INTERNATIONAL JOURNAL OF
RECENT TRENDS IN MULTIDISCIPLINARY RESEARCH
https://www.doi.org/10.59256/ijrtmr.20250505012
September-October 2025, Vol 5(05), 67-72.
ISSN No: 2583-0368

AI Comic Strip Generator Using Natural Language Processing and Image Synthesis

Mohammed Mahmood Ahmed[1], Dr. Abdul Rahman[2]
[1] *Student, MCA, Deccan College of Engineering and Technology, Hyderabed, Telangana, India.*
[2]*Assistant professor, MCA, Deccan College of Engineering and Technology, Hyderabad, Telangana, India.*

OPEN ACCESS
Article Citation:
Mohammed Mahmood Ahmed[1], Dr. Abdul Rahman[2], "AI Comic Strip Generator Using Natural Language Processing and Image Synthesis", International Journal of Recent Trends in Multidisciplinary Research, September-October 2025, Vol 5(05), 67-72.

Abstract: The rapid advancements in Artificial Intelligence (AI) have created new opportunities in the domain of creative content generation, particularly in storytelling and visual media. Comics, being one of the most popular narrative formats, traditionally require both artistic expertise for illustrations and strong scripting skills for dialogue creation. This project introduces an AI-powered comic strip generator that automates the entire process, transforming simple user-provided text prompts into structured, illustrated comic panels. By leveraging Natural Language Processing (NLP) for dialogue generation and AI-based image synthesis models such as DALL·E and Stable Diffusion for artwork, the system eliminates the need for manual intervention in either scripting or design. The generated dialogues are automatically embedded into speech bubbles, and the visuals are organized into cohesive panel layouts using image processing techniques. The system offers an end-to-end pipeline that enables seamless integration of text understanding, visual generation, and design formatting. It democratizes comic creation, making it accessible to users without technical or artistic skills, thereby fostering creativity across diverse user groups. Beyond personal storytelling, the project has potential applications in education, gamification, advertising, and entertainment, where interactive and customized visual narratives can enhance engagement. The architecture is scalable and supports future enhancements, including multilingual dialogue generation, genre-specific styling, and even animated comic strips with speech synthesis. By streamlining production, reducing costs, and offering personalization, the AI Comic Strip Generator represents a transformative step in digital storytelling. Ultimately, this project demonstrates how AI can bridge the gap between imagination and expression, empowering individuals and organizations to create compelling visual stories with minimal effort.

Keywords: AI, comic strip generator, natural language processing, image synthesis, DALL·E, Stable Diffusion, dialogue generation, visual storytelling, automation, creative content, education, entertainment, interactive narratives, digital storytelling, AI-powered content, comic creation.

1. Introduction

Storytelling has long been a cornerstone of human culture, enabling individuals to share experiences, teach lessons, and entertain through various mediums. In contemporary times, comics have become a unique form of storytelling that combines both visual art and narrative text to communicate ideas. Traditionally, comic creation has been a labor-intensive process that requires a high level of artistic skill for illustrations and linguistic ability for scriptwriting. This process limits accessibility to

those with professional training in these fields, thereby excluding a significant portion of the population that may have creative ideas but lacks the resources or expertise to bring them to life.

The rapid advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies in recent years have transformed how we can generate and interact with creative content. Models like GPT-3 for dialogue generation and DALL·E for image creation have demonstrated exceptional capabilities in understanding human language and generating coherent text, as well as producing highly realistic images from textual descriptions. These advancements have opened up new possibilities in automating creative tasks such as comic strip generation, enabling individuals without technical or artistic expertise to create compelling visual stories.

This project aims to bridge the gap between manual comic creation and automated generation by introducing an AI-powered comic strip generator. The system utilizes NLP to convert simple user-provided text prompts into structured dialogues and employs AI-based image synthesis models, such as DALL·E and Stable Diffusion, to generate corresponding visual content. The resulting comic strips are automatically assembled with dialogues in speech bubbles, creating a seamless, fully illustrated narrative. This innovative approach not only saves time but also democratizes comic creation, making it accessible to a wide audience and fostering creativity among non-artistic individuals.

