Grenze International Journal of Engineering and Technology, January Issue

ComicMath: Automated Comic Generation for Teaching Grade 9 Mathematics

Sanyukta Khetan[1] and Chandravva Hebbi[2]
[1-2]PES University, Bengaluru, India
Email: sanyuktakhetan123@gmail.com, chandravvahebbi@pes.edu

***Abstract*— Traditional methods of teaching Mathematics often result in reduced student engagement, particularly among younger learners. This paper presents an AI-powered educational framework aimed at enhancing the comprehension and interest of Grade 9 students in mathematics through comic-based storytelling. The system accepts a mathematical topic as input and employs the Mistral Large Language Model API to generate a Marvel-themed comic script, aligning with curriculum objectives. The generated narrative is converted into visual comic panels using prompt-based image synthesis via the Stable Diffusion Model. Dialogues are programmatically stored in structured JSON format for efficient manipulation. A face detection module is used to associate characters in the images with their corresponding dialogues, ensuring accurate placement of speech bubbles. Additionally, the system integrates a Text-to-Speech (TTS) engine with clickable panels to provide interactive audio feedback, enabling students to listen to the character dialogues and reinforcing learning through multimodal interactions. To ensure educational reliability, all generated scripts are reviewed by mathematics educators for accuracy and curriculum alignment. This framework demonstrates a novel application of multimodal artificial intelligence combining Natural Language Processing for script generation, Generative Image Modelling for visual content, Computer Vision for character mapping, and Speech Synthesis for interactive learning. The results are an engaging, narrative-driven educational tool that transforms traditional mathematics instructions into an accessible, interactive and visually enriched experience.**

***Index Terms*— AI in education, story generation, comic generation, text-to-image, dialogue mapping, face detection, OCR, Text-to-Speech, generative models, math education, and multimodal learning.**

I. INTRODUCTION

Conventional methods of teaching mathematics often struggle to maintain student engagement, especially at the middle and high school levels. The abstract nature of mathematical concepts, when conveyed through static textbooks or lectures, can limit both comprehension and interest. To address these challenges, we propose ComicMath, an AI-powered framework that simplifies Grade 9 mathematics through Marvel-style comic-based storytelling. ComicMath transforms complex topics into interactive, multi-panel comic strips by integrating natural language processing (NLP), generative image modelling, face detection, and text-to-speech (TTS) synthesis. The system accepts a Grade 9 math topic as input and employs the Mistral Language Model API to generate a curriculum-aligned narrative featuring Marvel Characters.

Grenze ID: 01.GIJET.12.1.95_1

This narrative is converted into descriptive prompts and processed through Stable Diffusion to create comic-styled visuals, with positive and negative constraints ensuring visual consistency and minimizing artefacts. Dialogue data are structured in JSON format, and a face detection model associates speech bubbles with the correct characters based on spatial positioning, thereby enhancing narrative coherence and visual accuracy. To promote interactivity, ComicMath integrates TTS functionality with clickable panels that vocalize dialogues upon user interaction.

This multi-modal design combines visual, textual and auditory instruction through an interactive, narrative-driven and engaging platform.

II. Problem Statement

Mathematics is widely regarded as a difficult and abstract subject, particularly at the secondary school level, where foundational concepts become increasingly complex. In Grade 9, students often face challenges in understanding and retaining mathematical principles due to the limitations of conventional teaching methods, which primarily rely on textbooks, lectures, and written exercises. These traditional approaches frequently fail to address diverse learning preferences and offer limited opportunities for interactive or personalized engagement, resulting in reduced interests, lower academic performance, and diminished long-term retention.

While multimedia tools and visual aids have been introduced to enhance mathematics education, they often lack contextual relevance, interactivity and adaptive engagement. There is a pressing need for an innovative solution that not only simplifies the complex mathematical content but also aligns with academic standards and engages learners through dynamic, relatable formats.

