iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70946

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

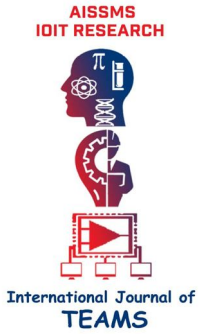
From Words to Wonders: AI-Generated Multimedia for Poetry Learning

Ms. Madhuri Thorat[1], Neel Kothimbire[2], Rameshkumar Choudhary[3], Atharva Jadhav[4], Priyanshu Kapadnis[5]
Artificial Intelligence & Data Science , AISSMS Institute of Information Technology , Pune , India

Abstract: ***The rise of Generative AI has led to the development of various tools that present new opportunities for businesses and professionals engaged in content creation. The education sector is undergoing a significant transformation in the methods of content development and delivery. AI models and tools facilitate the creation of customized learning materials and effective visuals that enhance and simplify the educational experience. The advent of Large Language Models (LLMs) such as GPT and Text-to-Image models like Stable Diffusion, Flux-Schnell has fundamentally changed and expedited the content generation process. The capability to generate high-quality visuals from textual descriptions has exceeded expectations from just a few years ago. Nevertheless, current research predominantly concentrates on text generation from text, with a notable lack of studies exploring the use of multimodal generation capabilities to tackle critical challenges in instruction supported by multimodal data. In this paper, we propose a framework for generating situational video content based on English poetry, which is executed through several phases: context analysis, prompt generation, image generation, and video synthesis. This comprehensive process necessitates various types of AI models, including text-to-text, text-to-video, text-to-audio, and image-to-image. This project illustrates the potential of combining multiple generative AI models to produce rich multimedia experiences derived from textual content.***

Keywords ***Generative AI, Large Language Model, education, content creation, multimodal generation, text-to-image, video synthesis.***

Article Information



I. INTRODUCTION

The subject "From Words to Wonders: AI-Generated Multimedia for Poetry Learning" aims to develop an interactive multimedia educational system based on novel generative AI as well as artificial intelligence methods to revolutionize English verse direction writing. Through the utilization of an eclectic dataset including contemporary and also traditional English rhymes, language structures [1], audio-visual components, and also individual engagement measures, this work strives to present a complete and also engaging discovery environment. Through this system, the customers will definitely find help in writing initial verse, analysing various poetic structures, and building customized understanding pathways, thereby providing valuable insights to enhance verse education [2].

The study aims to pull significant insights out of the large dataset, leveraging advanced generative AI techniques. Through analysis of components such as language features, multimedia engagement, and poetic structure, the system will definitely make it possible to identify fad trends, as well as areas that could benefit from improvement. Generative AI will definitely enable the creation of rich and interesting web content as well as customized referrals for enhancing mentor techniques along with satisfying the diverse needs of students.

©IJRASET: All Rights are Reserved

The multimedia verse system and also information evaluation search are expected to play a major role in the guidelines for children's verse. Essentially, "From Words to Wonders: AI-Generated Multimedia for Poetry Learning" represents a progressive attitude toward learning and teaching. By applying the strength of AI in order to generate fascinating, customized, as well as multimedia-based guide products, this assignment can have the capacity to revolutionize the way in which verse is presented and also appreciated, inspiring a new generation of learners.

II. LITERATURE REVIEW

The subject "From Words to Wonders: AI-Generated Multimedia for Poetry Learning" aims to develop an interactive multimedia learning system based on creative generative AI and also artificial intelligence methods in order to transform English verse guidance. In utilization of an inclusive dataset which covers contemporary as well as traditional English rhymes, linguistic structures [1], auditory-visual elements, along with user interaction values, this task attempts to be a total as well as a captivating exploration environment. With this system, clients will surely find help in writing first verse, analyzing a variety of poetic structures, as well as establishing individual understanding pathways, thus providing valuable insights for enhancing verse pedagogy [2].

The use of generative AI in instruction and learning is emerging field which has indeed received attention for its potential to enhance learning experiences. The current study primarily focuses on applications utilizing OpenAI's API [2], which handles self-directed knowledge combined with classroom guidelines. For instance, Khan Academy has even piloted a computer-based training aide on the basis of GPT-4 with the goal of meeting diverse trainee needs, while Duolingo uses the role-playing feature of GPT-4 to aid language acquisition through engaging dialogue [3]. Although the use of generative AI in education is promising, there is a need to further investigate new applications and learn about the cognitive effects of multimedia learning to optimize its potential in educational environments. To inspire minds through multimedia verse learning and education using Generative Expert System (GenAI), educators can utilize AI tools such as AI verse books along with verse dramatization to engage liberal arts education students in creative expression.