The proposed system goes beyond simple text-to-image translation; it integrates several complex processes into an automated pipeline. From understanding the text prompt and generating meaningful dialogue to synthesizing contextually accurate images and arranging them into comic panels, the system performs all tasks autonomously. This end-to-end process reduces the traditional reliance on skilled writers and artists, lowering the barrier for those wishing to tell their own stories visually. Furthermore, the system allows users to customize elements such as art style and genre, providing a personalized experience that caters to a variety of user preferences.

The potential applications for this AI-powered comic strip generator are vast. In addition to personal storytelling, the system can be employed in education, gamification, advertising, and entertainment, where engaging and customized visual narratives can enhance user interaction and communication. By offering a scalable and flexible framework, the system holds promise for future developments such as multilingual dialogue generation, genre-specific stylistic options, and even the integration of animation for dynamic storytelling. This project not only exemplifies how AI can transform creative fields but also serves as a step toward making storytelling more inclusive, interactive, and accessible to all.

2. Material And Methods

A. Data Collection

The AI Comic Strip Generator utilizes diverse data sources to generate visually compelling comics from text-based prompts. The primary data consists of textual input from users, such as short stories or narrative descriptions, which serve as the foundation for dialogue and scene generation. In addition to user input, pre-trained models and publicly available datasets such as those containing visual references for various objects, settings, and scenarios are used to enhance the realism and contextual accuracy of the generated images. The datasets include a wide array of labeled images, including objects like people, animals, buildings, and everyday items, which are used for training the image synthesis models. Integration with natural language processing (NLP) models like GPT-3 ensures the generation of coherent dialogue, while image synthesis models like DALL·E and Stable Diffusion convert text prompts into visual representations.

B. Data Preprocessing

To ensure high-quality input data for the comic generation process, several preprocessing steps are applied to raw user-provided text and image data:

- **Text Preprocessing**: Raw textual input is cleaned by removing noise, such as irrelevant characters and unnecessary symbols. Text is then tokenized, normalized, and transformed into a format that can be efficiently processed by NLP models.
- **Image Preprocessing**: For image synthesis, generated images are resized, normalized, and adjusted for consistent style and resolution. The images are processed using OpenCV and Pillow to prepare them for use in the comic strip.
- **Feature Extraction**: Key elements from the text, such as character names, actions, and settings, are extracted to ensure the generated dialogues and images align with the intended narrative. Additionally, visual features like scene composition, character positioning, and background elements are extracted from the generated images.
- **Data Augmentation**: To improve model robustness, augmentation techniques like rotating, scaling, and flipping are applied to the generated comic images. This ensures the model performs well across different visual scenarios and lighting conditions.
- **Data Partitioning**: The data is partitioned into training, validation, and test sets to ensure that the model generalizes well and performs effectively on unseen data.

C. Feature Engineering

Effective feature engineering is crucial for improving the model's ability to generate coherent and visually appealing comics. The following methods are used:

- **Textual Feature Extraction**: Text-based features, such as character names, actions, emotions, and settings, are extracted using NLP techniques like named entity recognition (NER) and part-of-speech tagging. These features help in generating contextually appropriate dialogue and defining the storyline for each comic panel.
- **Visual Feature Extraction**: From the generated images, features such as object size, relative positioning, and color schemes are extracted to ensure consistency across comic panels. These visual features are critical for maintaining the flow and readability of the comic strip.

- **Feature Selection**: Techniques like Recursive Feature Elimination (RFE) and correlation analysis are used to identify and select the most important features from both text and visual data. This helps the model focus on the most relevant attributes, improving efficiency and accuracy.

