preprint arXiv:2209.14491*.

[8] Li, Y., Gan, Z., Shen, Y., Liu, J., Cheng, Y., Wu, Y., ... & Gao, J. (2019). StoryGAN: A sequential conditional GAN for story visualization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6329–6338).

[9] Sachdeva, R., & Zisserman, A. (2025). From Panels to Prose: Generating Literary Narratives from Comics. *arXiv preprint arXiv:2503.23344*.

[10] Hugging Face Transformers. https://huggingface.co/transformers

[11] DALL·E 3 API. https://platform.openai.com/docs/guides/images