To address this gap, the proposed research aims to develop an AI-powered educational platform that transforms Grade 9 mathematics topics into interactive, Marvel-style comic stories. By combining advanced natural processing, generative image modelling, face detection, dialogue mapping, and text-to-speech(TTS) technologies, the system delivers a multimodal learning experience. This approach fosters improved comprehension and retention through visual storytelling, character-driven interactions and immersive audio feedback.

III. Literature Survey

This section reviews existing work in AI-driven comic creation, highlighting tools that integrate natural language processing and image generation for automated storytelling and visualization. Part A addresses story generation using large language models, Part B discusses text-to-image generation with Stable Diffusion, and Part C explores speech bubble placement and text-to-speech integration within comic panels.

A. Story Generation

Samo et al. [1] fine-tuned Mistral-7B Instruct v0.2 with QLoRA to achieve better Python code generation with scarce GPU resources. It outperforms the base model regarding BERTScore and HumanEval, is demonstrated with a Telegram bot, and is limited by a very small dataset in only Python. The authors recommend broader data for future improvements.

Suh et al. [2] proposed CodeToon, a tool that converts programming concepts into visual comic strips by parsing code into an Abstract Syntax Tree(AST) and mapping it to panels. The system employs Google's Quick Draw dataset for visuals but faces limitations due to low-quality drawings and restricted design options, such as visualizing non-executed code. He authors suggest improvements, including opacity adjustments for non-executed panels and emphasize the need for rigorous evaluation of its educational impact.

B. Stable Diffusion

Jin and Song[3] introduce an AI Comic creation pipeline, where ChatGPT synthesizes the stories and Stable Diffusion creates the images. Their system involves fine-tuning LoRA and ControlNet for preserving character consistency of the narratives.

This approach improves fidelity and structure, though limited by dataset size, prompt reliability and evaluation benchmarks.

Zhao et al. [4] propose a personalized text-to-image framework that integrates StyleGAN's W+ latent space with Stable Diffusion to enhance identity preservation and attribute control. A mapping network and a residual cross-attention module align the two latent spaces, trained on 100,000 synthetic and real images. Limitations include detail loss during image inversion to W+ vectors and reduced performance with multiple identities.

C. Speech Bubble Generation and Text-to-Speech

Yang et al.[5] proposed a fully automated system for generating comic books from videos. The framework includes keyframe extraction and stylization, multi-page layout generation using a genetic algorithm, and emotion-aware text balloon creation based on subtitles and audio signals. Evaluated on movie and TV datasets, the system improves visual quality but depends heavily on subtitle availability, requiring manual corrections when speech recognition is used.

N.-V. Nguyen [6] proposed a deep learning-based approach for comic character detection using CNN models such as YOLO for real-time object detection. The system was trained on the newly introduced Sequencity612 dataset comprising 612 annotated comic pages. While achieving state-of-the-art performance, challenges remain regarding dataset diversity and computational cost.

IV. Proposed Methodology

The proposed system follows a multi-stage pipeline designed to generate interactive Marvel-style comic stories that simplify Grade 9 Mathematics topics. It combines advancements in natural Language Processing, image generation and speech synthesis to create an engaging, educational experience. The complete methodology and system architecture are described as follows:

A. Math Topic Input and Narrative Generation

The system begins with users providing a Grade 9 math topic, which serves as a semantic seed for generating an educational comic narrative. Using the Mistral API, a state-of-the-art language model, the system produces a superhero-themed storyline that is concept-aligned, engaging, and dialogue-rich. Instead of fixing the number of panels, the API generates an n-panel, enhancing clarity, emphasizing character interactions, and ensuring better alignment with image generation and speech bubble placement in later stages.

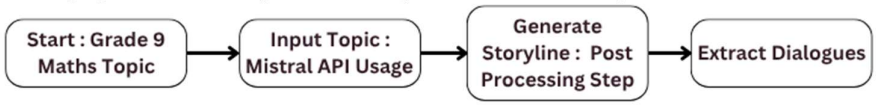

Fig. 1. Math Topic Input and Narrative Generation

The narrative generation process, illustrated in Fig. 1, states a math topic that is processed using the Mistral API to create a coherent storyline, followed by dialogue extraction for comic representation.