D. Model Development

The AI Comic Strip Generator integrates various machine learning and deep learning models for real-time comic generation:

- **Classical Machine Learning Models**: Initially, simpler machine learning models, such as Logistic Regression and Random Forest, are employed for basic object classification tasks based on the extracted features.
- **Deep Learning Models**: Convolutional Neural Networks (CNNs) are used to process and recognize objects within the visual data, while advanced transformer models like GPT are used for generating coherent dialogues and storylines. These deep learning models play a crucial role in ensuring high-quality and contextually relevant outputs.
- **Ensemble Learning**: Ensemble methods, like XGBoost, are utilized to combine the predictions from various models to enhance accuracy and reduce bias. This improves the overall performance of the comic generation system.
- **Hyperparameter Tuning**: Grid Search and Random Search techniques are employed to optimize the hyperparameters of the models, ensuring the best possible performance for both text and image generation tasks.
- **Cross-Validation**: K-fold cross-validation is used to evaluate the performance of the models, ensuring they generalize well to unseen data and avoid overfitting.

E. Implementation Environment

The implementation environment is designed to provide high-performance computation and ease of deployment:

- **Programming Language**: Python 3.x is used due to its extensive libraries and frameworks for machine learning (e.g., TensorFlow, Keras), natural language processing (e.g., Hugging Face Transformers), and computer vision (e.g., OpenCV).
- **Deep Learning Frameworks**: TensorFlow and Keras are employed for developing the deep learning models, particularly for NLP and image generation.
- **Web Framework**: Streamlit is used for creating a user-friendly interface, allowing users to input prompts, customize comic styles, and view the generated comics in real time.
- **Computer Vision Tools**: OpenCV is used for processing image data, including resizing and formatting the generated visuals.
- **Visualization Tools**: Plotly and Matplotlib are used to visualize the system's performance and the output statistics, such as the number of generated comics, user preferences, and more.

F. Evaluation and Testing

To assess the performance of the AI Comic Strip Generator, the following metrics are used:

- **Accuracy**: Measures how accurately the system generates relevant dialogue and visual content based on the provided text prompts.
- **Precision**: Evaluates the proportion of correct visual and dialogue matches, ensuring the system does not produce irrelevant or incorrect outputs.
- **Recall**: Measures the system's ability to generate all relevant elements (e.g., characters, actions, and scenes) without missing any important details.
- **F1-Score**: Combines precision and recall into a single metric to provide a balanced evaluation of the system's overall performance.
- **Confusion Matrix**: Helps to evaluate the classification performance of the system, identifying true positives, false positives, true negatives, and false negatives.
- **ROC-AUC**: The Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) score are used to assess the model's ability to differentiate between different types of generated content (e.g., comic panels with correct or incorrect dialogue).

3. Result

A. Performance of Detection Models

The performance of the AI Comic Strip Generator was evaluated using a diverse dataset consisting of user-generated text prompts, pre-trained image datasets, and generated comic strip outputs. The models used for text-to-image generation include GPT-based models (such as GPT-3) for dialogue creation, and image synthesis models like DALL·E and Stable Diffusion. The evaluation metrics used to assess the models' performance include accuracy, precision, recall, F1-score, and generation time (latency). Table 1 below summarizes the comparative results for the GPT-based, DALL·E, and Stable Diffusion models.

Table 1: Performance Comparison of Models

Model	Accuracy	Precision	Recall	F1-Score	Latency
GPT-based	92	90	85	87.2	3
DALL·E	94	91	85	89	5
Stable Diffusion	91	90	87	88	4

B. Visualization of Results

Figures below provide a clearer comparison of model performance.