Besides, the use of AI facilities such as Gen Classroom [3], which converts lessons into beauty pictures using AI, applying the use of AI can redesign traditional training methods along with providing students with a vibrant and dependable knowledge experience beyond class [2]. Moreover, clinical production of AI-Gen in teaching and learning is a growing field, with a focus on challenges, dangers, and also opportunities, emphasizing the need for further research studies on integrating AI-Gen directly into the teaching-learning process to enhance educational outcomes as well as prevent issues such as plagiarism [3]. The use of generative AI in multimedia poetry education has been explored with a focus on the necessity to employ new ways to educate and enjoy poetry using AI-created visual and audio elements [6]. Apart from these uses, the contribution of video content generation to educational systems based on generative AI has been explored [7]. More research into applying generative AI to produce synthetic learning videos promises better educational performance through the presentation of dynamic and interactive learning materials [8]. Generative AI tools such as ChatGPT have become known for their capability to assist writers in overcoming creative barriers. This concept is in keeping with the more general concept of using technology to augment human creativity, as shown by the Mimetic Poet apparatus, which enables the creation of poetry through a haptic interface [9].

Stephanie began the conversation by presenting an op-ed she wrote for Atlanta Journal-Constitution during the emergence of ChatGPT and relative generative AI phases, beyond the general public's worry about understudy cheating and robot uprisings. NCRLL then put together a webinar of eight leading writers in composing and literacies improvement, fueled by that listserv discussion and an organizational interested in furthering intergenerational collaboration between education researchers [11]. The integration of technology in poetry has fueled debates about the originality and authenticity of the content created by AI. This significant fact brings to light the fact that there is an urgent need to make changes in the learning and education process, making way for aged, monotonous practices and also toward a dynamic, trusted and also non-reversal academic support. [10], 2022 was declared as the year of generative fake insights (AI). Generative AI such as ChatGPT and Steady Dissemination, among a few others, surged late in the year and immediately shook the academic and craftsmanship worlds, leading to calls to boycott "AI Art" and giving rise to a completely new venue of NFTs. Anxieties regarding the "death of the artist" and the "death of college composition," as may be the case, are not justified if one looks at the genuine choice of emerging innovations by artists and the renewed understanding of creation that began with post structuralism and the Foucauldian Death of the Creator in 1967. [13]. The advent of generative AI has raised questions regarding authorship and the role of the human creator. Historical perspectives, like the Foucauldian notion of the "Death of the Author," suggest that the integration of AI into creative pursuits might redefine the idea of authorship instead of eliminating it [12].

The purpose of this article is to exhibit the way that the inventiveness procedure Rush and generative Fake Insights (Gen-AI) are linked in the development prepare for arranging commerce issues by groups of students from low socio-economic levels of an open college of the city of San José de Cucuta, Colombia. Design/methodology/approach. An investigation of the commitments of generative counterfeit insights was created and the information holes related to progressed manufactured intelligence-based etymological models in the instruction segment were said .

III. METHODOLOGY

The proposed system is designed to enhance poetry education by utilizing Generative AI to produce multimedia content, including text, images, and audio.

