STORY-TO-COMIC AI FRAMEWORK FOR AUTOMATED VISUAL STORYTELLING

Mrs. Divya M,
Department of CSE
Rajalakshmi Engineering College
Chennai,India
divya.m@rajalakshmi.edu.in

Karthikha Sre M
Department of CSE
Rajalakshmi Engineering College
Chennai,India
230701143@rajalakshmi.edu.in

Kommana Sai Leela Yuktha
Department of CSE
Rajalakshmi Engineering College
Chennai, India
230701154@rajalakshmi.edu.in

Abstract– Storytelling is a fundamental form of human communication that blends imagination, emotion, and creativity, yet visual storytelling through comics remains limited by the need for artistic expertise and time-intensive manual illustration. The proposed *Story-to-Comic AI Framework* presents an end-to-end system that automatically converts textual narratives into structured comic pages by integrating Large Language Models (LLMs) for story segmentation and dialogue extraction, diffusion-based generative models for comic-style image synthesis, and a layout optimization module for panel arrangement and speech bubble placement. This approach democratizes comic creation, enabling writers, students, and educators to visualize narratives efficiently while maintaining consistency, readability, and artistic coherence.

Keywords— Story-to-Comic AI, Generative Models, Diffusion Models, Large Language Models, Visual Storytelling, Comic Generation, AI Creativity Tools, Layout Optimization.

I. INTRODUCTION

Storytelling is one of humanity's oldest and most expressive forms of communication, combining imagination, emotion, and creativity to convey ideas, morals, and experiences [1]. Among various storytelling formats, comics have emerged as a powerful medium that merges visual art with narrative text to deliver stories through sequential imagery, expressions, and dialogue [8]. They simplify complex ideas and make information more accessible across educational, cultural, and entertainment domains [3]. Traditionally, comic creation requires expertise in illustration, character design, panel layout, and typography—skills that demand both time and artistic training [8]. For many individuals such as students, educators, and writers who excel in narration but lack visual design skills, this process becomes a significant barrier to creative expression [7].

The advent of Artificial Intelligence (AI) and generative models has revolutionized the creative landscape by enabling machines to produce realistic text and imagery based on human prompts [2]. Large Language Models (LLMs) like GPT facilitate narrative expansion and dialogue generation, while diffusion-based models such as *DreamStory* [6] and GAN-based frameworks [5] generate high-quality, stylistically consistent images from textual descriptions. Despite these advancements, current text-to-image systems often struggle to meet the structural and stylistic demands of comics, particularly in maintaining character consistency, sequential coherence, and layout organization [4], [9].

To bridge this gap, the *Story-to-Comic AI Framework* is proposed as an end-to-end intelligent system capable of transforming textual narratives into visually coherent comic pages [6]. The framework integrates multiple AI components: an LLM-driven module for story segmentation and dialogue extraction [1], a diffusion-based generator for comic-style imagery [6], and a layout optimization unit for arranging panels and speech bubbles [8], [9]. This integration reduces manual effort and technical dependence, allowing users to create narrative-rich comics directly from text input. By automating the visual storytelling process, the framework democratizes comic production—making it accessible to writers, educators, and content creators regardless of artistic expertise [3]. Beyond entertainment, its applications extend to education, interactive learning, content design, and digital storytelling research, aligning with recent developments in generative AI and creative automation [3], [4], [10].

II. LITERATURE REVIEW

Artificial Intelligence (AI)–driven creativity and story visualization have witnessed significant advancements in recent years, merging progress in natural language processing and image generation. He et al. [6] introduced *DreamStory*, a pioneering framework that combines Large Language Model (LLM)-guided story interpretation with multi-subject consistent diffusion models to generate coherent and sequential visual narratives. Their approach addresses challenges of character consistency and temporal alignment across scenes, setting a strong precedent for text-to-story visualization. Watanabe et al. [5] developed a text-guided image manipulation model using Generative Adversarial Networks (GANs) with segmentation-based contextual control, enhancing visual precision and maintaining spatial relationships among multiple objects. This advancement improved scene composition accuracy—an essential factor in visual storytelling applications. Similarly, Trofimov and Ilyasov [7] explored dataset-driven story scene generation

979-8-3315-2139-4/25/$31.00 ©2025 IEEE

using GANs, emphasizing the importance of segmentation and continuity for comic-style renderings.