B. Image Generation Using Stable Diffusion

Once the storyline is generated, the system proceeds to create a high-quality visual representation for each panel corresponding to the textual narrative. The story is parsed into structured prompts that capture scene descriptions, character names, and associated actions. To preserve the visual fidelity of Marvel characters, a structured JSON file is maintained containing detailed descriptions of 31 characters. Each character entry looks like:

Fig. 2. Character Dataset

To ensure consistent character traits and backstories, a structured dataset comprising character metadata is employed. As shown in Fig.2, each record includes attributes such as name, powers, affiliations, origin, physical features and roles(hero/villain), enabling the language model to maintain contextual coherence across narratives. The visual description field from the JSON file is embedded into prompts for accurate rendering. Stable diffusion is then utilized to generate high-quality comic-style images, with negative prompts incorporated to enhance visual accuracy and consistency.

Fig. 3. Image Generation Using Stable Diffusion

Fig.3 illustrates the image generation process utilizing Stable Diffusion. The input script is first processed through prompt engineering to create detailed textual prompts. These prompts are then fed into the Stable Diffusion model to generate relevant comic-style images. The generated images are subsequently reviewed, and the most contextually appropriate ones are selected and arranged into panels that align with the narrative flow.

C. Dialogue Annotation and Mapping

After image generation, character dialogues are aligned with visual representations in each panel. The dialogue data are structured into a JSON format containing the panel number, character name, and dialogue. A face detection model is then applied to identify and localize character faces, enabling dialogues to be mapped to specific facial regions based on spatial positions and dialogue sequence. To enhance robustness, a fallback mechanism is implemented: if RetinaFace fails, the system employs OpenCV's DNN-based face detector, which uses an SSD architecture with a ResNet-10 backbone, integrated via the readNetFromCaffe() function. This ensures graceful degradation and continued functionality.

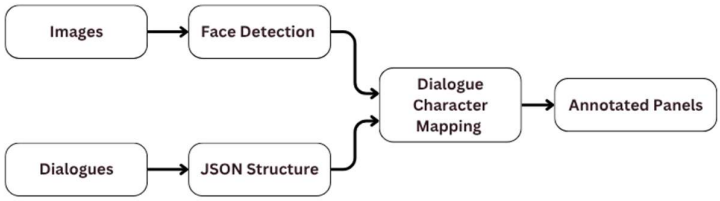

Fig. 4. Dialogue Annotation and Mapping

Fig. 4 presents the dialogue annotation and mapping workflow. Initially, face detection is performed on the generated images, while dialogues are structured in JSON format. These two inputs are then processed in a dialogue-character mapping module, which assigns dialogues to the corresponding characters based on position and structure. The resulting output consists of annotated comic panels with appropriate character dialogues.

D. Speech Synthesis Integration For Interactivity

To enhance accessibility and engage users interactively, the system integrates browser-based Text-to-Speech(TTS) using the native Web Speech API. When a user clicks on a comic panel, a corresponding event listener retrieves the associated dialogues from a structured JSON file. The dialogue is then passed to the speechSynthesis interface for real-time audio generation using the default English voice. The narration is synchronized with the visual panel sequence, maintaining storyline coherence. This functionality supports both visual and auditory learning styles and enhances usability for users with reading difficulties.

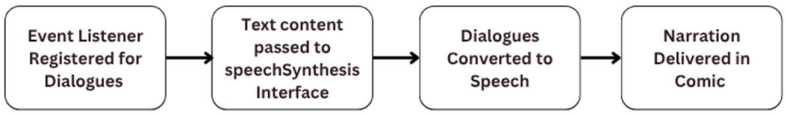

Fig. 5. Text-to-Speech Integration

Fig.5 illustrates the TTS integration process. Dialogues in JSON format are linked to speech bubbles via JavaScript event Listeners and passed to the Web Speech API's speechSynthesis interface for real-time audio narration, enhancing both visual and auditory learning.