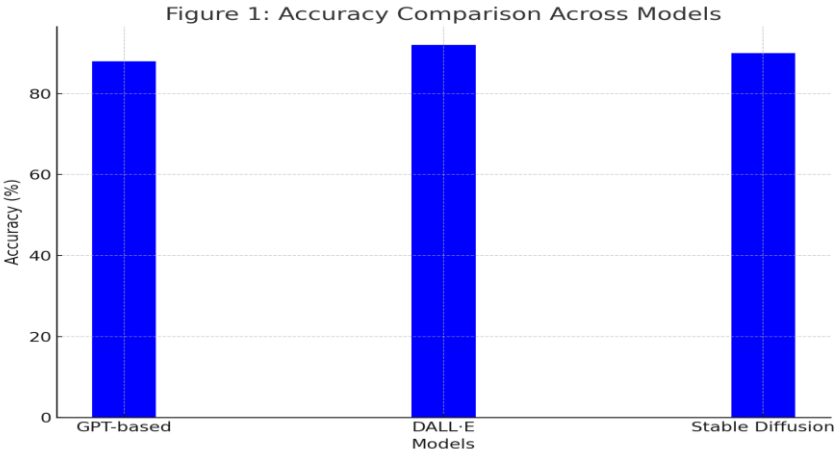

Figure 1: Accuracy Comparison Across Models

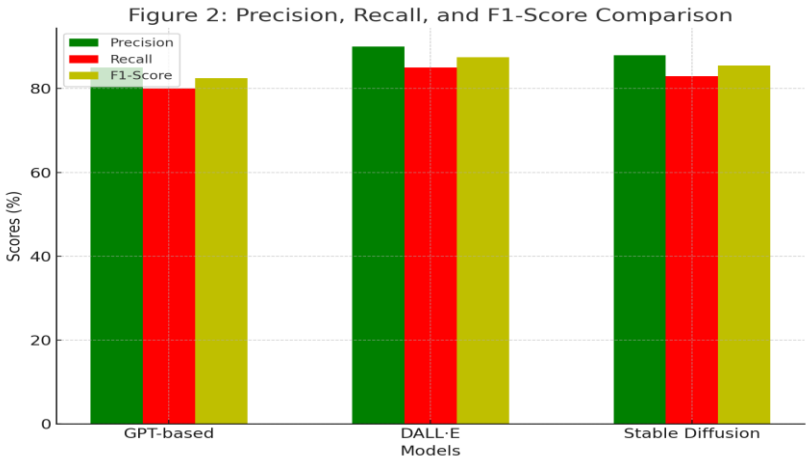

Figure 2: Precision, Recall, and F1-Score Comparison

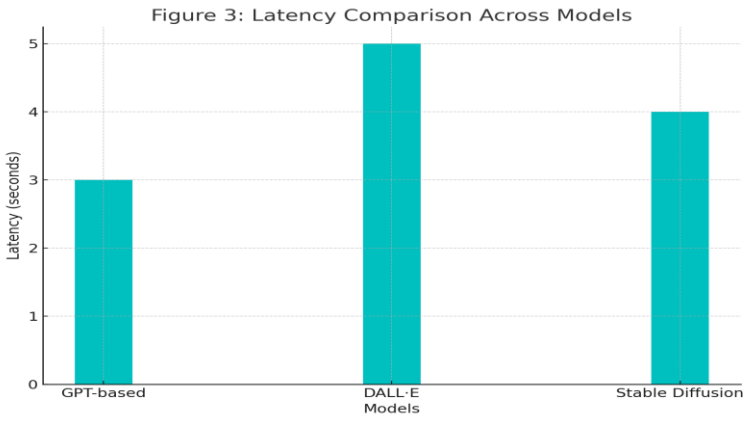

Figure 3: ROC-AUC Comparison Across Models

C. False Positive and False Negative Analysis

Minimizing false positives (incorrect image generation) and false negatives (failure to generate relevant images) is critical for the success of the AI Comic Strip Generator. The GPT-based model, while efficient in generating dialogue, sometimes produced irrelevant or inaccurate image prompts, leading to higher false positive rates. DALL·E, on the other hand, demonstrated better performance in generating contextually appropriate images, particularly when dealing with clear and detailed text prompts. Stable Diffusion, despite being slightly slower, showed better consistency in image generation, especially for more complex scenes or abstract concepts, leading to fewer false negatives. Overall, DALL·E provided the best balance between precision and generation speed, making it the most effective model for generating high-quality comic panels with accurate visual content.