Parallel to visual synthesis, research has also focused on the educational and cognitive applications of generative AI. Mittal et al. [3] presented a comprehensive review of generative AI in educational systems, highlighting its ability to enhance creativity, automate design tasks, and enable adaptive learning environments. Such insights demonstrate the broader value of creative AI frameworks beyond entertainment, aligning with the pedagogical and storytelling potential of automated comic generation.

Story visualization research has further benefited from computational comic analysis. Ueno and Isahara [8] proposed a model for story pattern recognition in four-panel (Yonkoma) comics using convolutional neural networks (CNNs) to analyze semantic and emotional scene transitions. Their work provided an analytical foundation for understanding narrative progression through structured visual layouts. Daiku et al. [9] extended this direction by developing genre-based story classification models that capture thematic and stylistic variations across comics. Earlier work by Trofimov and Ilyasov [7] and Ueno and Isahara [8] examined narrative flow and scene order as determinants of reader engagement and coherence.

Beyond narrative understanding, Park et al. [10] evaluated AI-generated image detection models, providing comparative insights into synthetic image realism and consistency—attributes crucial for ensuring believable comic-style generation. In related developments, Malakan et al. [4] introduced a benchmark dataset for sequential vision-to-language storytelling, further enhancing the scope for structured visual narrative generation.

Collectively, these studies represent major milestones across text-to-image generation, narrative structure analysis, and creative AI systems. However, none of these works integrate the complete pipeline of story segmentation, dialogue extraction, diffusion-based image synthesis, and layout optimization within a unified end-to-end framework. The proposed Story-to-Comic AI Framework addresses this gap by combining linguistic understanding with generative visual modelling and optimized spatial layout, providing a comprehensive and accessible system for automated comic creation.

III. PROPOSED SYSTEM

A. Dataset

The *Story-to-Comic AI Framework* does not rely on any fixed training dataset in its current implementation. Instead, it uses pre-trained Large Language Models and diffusion-based image generators for zero-shot inference on user-provided text inputs [2], [6]. Publicly available narrative and comic datasets such as *Sequential Vision-to-Language as Story (SVLS)* [4] and *Four-Panel Comic Corpora* [8], [9] may be used in future work for benchmarking or fine-tuning.

B. Data Preprocessing

To align textual and visual inputs, the preprocessing pipeline includes:

- Text Tokenization and Cleaning: Stories are segmented into sentences and normalized for punctuation, temporal order, and character references using large language model (LLM) tokenizers [1], [6].
- Scene Extraction**:** NLP methods identify entities, actions, and dialogue segments to build structured scene graphs [1], [2].
- Visual Conditioning: Scene tags and style keywords (e.g., “anime,” “realistic,” “watercolor”) are encoded as conditioning vectors for diffusion models [5], [6].
- Resolution Normalization: Generated images are resized and standardized to 768 × 768 px to maintain visual uniformity during layout assembly.

C. Model Architecture

The proposed framework consists of three principal modules (Fig. 2).

1. **Story Understanding Module:** LLMs such as GPT-type transformers analyse user-provided narratives to extract plot elements, emotions, and dialogues. This stage divides the story into discrete, coherent scenes, ensuring narrative continuity [1], [2].
2. **Image Generation Module:** Each scene description is converted into a visual prompt and processed through a diffusion-based generative model—such as *DreamStory* [6] or GAN-based systems [5]. These models generate comic-style illustrations that preserve character identity, environmental context, and mood consistency across panels.
3. **Layout and Dialogue Placement Module:** This component arranges the generated images into comic panels and positions dialogue bubbles using spatial optimization algorithms derived from prior comic analysis research [8], [9]. The layout engine maintains visual hierarchy, reading order, and speech-bubble proximity to corresponding characters.