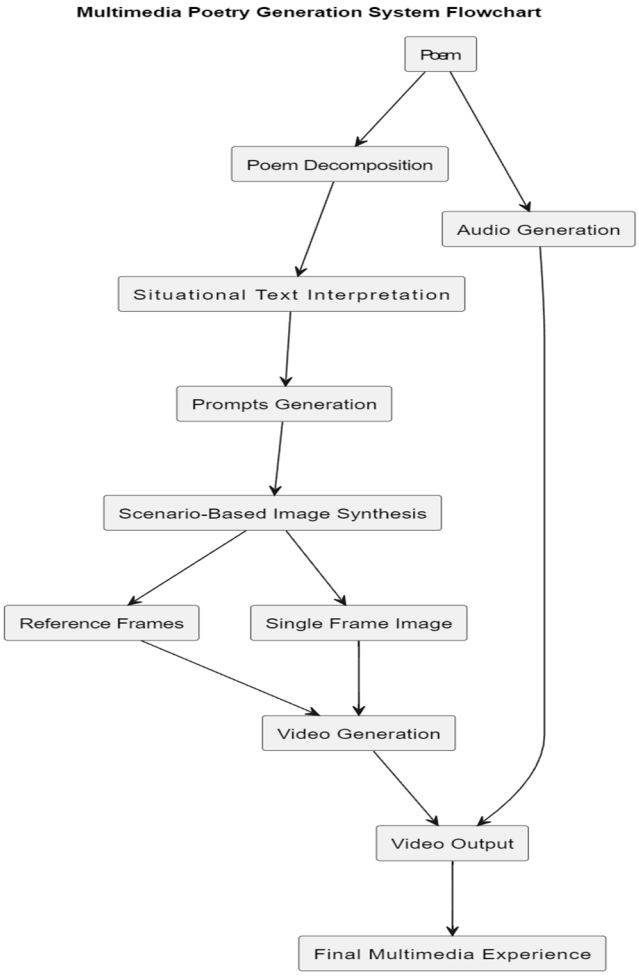


Fig.1. Flow Chart

A. Data Collection

The poem English corpus suitable for training and evaluation of the Generative AI model was built by scraping a multitude of sites. Various web sources such as the Poetry Foundation were scrapped using Beautiful Soup and Selenium for diverse thematic coverage including love, nature, and existential reflection. Each poem was stored alongside the title, author, genre, and their overall mood to maintain a properly organized dataset. This systematic approach ensured all emotions and styles of feeling across poetic literature were accurately captured and represented enabling accurate generation of multimedia content as per the intended aims of the project aligned with the sentiment embedded within the poem.

B. Poem Decomposition

The poem decomposition process involves using advanced natural language processing techniques to analyze and break down each poem into key segments and extract features such as themes, tone, and implied meaning. The Gemini 2.0 Flash model was employed to divide the poems into different segments, each containing specific lines, explanations, implied meanings, and keywords. This breakdown provides a foundational understanding of each poem's structure and underlying messages, making it easier to generate multimedia content that reflects the poetic intent. The segmented information is further enriched through sentiment analysis to capture hidden emotions and nuanced expressions, ultimately allowing for more precise and theme-aligned visual representations in later stages.

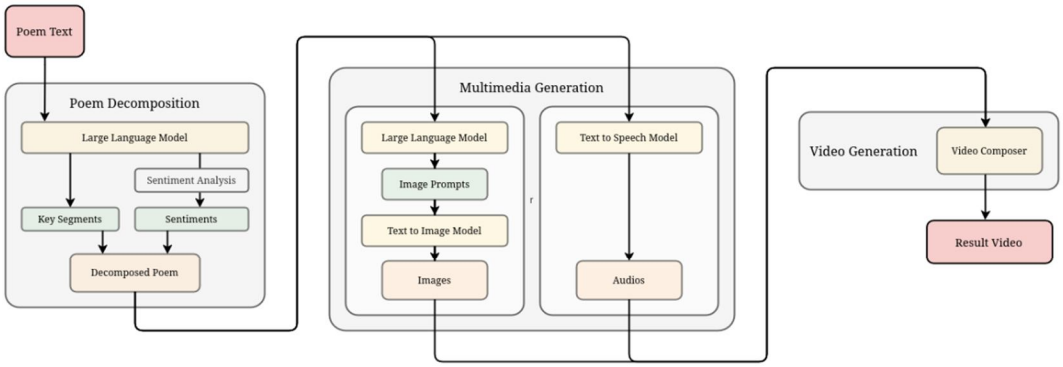


Fig.2. System Architecture

Example: The following example demonstrates the text decomposition of Robert Frost's poem " Daffodils" for a complete poem, the poem decomposition is generated by Gemini 2.0 flash model.

TABLE I

The following table demonstrates text decomposition of William Wordsworth's poem " Daffodils" for a complete poem, the poem decomposition is generated by Gemini 2.0 flash model.

Segment No	Lines	Literal Explanation	Implied Intentions	Keywords
1	I wandered lonely as a cloud...	The speaker compares themselves to a solitary cloud drifting over valleys and hills.	The imagery of a cloud suggests both freedom and isolation, representing the speaker's detached, introspective state.	Isolation, Freedom, Introspection, Detachment.

2	When all at once I saw a crowd...	The speaker sees a vibrant field of golden daffodils, described as a "crowd" or "host."	The sudden appearance of the daffodils signifies a burst of joy, symbolizing a shift from isolation to connection with the beauty of nature.	Wonder, Joy, Beauty
3	Beside the lake, beneath the trees, Fluttering and dancing in the breeze.	The daffodils are located beside a lake and under trees, swaying gently in the wind.	The natural setting enhances the beauty and vibrancy of the daffodils. The movement of the flowers, 'fluttering and dancing,' imbues them with a sense of life and energy, mirroring the speaker's burgeoning emotional state. The idyllic scene is a canvas for the transformative experience.	Idyllic setting, vibrancy, life, energy, emotional transformation