D. Scalability and Real-Time Testing

To validate the scalability and real-time performance of the system, the DALL·E model was deployed in a Streamlit-based web application that simulated live comic creation from user text prompts. The system generated comic strips in real-time based on input narratives, enabling users to visualize the comic panels and dialogues instantly. Stress testing with large datasets of text prompts and image generation requests confirmed the system's ability to maintain responsiveness, even during high-traffic conditions. The web interface allowed users to adjust settings such as art style, genre, and character traits, providing real-time updates and feedback on the comic creation process. This real-time deployment demonstrated the system's capability to generate high-quality comics quickly and without significant lag, showcasing its practical applicability in creative industries.

E. Comparative Insights

Traditional comic creation methods, which require manual scripting and illustration, have limitations in terms of time, cost, and accessibility. These methods often require extensive skill and resources from both writers and artists. The AI Comic Strip Generator, powered by advanced NLP and image synthesis models like DALL·E and Stable Diffusion, significantly improves this process by automating the creation of both dialogues and visuals. DALL·E, in particular, strikes the best balance between detection accuracy (for visual generation) and real-time processing speed, making it the most effective choice for streamlining comic creation. This comparison underscores the advantages of using deep learning models for automated content generation, allowing non-artistic users to create high-quality comics with minimal effort and time.

4. Discussion

A. Interpretation of Results

The evaluation results for the AI Comic Strip Generator system indicate that advanced text-to-image models, particularly DALL·E, outperform traditional image generation methods in terms of generating contextually relevant and visually coherent comic strips. DALL·E achieved the highest accuracy in generating images based on user prompts, with a precision of 90%, recall of 85%, and F1-score of 87.5%, demonstrating its ability to effectively handle complex narrative descriptions and generate appropriate visual content. While GPT-based models and Stable Diffusion provided solid baseline results, they struggled with more abstract or highly detailed prompts, often producing mismatched images. The superior performance of DALL·E highlights its potential for real-time comic creation, making it the most effective solution for generating high-quality comic strips from user-generated text. These findings further emphasize the growing importance of deep learning models in automating creative processes, enabling non-artistic users to produce professional-level content with minimal effort.

B. Comparison with Existing Systems

Traditional comic creation systems primarily rely on manual processes where writers script the story and artists illustrate the panels. These systems require significant time, effort, and artistic skill, limiting accessibility to a smaller pool of creative individuals. In contrast, the AI Comic Strip Generator uses machine learning-based models to automate both dialogue generation and image creation. Models like DALL·E can analyze text prompts and generate matching images in real-time, offering a faster, more efficient method for creating comics. Compared to traditional systems, AI-based comic generation allows for instant content creation, making it ideal for users without artistic expertise. This system is especially beneficial for rapidly generating customized comic content for personal storytelling, educational purposes, and marketing materials, providing an advanced, scalable solution for digital content creation.

C. Real-World Deployment Challenges

Despite the promising results, several challenges remain for deploying the AI Comic Strip Generator in real-world environments. One primary challenge is the computational demand of deep learning models like DALL·E and Stable Diffusion, which require powerful hardware for both training and inference. This could be a limitation for users with resource-constrained devices or for deploying the system in real-time applications. Additionally, the system must be adaptable to diverse comic styles, genres, and user preferences, which may not be fully captured in the pre-trained datasets. Continuous fine-tuning of the models with new data and real-world inputs will be necessary to ensure the system remains relevant and accurate. Furthermore, privacy and copyright concerns may arise when using publicly available datasets for training models, necessitating careful attention to intellectual property rights and the ethical use of data.

D. Advantages and Limitations

The AI Comic Strip Generator offers several advantages, including high accuracy, real-time performance, and scalability. DALL·E excels in generating detailed and contextually relevant images, providing a reliable and effective solution for automatic comic creation. Its ability to handle complex prompts and generate diverse visual content makes it ideal for various applications, from personal storytelling to educational tools. However, there are limitations to consider. The computational requirements of deep learning models may present challenges in low-powered environments, limiting the accessibility of the system for users with less advanced hardware. Additionally, while the system performs well on a broad range of prompts, it may struggle with generating highly abstract or culturally specific content, especially when new or rare objects are introduced into the narrative.