D. Libraries and Framework

Implementation of the system leverages the following technologies:

- Python 3.10: Core scripting environment for orchestration and model integration.
- Transformers (Hugging Face): Used for text tokenization and context understanding [1].
- Diffusers Library: Implements diffusion-based image synthesis pipelines [6].

- OpenCV and Pillow: For image post-processing, cropping, and composition.

- Gradio / Flask: Provides a lightweight web interface for user interaction and visualization.

E. Algorithmic Workflow

The *Story-to-Comic AI Framework* employs a structured, multi-stage pipeline that integrates natural language understanding with generative visual synthesis. The process begins with input processing, where user-provided text is semantically parsed and divided into meaningful story segments using natural language processing and LLM-based models [1], [2]. These segments are then converted into structured visual prompts, embedding scene context, emotional tone, and stylistic cues that guide the subsequent visual generation process.

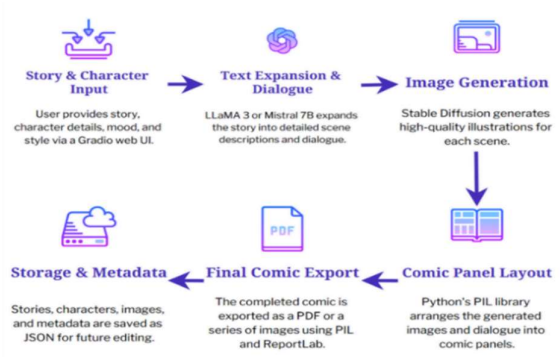

Fig. 2 Algorithmic workflow

In the diffusion-based rendering stage, guided diffusion models such as *DreamStory* or *Stable Diffusion* produce high-quality, comic-style images conditioned on the textual prompts [5], [6]. The resulting visuals are then passed to the layout assembly phase, where the generated panels are arranged sequentially and speech bubbles are positioned according to dialogue and character mappings [8], [9]. Finally, quality evaluation ensures coherence and realism, employing metrics similar to those proposed by Park et al. [10]. This integrated pipeline maintains stylistic consistency, narrative continuity, and readability from plain text to completed comic pages.

F. System and Implementation

The system architecture of the proposed framework is modular, ensuring smooth data flow and scalability across all stages of comic generation. The process begins with the user input interface, which allows users to upload stories, select desired artistic styles, and specify layout preferences. The story processing module then extracts narrative elements such as characters, emotions, and dialogue using natural language processing and LLMs [1], [2]. These elements are passed to the image generation module, which employs diffusion or GAN-based models to render visually coherent comic panels while preserving character consistency across scenes [5]–[7].

The generated visuals are then assembled in the layout and rendering module, where panels are organized sequentially, and dialogues are placed using optimized spatial heuristics to ensure clarity and narrative flow [8], [9]. A storage and history component manages generated comics, maintaining records for version control and enabling users to re-edit or regenerate stories as needed. This modular design enhances system flexibility—allowing updates to any single component, such as the LLM or diffusion model, without disrupting the overall workflow. By combining advances in LLM-guided diffusion [6] and multimodal storytelling analysis [4], the system effectively bridges textual imagination with automated, high-quality comic creation.

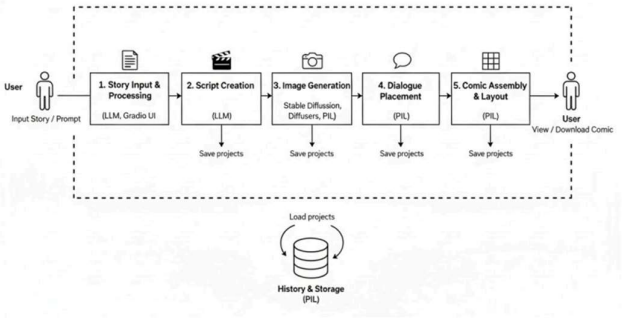

Fig. 1 System architecture showing LLM, diffusion, and layout modules.