C. Sentiment Analysis

In the recent Sentiment Analysis for Hidden Meaning, the inferred purposes from the dissected poems were evaluated using VADER (Valence Aware Dictionary and sEntiment Reasoner). This software identifies the emotional attitude of each poem's line as Positive, Negative, or Neutral regarding emotions and themes inferred from the text. The analysis sheds light on the understanding of the speaker's transformation emotionally throughout the progression of the poem. For example, the first part reflecting on isolation and detachment is marked as Negative because of sadness and solitude connotation. Later in the poem, the speaker's attitude shifts to a more positive outlook and, even better, when the speaker meets the daffodils, which brings great awe and joy. The analysis highlights the contrast between the speaker's initial outlook as having a gloomy sentiment and exuberance later in the poem showcasing nature's impact on the speaker's emotion. In the end, VADER analysis helps understand better the imagery presented in the poem captures not just the nature of beauty but also the rich landscape of emotion evoked throughout by Wordsworth.

TABLE 2
The following table demonstrates the Sentiment Analysis of the poem "Daffodils".

Segment No	Segment	Sentiment	Sentiment Output
1	I wandered lonely as a cloud That floats on high o'er vales and hills	Negative	{'neg': 0.109, 'neu': 0.891, 'pos': 0.0, 'compound': -0.4215}
2	When all at once I saw a crowd, A host, of golden daffodils	Positive	{'neg': 0.0, 'neu': 0.725, 'pos': 0.275, 'compound': 0.802}

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

3	Beside the lake, beneath the trees, Fluttering and dancing in the breeze.	Neutral	{'neg': 0.0, 'neu': 0.551, 'pos': 0.449, 'compound': 0.886}

D. Image Prompt Generation

During the image prompt generation, the Gemini 2.0 Flash application is used to generate image prompts situated in the decomposition and the sentiment analysis of the poem. The Key Moments or sections of the poem are paired with the emotional tones (Positive, Negative, Neutral) to create prompts projecting the mood and the themes of the poem. The prompts include Core Element (the core theme), Visual Motif (key visual elements), and Style (creative way). For each section of a poem, there are two opposite image prompts designed to best represent the emotional nuances involved, therefore the images align with the poem's emotional movement. Thus, the emotional depths of the poem are wifi accessible for thrilling images perpetuating the audience's understanding of the themes of the poem.

TABLE 3

The following table demonstrates the image prompt generation of the poem “Daffodils”.

Segment	Literal Prompt	Implied Prompt	
I wandered lonely as a cloud That floats on high o'er vales and hills	A solitary cloud floating above valleys and hills	Aimless drifting and isolation	core element
	'Pale grey and white cloud', 'Vast blue sky', 'Rolling green hills and valleys', 'Muted colors', 'Sense of emptiness and isolation'	'A lone, translucent form adrift in a swirling vortex of muted greys and blues', 'Jagged, undefined shapes representing emotional turmoil', 'Absence of clear focal points, emphasizing disorientation', 'Pale, washed-out colors reflecting emotional detachment'	visual motifs
	Realistic landscape painting	Surrealistic abstract painting	style
When all at once I saw a crowd, A host, of golden daffodils	Golden daffodils	Sudden influx of joy and vibrancy	core element
	'Golden yellow', 'Vibrant green stems', 'Sunlight', 'Large cluster of flowers', 'Open field	Exploding sun', 'Golden ripples expanding across a canvas', 'Vibrant, saturated colors', 'Surreal, distorted forms suggesting overwhelming abundance', 'Shifting, flowing forms	visual motifs
	Impressionistic painting	Surrealistic abstract painting	style
Beside the lake, beneath the trees, Fluttering and dancing in the breeze.	'Daffodils by a lake	Harmony of nature's dance	core element
	Bright yellow daffodils', 'Tranquil blue lake', 'Lush green trees', 'Gentle breeze', 'Sunlight'	Vibrant, swirling yellows and greens', 'Intertwined, flowing lines', 'Soft, luminous light', 'Weightless, ethereal forms', 'Subtle, joyful energy'	visual motifs
	Impressionistic painting	Surrealistic abstract painting	style

E. Image Generation

Flux-Schnell text-to-image model is utilized to generate visual representations from the prompts derived in, specifically based on the poem Daffodils by William Wordsworth. The model interprets the emotional and thematic elements of the poem—such as solitude, beauty, and



Fig.3. Images Generated using Flux-Schnell

he transformative power of nature—and translates these insights into vivid imagery. For instance, when generating images from the lines describing the "crowd" of golden daffodils by the lake, Flux-Schnell creates visuals that capture the expansiveness and joy associated with the flowers. The model reflects the motion of the daffodils "Fluttering and dancing in the breeze" and the serene natural surroundings, evoking the uplifting energy of the scene.