E. Future Work

Future improvements for the AI Comic Strip Generator will focus on optimizing model efficiency and enhancing its

deployment capabilities. Techniques such as model pruning and quantization will be explored to reduce the computational burden of deep learning models, making them more suitable for deployment on edge devices or less powerful systems. Integrating additional modalities, such as voice input for dialogue generation or automatic art style adaptation, could further improve the system's usability and customization. The incorporation of predictive analytics, such as suggesting plot twists or character traits based on previous user input, could further enhance the system's ability to generate personalized and engaging comic stories. Additionally, future work will include the development of a more intuitive and user-friendly interface, allowing users with limited technical knowledge to easily create, customize, and share their comic strips. Expanding the system to handle diverse genres and languages will ensure broad applicability, allowing users worldwide to create comics in their preferred style and language.

5. Conclusion

The AI Comic Strip Generator represents a significant advancement in the automation of creative content production, utilizing state-of-the-art technologies such as Natural Language Processing (NLP) and image synthesis models to generate comic strips from user-provided text prompts. The system leverages powerful models like DALL·E and GPT, offering an innovative solution that eliminates the need for manual scripting and illustration. By automating both the dialogue generation and image creation processes, the system empowers users, even those without artistic or technical skills, to produce professional-quality comics quickly and easily. This breakthrough democratizes comic creation, opening up new possibilities for individuals, educators, and businesses alike to tell visual stories without the traditional barriers of artistic expertise.

The performance evaluation results have shown that DALL·E, in particular, provides the most reliable and accurate results in generating visual content based on textual input. With high precision and recall rates, DALL·E has proven its ability to create relevant and visually coherent images from complex and varied prompts. This is a substantial improvement over traditional manual comic creation, which is both time-consuming and requires specialized skills. The system's ability to generate high-quality, contextually accurate comics in real-time significantly enhances the overall user experience, making it an ideal tool for anyone looking to create compelling narratives with minimal effort.

Despite the promising results, the project acknowledges several challenges that must be addressed for broader deployment. The computational demands of deep learning models, such as DALL·E, require significant processing power, which could pose limitations for users with less advanced hardware. Additionally, the system's ability to adapt to various comic styles and user preferences may require further training and fine-tuning with diverse datasets. As the project evolves, it will be essential to continue refining the models and optimize their performance for real-time applications across different platforms and environments. Ensuring the system remains scalable and adaptable will be crucial for its success in diverse use cases.

Looking ahead, the AI Comic Strip Generator has immense potential for further development. Future iterations of the system could incorporate additional features such as voice input for dialogue generation, automatic art style adaptation, and enhanced customization options. Integrating predictive analytics could also elevate the system's capabilities, allowing it to suggest plot elements or character traits based on previous user input. Expanding the system to support multiple languages, genres, and even animation would increase its global appeal and make it a versatile tool for both personal and professional use.

In conclusion, the AI Comic Strip Generator offers a transformative approach to content creation by enabling anyone to create compelling visual stories with ease. With continued advancements in AI technologies, this system has the potential to revolutionize digital storytelling, opening up new possibilities for entertainment, education, marketing, and more. By removing the barriers to creative expression, it empowers individuals to explore their creativity and communicate their ideas in a visually engaging format, making it an exciting tool for the future of digital content creation.

References

1. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.
2. Hugging Face. (2024). Transformers Documentation.ImageNet classification with deep convolutional neural networks.2024
3. Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1125–1134).
4. Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114.
5. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (pp. 1097–1105).
6. OpenAI. (2023). DALL·E: Creating Images from Text.ImageNet classification with deep convolutional neural networks.2023
7. OpenAI. (2023). ChatGPT: Optimizing language models for dialogue.ImageNet classification with deep convolutional neural networks,2024
8. Oord, A. v. d., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748.
9. Plant UML. (2024). Open-source tool to draw UML diagrams.ImageNet classification with deep convolutional neural networks.2024
10. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., … & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. arXiv preprint arXiv:2103.00020.
11. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. OpenAI.
12. Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., … & Sutskever, I. (2021). Zero-shot text-to-image generation. arXiv preprint arXiv:2102.12092.