E. Final Comic Rendering and User Experience

Once the comic panels are generated with integrated speech narration, the final output consists of a sequence of images arranged in a logically ordered format. Each image represents a panel that visually and textually conveys part of the story. The images are compiled and presented in the correct narrative flow, allowing users to view the entire comic as a continuous story. The structured format ensures clarity in storytelling and supports offline access or sharing of the educational content in image form.

F. System Architecture

The system architecture consists of the following components :

- Input Stage: The User enters a Grade 9 math topic. The topic is sent to the Mistral API.
- Story Generation: Mistral API returns a superhero-style narrative with dialogues.
- Prompt Generation: The narrative is segmented panel-wise. Character visuals are embedded into prompts using a structured JSON dataset.
- Image Generation: Prompts are passed to Stable Diffusion to generate comic-style images.
- Face Detection: RetinaFace or OpenCV detects character faces within the generated images.
- Dialogue Mapping: Dialogues are mapped to detect faces using spatial and sequential logic.
- Text-to-Speech: Browser TTS (Web Speech API) enables narration on bubble click.
- Final Output: An interactive comic viewer with speech-enabled dialogues is rendered.

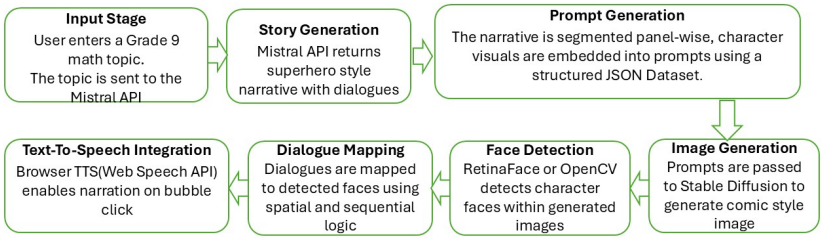

Fig. 6. System Architecture

Fig.6, the architecture comprises a sequence of components that transform a user-defined math topic into an interactive comic experience enriched with narration and visual storytelling.

V. Results and Discussions

Fig. 7. Final Interactive Comic Panel

The proposed system successfully generates an educational comic narrative using a multi-stage pipeline involving Mistral-7B-Instruct-v03 for story generation, structured prompt construction for panel segmentation, and Stable Diffusion for comic-style image synthesis. Each story is broken into individual panels, where one dialogue per frame ensures clarity and visual coherence. Character attributes from a custom dataset are embedded into prompts to maintain visual consistency. Results demonstrate that the system produces coherent instructional narratives for Grade 9 math topics, enhanced through character-driven storytelling. However, performance is influenced by prompt quality, model availability, and occasional repetition in generated plot segments.

The final comic output is a generated panel image merged into one single page layout, with dialogues mapped to the characters through a face detection and spatial alignment pipeline. Interactive text-to-speech allows for accessibility and better user engagement. The compiled sequence of panels visually represents the complete educational storyline from the input topic to the final comic. The completed output for the topic "Algebra" is shown in Fig. 7, illustrating the integration of narrative generation, visual synthesis, and dialogue placement into one unified comic-based learning artifact.

Fig. 7 illustrates the complete comic layout, where all panel images with corresponding dialogues are arranged in a logical sequence to present the topic in a coherent and easily understandable manner.

VI. CONCLUSION

Accurate identification of sources contributing to power quality disturbances is crucial for maintaining the reliable operation of electrical grids, especially as modern systems experience increasing complexity due to the integration of non-linear loads, electric vehicle (EV) chargers, and fault events. This study presents an advanced approach for harmonic source identification using the Tunable-Q Wavelet Transform (TQWT), which provides high-resolution time-frequency analysis for isolating harmonic components. By applying TQWT to voltage and current signals, we effectively capture distinctive harmonic signatures, such as high-order harmonics from non-linear loads, low-frequency distortions from faults, and switching harmonics from EV chargers. The proposed method demonstrates superior accuracy in classifying power quality disturbance sources through the analysis of simulated power system signals, effectively isolating and identifying each source based on its unique harmonic characteristics. Moreover, the TQWT-based approach outperforms traditional Fourier-based methods, particularly in handling non-stationary signals and transients that are common in real-world power systems. Additionally, the study integrates a deep learning-based solution by employing a Convolutional Neural Network (CNN) model for the classification of power quality events and location identification. The dual-task CNN architecture presented in the previous sections offers a robust solution for both classification and regression tasks, ensuring high accuracy and faster convergence in comparison to traditional models. The CNN model, when combined with the TQWT-based feature extraction, significantly enhances the detection and classification of power quality disturbances with minimal loss and high efficiency.