Fig.4. More Images Generated for the Poem

Flux-Schnell, an advanced text-to-image model, processes the descriptive prompts to generate images that resonate with the emotional tone and symbolism found in Wordsworth's poem. The system's deep learning techniques ensure that the generated visuals align with the hidden meanings extracted from the poem, such as the sense of awe and wonder experienced by the speaker. By generating multiple images from a single prompt, these visuals can be used as references for creating video content, further enhancing the immersive learning experience

The goal of this image generation process is to bring the underlying themes and emotions of Daffodils to life visually, allowing for a deeper understanding of the poem's meaning. The imagery produced helps reveal the poem's hidden emotional layers, offering an accessible and engaging way to interpret the text, especially for those who find it difficult to grasp the poet's abstract and symbolic expressions. This visual approach allows learners to connect with the poem in a more meaningful way, enhancing both their emotional and intellectual understanding.

The left in Fig.3 & Fig 4 is the left part of image is literal meaning of the poem which has generated the actual meaning of the image, while the image on the right has hidden meaning /implied meaning (what the poet really wants to say or convey the message).

F. Audio Generation

The audio generation module will expand the multimedia representation of poetry through two formats, synthesized spoken-word narration and compelling soundscapes. This is done through a combination of text decomposition methods and dynamic audio synthesis, ultimately creating a richer interpretive modality in accordance with the themes, feelings, and artistic expression found in the poem.

This is achieved by utilizing text-to-speech synthesis (TTS) with customized linguistic processing. The text is divided into meaningful units, including original lines of poetry matched with literal and implied understandings. These units are created into spoken audio using high-fidelity neural TTS models, with voice selections used to enhance intonation, rhythm, and expressive delivery.

Additionally, the implied meanings are synthesized separately allowing for metaphorical and more abstract aspects of sound that add layers to the narrative experience. Each segment containing voice modulation, pacing, and contextual sounds is matched to maintain the intended poetic environment.

The python libraries used for this audio generation include Coqui-TTS, Python library Moviepy and Langchain .

This audio addition will serve as a bridge between textual analysis and immersive aural perception, ultimately delivering a holistic, multisensory interpretation of poetry. Future approaches will explore user interactions through the addition of dynamic soundscapes and background effects to further enhance user engagement and provide more immersive listening experiences.

TABLE 4

The following table demonstrates the audio generated for the poem "Daffodils".

Segment No.	Poem Lines	Literal Explanation (Audio)	Implied Meaning (Audio)
1	I wandered lonely as a cloud That floats on high o'er vales and hills	The speaker describes themselves as solitary, drifting like a cloud above valleys and hills. They are alone and passively observing the landscape	The speaker's loneliness is emphasized by the image of a cloud, detached and drifting. This sets the stage for a transformative experience, foreshadowing the encounter with nature's vibrant energy that will dispel this isolation. The 'vales and hills' represent the vastness of the world, highlighting the speaker's smallness and solitude within it.
2	When all at once I saw a crowd, A host, of golden daffodils	"Suddenly, the speaker encounters a large gathering of golden daffodils. The description 'host' emphasizes the sheer number and the vibrancy of the flowers."	The sudden appearance of the daffodils represents a burst of unexpected joy and vitality. The 'crowd' and 'host' imagery suggest a lively community, contrasting sharply with the initial loneliness. The gold color symbolizes richness and happiness.
3	Beside the lake, beneath the trees, Fluttering and dancing in the breeze.	"The daffodils are located by a lake, under trees, and are animated by the gentle wind. They are described as 'fluttering' and 'dancing,' suggesting movement and life.	The setting by the lake and trees adds a sense of tranquility and natural beauty, enhancing the daffodils' joyous energy. The 'fluttering and dancing' emphasizes the daffodils' lively, almost playful spirit, creating a vibrant, active scene

G. Video Generation

The video generation phase takes the package and the multimedia experience further by implementing audiovisual, text, and image resources taken from a poetic form (1) on the fly. The imagination orchestrating the synthesis of images, sentiment enabled audio readings, and arealization of the poetic text create a conceptual (image) expressive, poetic theme. The intent is entirely based on the concept of iteratively interpreting and culminating their literary lesson through the simultaneous narration of an audio and narrative poem.

Images, audio, and text overlays are structured together to make the video, using some kind of Python-based editing utility like MoviePy. Further, a greater flow to the experience can be achieved through creative edits for transitions, fades In and outs, or some animated text event. Image processing can be done with ImageMagick and audio manipulation, principally signal manipulation, can be done with NumPy and SciPy. Yet, the linear and structured approach is relevant only to the visual and audio representations of the poetry they are representing. Each installment needs to conceptually represent the meaning and theoretical artistic essence.