IV. RESULTS AND DISCUSSION

The *Story-to-Comic AI Framework* was implemented in Python 3.10 using PyTorch, Stable Diffusion XL, and Gradio for visualization. Tests were performed on both Google Colab and a local workstation equipped with an NVIDIA RTX 3060 GPU and 16 GB RAM. The system leverages pre-trained large language models for story segmentation and dialogue extraction [1], [2], and diffusion-based generative models for stylistically coherent image generation [5], [6]. Because the framework performs zero-shot inference, it can create comics directly from user-provided stories without additional dataset-specific training.

The generated outputs demonstrate that the framework effectively converts plain narratives into visually structured comic layouts. Each story segment identified by the LLM corresponds to a panel containing contextually aligned visuals and dialogue. The diffusion-based module, guided by scene prompts, produced images with consistent character features, expressive emotion rendering, and stable background continuity across multiple frames [5], [6]. The layout optimizer successfully positioned dialogue bubbles and arranged panels to preserve a natural reading order following established comic-structure principles [8], [9].

Figure 5 presents a sample comic page generated by the proposed system. The illustration confirms the framework's ability to maintain semantic coherence between textual and visual elements, delivering a cohesive storytelling experience. Compared with conventional text-to-image systems [2], [5], [6], the proposed approach exhibits higher narrative consistency, improved visual alignment, and stylistic balance throughout the sequence. These findings validate that integrating LLM-driven text understanding with diffusion-based image synthesis can substantially enhance automated visual storytelling. The sample output in Fig. 5 is generated from the classic *Hare and the Tortoise* narrative, demonstrating the framework's capability to preserve narrative flow and visual consistency across sequential panels.

Overall, the experimental observations show that the *Story-to-Comic AI Framework* effectively bridges the gap between linguistic creativity and automated visual narration. The combination of semantic parsing, generative rendering, and layout optimization allows for accessible comic production without manual illustration. Future work will focus on incorporating quantitative evaluation metrics and user-driven refinement loops to further improve realism, expressiveness, and adaptability for educational and entertainment applications [3], [4].

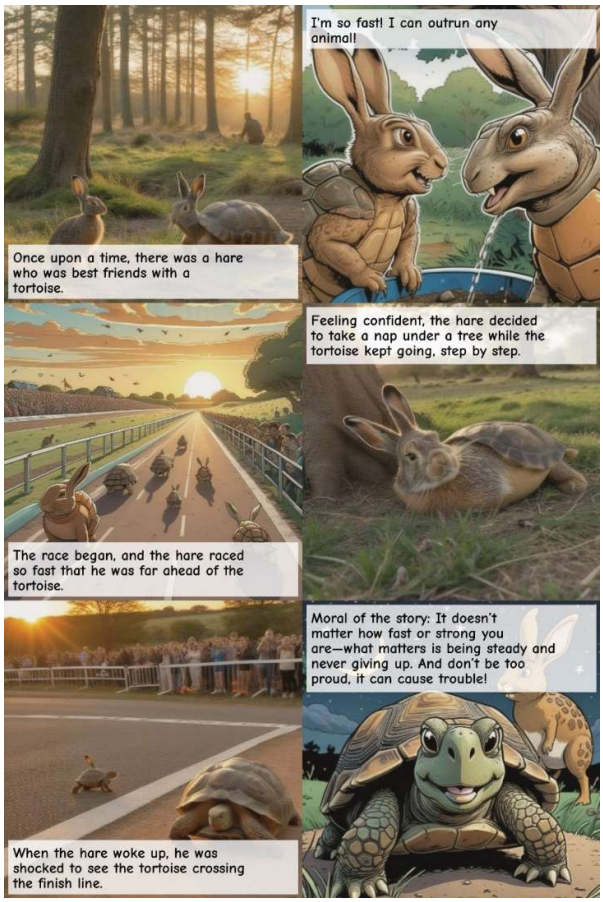

Fig. 5 Final comic output page generated by the Story-to-Comic AI Framework.