REFERENCES

[1] H.Samo, K.Ali, M.Memon, F.A.Abbasi, M.Y.Koondhar and K.Dahri, "Fine-Tuning Mistral 7B for Python Code Generation using QLoRA," in Proc. Int. Conf. Artificial Intelligence and Applications, pp. 205-217, Jun. 2024, doi:10.21015/v12i1.1885.

[2] S. Suh, J. Zhao, and E. Law, "CodeToon: Story Ideation, auto comic generation and structure mapping for code-driven storytelling," in Proc. 35th ACM Symp. User Interface Software and Technology (UIST '22), Bend, OR, USA, Oct. 29-Nov.2, 2022.

[3] Z. Jin and Z.Song," Generating coherent comics with rich story using ChatGPT and Stable Diffusion," arXiv: 2305.11067, May 2023.

[4] X.Li, X. Hou, and C.C.Loy, "When StyleGAN Meets Stable Diffusion: A W+ Adapter for Personalized Image Generation," arXiv preprint arXiv:2311.17461, 2023.

[5] Y. Yang, I. Kotsia, S. Zafeiriou, and M. Pantic, "Automatic comic generation with stylistic multi-page layouts and emotion-aware text-balloons," in Proc. IEEE Int. Conf. Computer Vision Workshops, 2019.

[6] N. -V. Nguyen, "Comic character detection using YOLO-based models trained on Sequencity612 dataset," in Proc. ACM Int. Conf. Multimedia Retrieval, 2022.

[7] G.G.Sayyad, V.G.Dhumal, V.D. Khandekar, and K.P. Thorat, " A Survey Report on Tect-to-Image Generator Using Stable Diffusion," Int. J.Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 9, no. 10, Oct. 2023.

[8] M. P. M. Perera, et al., "AI-Generated Comic Strips," IRJIET, vol. 7, no. 9, pp. 74–82, Sep. 2023.

[9] N.-V. Nguyen, C. Rigaud, and J.-C. Burie, "Comic characters detection using deep learning," in Proc. 14th Int. Conf. Document Analysis and Recognition (ICDAR), Kyoto, Japan, Nov. 2017.

[10] S. Chen, B. Liu, J. Fu, R. Song, Q. Jin, P. Lin, X. Qi, C. Wang, and J. Zhou, "Neural storyboard artist: Visualizing stories with coherent image sequences," in Proc. 27th ACM Int. Conf. Multimedia (MM'19), Nice, France, Oct. 2019.
[11] Y. Topkaya, V. Batdı, D. Burak & A. Özkaya, "The Effectiveness of Using Comics in Education: A Meta-analytic and Meta-thematic Analysis Study," Ahmet Keleşoğlu Eğitim Fakültesi Dergisi, vol. 5, no. 3, pp. 922–940, Oct. 2023.
[12] M. Şentürk and U. Şimşek, "Educational comics and educational cartoons as teaching material in the social studies course," African Educational Research Journal, vol. 9, no. 2, pp. 515–525, May 2021.
[13] H. Walkington, "The implications of generative artificial intelligence for mathematics education," Social Sciences & Medicine, 2024.
[14] "Artificial Intelligence in Mathematics Education," International Journal of Current Research & Trends (IJCRT), paper IJCRT2408054, Aug. 2024.
[15] A narrative review chapter: "AI-Powered Tools, Thematic Trends and Transformative Applications in Mathematics Education," SSRN, Oct. 2025.