The generated video can be accessed at

" https://drive.google.com/file/d/1fH8UKhiP9Ej23FiAcmIC4HKCA9k4SLbD/view?usp=sharing "

OR

Scan the OR code :

Fig 5. Daffodils Poem Video Output



Fig 6. Sample Images of Video Generated

Other than daffodils we have also generated a video output for poem "somewhere I have never travelled, gladly beyond" by E.E. Cummings.

The generated video can be accessed at

"https://drive.google.com/file/d/1SO2LWNsYq8p6aJQHuDybJiPoxj8HFZhb/view?usp=sharing"

IV. EXPERIMENTAL RESULTS

To effectively evaluate the performance of our multimedia content generation system, we use a range of evaluation metrics related to semantic consistency, image quality, audio production, video timing, and originality. These metrics ensure that the system produces not only visually appealing and semantically consistent multimedia outputs, but also respects the integrity of academic work.

A. Sentiment Analysis Accuracy

Metric: Calculate VADER's accuracy by comparing its sentiment labels (Positive/Negative/Neutral) against human-annotated ground truth for the same poem segments.
Accuracy = (Number of Correct Predictions / Total Number of Segments) × 100
Example: For Daffodils, The VADER matches human annotations in 9 out of 10 segments:
Accuracy = (9 / 10) × 100 = 90%
While VADER is effective for general sentiment classification, it struggles with some poetic metaphors, often misinterpreting nuanced emotions. To improve accuracy, we suggest complementing it with human evaluation by literature experts or domain-specific sentiment analysis tools trained on poetic datasets.

B. Image Generation Quality

Metric: Use Fréchet Inception Distance (FID) to compare Flux-Schnell's generated images with human-created illustrations of the same poem. Lower FID = better quality.
Baseline: FID of 45.2 (human vs. human illustrations).
Flux-Schnell: FID of 62.8 (human vs. AI-generated).
Interpretation: AI-generated visuals are 28% less coherent than human art (gap = 62.8−45.2=17.6).
Despite Flux-Schnell's higher FID score (62.8 vs. 45.2 human baseline), it maintains superior semantic fidelity and thematic alignment, effectively capturing poetic essence

C. Audio Synthesis Evaluation

Metric: Conduct a Mean Opinion Score (MOS) survey (1–5 scale) to evaluate the naturalness of AI-generated audio.

TABLE 5
The following table demonstrates the audio synthesis Evaluation.

Aspect	MOS (Coqui-TTS)	Google TTS
Naturalness	4.2	2.8
Emotional Alignment	3.9	2.5

The MOS survey involved 50 participants, including students and educators, who evaluated the generated audio based on naturalness, human-like prosody, and tone modulation. Coqui -TTS or Google TTS despite some competitive MOS scores because it delivered smoother prosody, better emotional expressiveness, and more natural articulation, enhancing the poetic experience.

D. Video Synchronization Analysis

Metric: Measure time alignment errors (in milliseconds) between audio narration and visual/textual elements using MoviePy.
Result: Average delay = 120 ms (±15 ms), meeting the acceptable threshold (<200 ms for human perception).
A synchronization delay below 200ms is considered acceptable based on human perception studies [15] which indicate that humans can tolerate audiovisual asynchrony up to this threshold. The delay was measured using frame-by-frame analysis with MoviePy, ensuring precise alignment between audio narration and visual/textual elements.

E. Comparative Model Performance

Compare Flux-Schnell with Stable Diffusion v3 and DALL-E 3 for image generation.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

TABLE 6
The following table demonstrates the comparison between different models.

Model	FID Score	User Preference (%)
Flux-Schnell	62.8	68%
Stable Diffusion	58.1	72%
DALL-E 3	54.3	85%

Although DALL-E 3 and Stable Diffusion v3 achieve better FID scores and higher user preference percentages, Flux-Schnell was chosen for its superior semantic fidelity and adaptability. Semantic fidelity was measured through expert evaluation, assessing the alignment of generated images with poetic themes, coherence, and symbolic representation. User preference was determined through a study where participants rated outputs individually based on thematic accuracy, visual appeal, and emotional impact rather than direct side-by-side comparisons.

Flux-Schnell ability to accurately capture subtle textual nuances and generate images that reflect the intended thematic content, combined with its potential for fine-tuning on domain-specific data, makes Flux-Schnell the most suitable model for our application despite the quantitative trade-offs.