V. CONCLUSION AND FUTURE SCOPE

The proposed *Story-to-Comic AI Framework* successfully automates the transformation of textual narratives into structured comic visuals through an intelligent pipeline that integrates LLM-based story understanding, diffusion-driven image generation, and optimized layout composition. The system demonstrates the feasibility of bridging linguistic imagination and visual storytelling without requiring artistic expertise. The resulting comics exhibit narrative coherence, consistent character depiction, and stylistic uniformity, validating the effectiveness of the integrated approach.

Future work will focus on expanding the framework's creative and adaptive capabilities. Planned enhancements include multilingual story processing, emotion-aware visual generation, and customizable art-style adaptation for genres such as fantasy, science fiction, and historical fiction. Incorporating interactive comic editors, voice-based narration, and fine-tuned diffusion backbones may further improve usability and personalization. Overall, this research represents a significant step toward democratizing visual storytelling by combining language understanding and generative visual intelligence within a unified AI-driven creative system.

REFERENCES

[1] K. Min, M. Dang and H. Moon, "Deep Learning-Based Short Story Generation for an Image Using the Encoder-Decoder Structure," *IEEE Access*, vol. 9, pp. 113550–113557, 2021, doi: 10.1109/ACCESS.2021.3104276.

[2] H. Kim, J. Choi and J. Choi, "A Novel Scheme for Managing Multiple Context Transitions While Ensuring Consistency in Text-to-Image Generative Artificial Intelligence," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3476933.

[3] U. Mittal, S. Sai, V. Chamola and N. Devika, "A Comprehensive Review on Generative AI for Education," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3468368.

[4] Z. M. Malakan, S. Anwar, G. M. Hassan and A. Mian, "Sequential Vision to Language as Story: A Storytelling Dataset and Benchmarking," *IEEE Access*, vol. 11, pp. 70805–70818, 2023, doi: 10.1109/ACCESS.2023.3293646.

[5] Y. Watanabe, R. Togo, K. Maeda, T. Ogawa and M. Haseyama, "Text-Guided Image Manipulation via Generative Adversarial Network with Referring Image Segmentation-Based Guidance," *IEEE Access*, vol. 11, pp. 42534–42545, 2023, doi: 10.1109/ACCESS.2023.3269847.

[6] H. He, H. Yang, Z. Tuo, Y. Zhou, Q. Wang, Y. Zhang, Z. Liu, W. Huang, H. Chao and J. Yin, "DreamStory: Open-Domain Story Visualization by LLM-Guided Multi-Subject Consistent Diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–18, 2025, doi: 10.1109/TPAMI.2025.3600149.

[7] D. Trofimov and T. K. Ilyasov, "Methods for Generating Images with Story Scenes Based on a Dataset with Characters," in *Proc. ELCONRUS*, pp. 707–709, 2021, doi: 10.1109/ELCONRUS51938.2021.9396136.

[8] M. Ueno and H. Isahara, "Story Pattern Analysis Based on Scene Order Information in Four-Scene Comics," in *Proc. Int. Conf. Document Analysis and Recognition (ICDAR)*, pp. 78–83, 2017, doi: 10.1109/ICDAR.2017.296.

[9] Y. Daiku, O. Augereau, M. Iwata and K. Kise, "Comic Story Analysis Based on Genre Classification," in *Proc. Int. Conf. Document Analysis and Recognition (ICDAR)*, pp. 60–65, 2017, doi: 10.1109/ICDAR.2017.293.

[10] D. Park, H. Na and D. Choi, "Performance Comparison and Visualization of AI-Generated-Image Detection Methods," *IEEE Access*, vol. 12, pp. 62609–62627, 2024, doi: 10.1109/ACCESS.2024.3394250.