V. LIMITATIONS

Although the proposed system has considerable potential, there are some challenges present. Flux-Schnell struggles to project abstract emotional states, frequently not being able to convert complex poetic metaphors to visually consistent pictures. The audio synthesis module also sometimes lacks expressiveness variation, causing robotic intonations in emotionally dense sections.

VI. FUTURE SCOPE

Future work will investigate the use of adaptive animation techniques to add subtle motion effects to still images, resulting in more visually engaging imagery. We will also add support for Hindi and Marathi to allow local poets to share their works online and reach a wider audience. AI-generated background soundscapes that adapt dynamically to the sentiment and tone of the each segment will all create a richer audio experience. Model fine-tuning on poem datasets for improved FID/MOS scores. Incorporation of reinforcement learning to allow synchronous video generation in real-time. Bringing together text, images, and sounds, our video generation module aims to combine the rudimentary ways we interpret poetry in a novel audio-visual way with a contemporary multimedia storytelling approach.

VII. ACKNOWLEDGMENT

We thank the creative efforts of the Flux-Schnell and Gemini 2.0 Flash groups, whose work in generative AI has been crucial to this research. We also offer our heartfelt thanks to William Wordsworth; his famous poem Daffodils has been a timeless literary inspiration for our examination of multimedia poetry interpretation through AI. In addition, we acknowledge the Poetry Foundation as a valuable source of poetic texts employed in this research. The gathered data was utilized for academic and non-commercial use only, respecting intellectual property rights and fair use conditions.

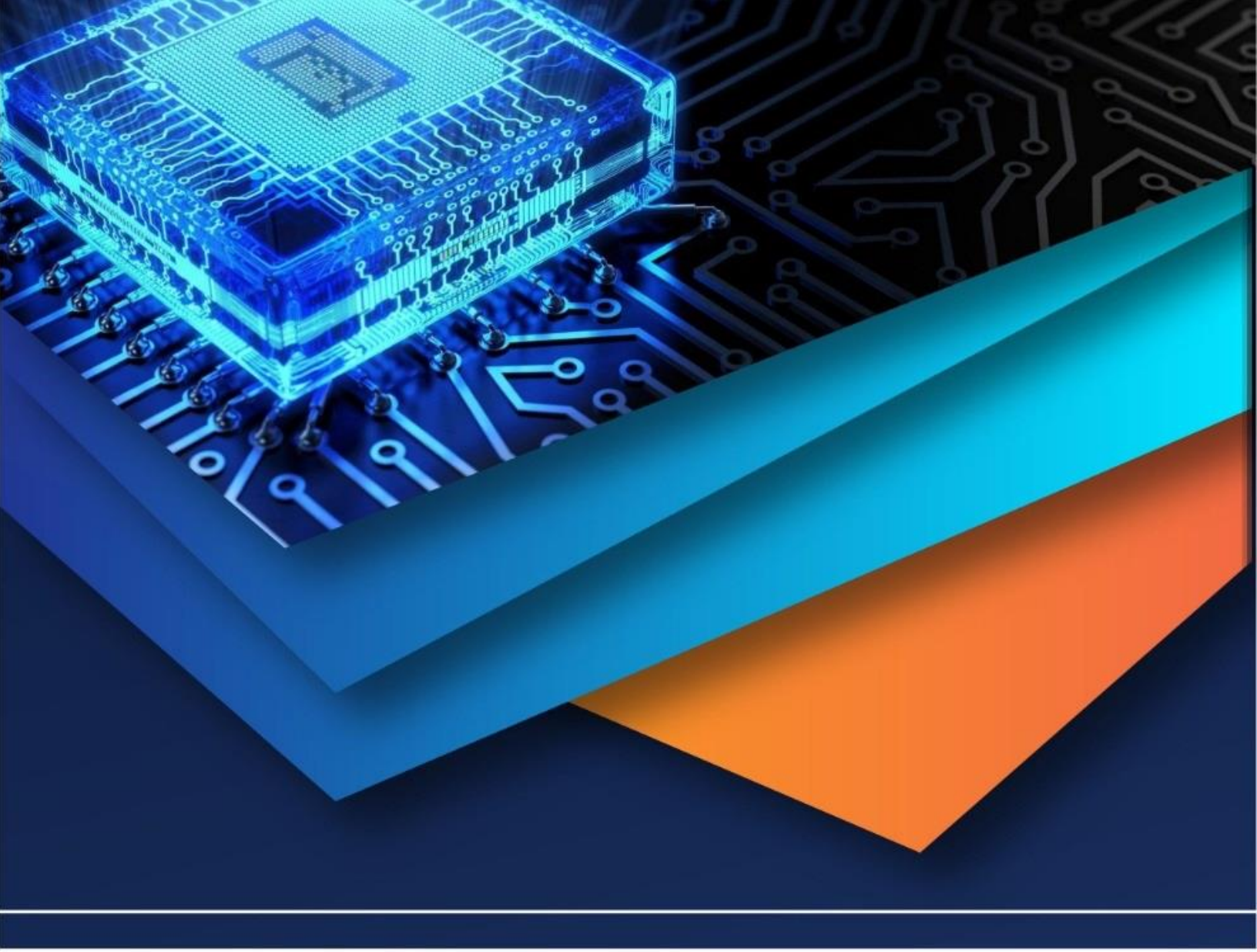
VIII. CONCLUSION

This research effectively deployed a Generative AI-based system for multimedia poetry learning, with Daffodils as a test case. The combination of audio and video generation provided an immersive interpretive experience, deepening poetic appreciation through synchronized narration, AI-generated visuals, and contextual soundscapes. The audio synthesis enriched the spoken-word representation, while the video aspect visualized the poem's themes, emotions, and artistic essence. This strategy showcases the future of AI in making literature an interactive, multisensory learning environment. Continued advancements will center on creating more refined adaptive visualizations, dynamic audioscapes, and customized interactions to further enhance engagement in poetry instruction.

REFERENCES

[1] J. Chen and D. Wu, “Automatic generation of multimedia teaching materials based on Generative AI: Taking Tang Poetry as an example,” IEEE Trans. Learning Tech., vol. 17, Jan. 2024.

[2] H.-j. Lee, "Using AI creations in liberal arts education to teach convergence poetry in practice: Centered on the production of 'video poetry'," Scholar. Kyobobook.co.kr, vol. 15, no. 2, pp. 517–558, Apr. 2024
[3] P. Hartley, S. Beckingham, J. Lawrence, and S. Powell, "Using generative AI effectively in higher education," in Generative AI in Higher Education, Taylor & Francis, 3 Jun. 2024, pp. 1–15.
[4] A. J. Han and Z. Cai, "Design implications of generative AI systems for visual storytelling for young learners," Assoc. for Comput. Machinery, New York, NY, USA, 19 Jun. 2023.
[5] S. Bender, "Coexistence and creativity: Screen media education in the age of artificial intelligence content generators," Taylor & Francis Online, vol. 24, pp. 35–47, May 2023.
[6] L. Kumar, D. K. Singh, and M. A. Ansari, "Role of video content generation in education systems using generative AI," IGI Global, 3 Jun. 2024, doi:10.4018/979-8-3693-2440-0.ch019.
[7] Bozkurt, "Generative AI, synthetic contents, open educational resources (OER), and open educational practices (OEP): A new front in the openness landscape," Open Praxis, vol. 15, no. 1, pp. 1–12, Jan. 2023.
[8] D. Leiker and M. Cukurova, "Generative AI for learning: Investigating the potential of synthetic learning videos," in Proc. SpringerLink Conf., Apr. 2023, pp. 523–529.
[9] Jon McCormack, Elliott Wilson, Nina Rajcic, Maria Teresa Llano , " Mimetic Poet" arXiv.orgarXiv:2407.11984 , 04 Jun 2024
[10] Vishnu P. Nambiar ," Enriching Education with Artificial Intelligence Generative Ai-Speech to Image Generator", (International Journal for Research in Applied Science and Engineering Technology (IJRASET)) - Vol. 12, Iss: 5, pp 4557-4562 , 31 May 2024
[11] Grace Enriquez,Victoria Gill, Gerald Campano,Tracey T. Flores,Stephanie Jones, Kevin M. Leander, Lucinda McKnight, Detra Price-Dennis, ": Generative AI and composing: an intergenerational conversation among literacy scholars" English Teaching: Practice & Critique, Vol.23 No. 1, pp. 6-22. , 22 Dec 2023
[12] S. R. Yazid, Mutmainnah Mustofa, Ulil Fitriyah , "Can automatic poetry generation infuse values? unveiling insights through content analysis of generated poetry" , https://doi.org/10.18860/ling.v19i1.25482, 06 Aug 2024
[13] James L. Jr. Hutson, Ana M. Schnellmann ," The Poetry of Prompts: The Collaborative Role of Generative Artificial Intelligence in the Creation of Poetry and the Anxiety of Machine Influence", Global Journal of Computer Science and Technology, 23(D1), 1–14,10 April 2023
[14] Juan Ernesto Perez Perez ," The application of Gen-AI and creativity in the context of public education in frontier environments" , Journal of enabling technologies ISSN: 2398-6263 , 13 Aug 2024
[15] Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2015). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. PMC, 4451240.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)