Grenze International Journal of Engineering and Technology, January Issue

Image Generation using Quotes by Llama 3.2: A Novel Approach to Quote-to-Image Generation

Sanket Dangle[1], Aditya Dhadway[2], Dev Mehta[3], Sarat Alane[4] and Pradnya Patil[5]
[1-5]K. J. Somaiya Institute of Technology/Computer Engineering Department, Mumbai, India
Email: sanket.dangle@somaiya.edu, aditya.dhadway@somaiya.edu, dev.m@somaiya.edu, s.alane@somaiya.edu, pradnya08@somaiya.edu

***Abstract*—A unique approach to text-to-image generation using quotes, employing Llama 3.2 as a language model for prompt generation, alongside Stable Diffusion XL for high-quality image synthesis. By integrating natural language processing (NLP) with sentiment analysis, the system interprets the emotional tone and thematic content within quotes, producing images that capture the intended sentiment and context. Our method demonstrates the potential of AI for creative applications, providing artists and designers with a powerful tool for translating language into expressive visuals. The model achieves high-quality, thematically aligned outputs, showcasing advancements in AI for creative expression.**

***Index Terms*— Text-to-Image Synthesis, Quote Sentiments, Prompt Modifier, Generative AI, Visual Representation of Quotes.**

I. INTRODUCTION

The intersection of artificial intelligence (AI) and creative expression has evolved rapidly in recent years, especially with the advent of models that transform textual descriptions into unique visual representations. This project, titled "Image Generation Using Quotes by Llama 3.2," aims to explore the potential of AI in bridging the gap between language and visual art by converting complex, sentiment-laden quotes into emotionally resonant images. While existing models like DALL-E, Midjourney, and Stable Diffusion have showcased impressive abilities in generating high-quality images from simple descriptions, they often struggle with nuanced emotional interpretation and thematic depth, particularly when processing abstract or poetic language.

Current text-to-image models typically generate visually compelling outputs but may overlook subtle emotional cues embedded within complex texts. They may struggle to convey a sense of nostalgia, hope, or introspection that often accompanies carefully crafted quotes. This project addresses these limitations by employing Llama 3.2 for advanced prompt generation and Stable Diffusion XL for image synthesis. Llama 3.2, known for its proficiency in natural language processing (NLP) and contextual understanding, captures the subtleties within quotes, while Stable Diffusion XL transforms these refined prompts into detailed, high-resolution images.

The core innovation in this project lies in the integration of sentiment analysis and thematic extraction into the text-to-image generation pipeline. By analyzing each quote's sentiment—whether joy, melancholy, calmness, or dynamism—alongside thematic elements, the system creates prompts that guide the image generation model in capturing not only the literal meaning but also the intended emotional undertone. This is particularly valuable for artistic and educational applications, where aligning the visual output with the intended sentiment enriches viewer engagement and comprehension.

Grenze ID: 01.GIJET.11.1.37_1

The objectives of this project are threefold: first, to develop a robust end-to-end pipeline that accurately interprets quotes' sentiment and thematic content; second, to create an intuitive user interface that allows users to input quotes and visualize their themes; and third, to evaluate the effectiveness of the generated images in terms of quality, coherence, and emotional alignment with the original text. By achieving these objectives, this project contributes to AI-assisted creative expression, offering a tool that could be beneficial for artists, educators, and designers seeking to visualize complex ideas and emotions embedded within text.

This paper explores each aspect of the system, from sentiment analysis and prompt generation to image synthesis, with a focus on the unique contributions of Llama 3.2 and Stable Diffusion XL in addressing the challenges of emotional and thematic representation. We present a comprehensive methodology, including a detailed system architecture and evaluation of image quality and relevance. In the following sections, we discuss the foundational literature, technical architecture, experimental results, and future implications of using AI-driven image generation to visualize language in ways that are contextually accurate and creatively engaging.

II. Literature Survey

In recent years, the field of text-to-image generation has seen remarkable advancements, with researchers developing innovative models that leverage deep learning techniques. Notable contributions include OpenAI's DALL-E, which uses transformers to create imaginative visuals, and the combination of VQGAN and CLIP to produce high-resolution images from textual prompts. Researchers like S. S. AlamA et al. have pioneered conditional diffusion models [1][3][9], while attention mechanisms introduced by X Duan et al. enhance the relevance and quality of generated images[4][6]. These developments mark significant progress toward bridging language and visual content, laying the groundwork for future exploration in generative modelling.

A. Image Generation with Stable Diffusion AI [1]

This paper proposes a system for generating facial images of suspects using Stable Diffusion AI, aiming to improve the accuracy and efficiency of suspect identification in law enforcement. The system utilizes text descriptions of suspects' appearances to generate realistic facial images and incorporates user feedback for refinement.

B. A Taxonomy of prompt modifiers for text-to-image generation [2]

This paper identifies and presents a taxonomy of six prompt modifiers used by practitioners in online text-to-image communities to control the output of AI image generation models. The paper explains how practitioners use these modifiers in the practice of prompt engineering, which is conceptualized as an iterative and experimental conversation with the AI model, to generate images with desired subjects, styles, and qualities.

C. Stable Diffusion Text-Image Generation [3]

This research paper describes Stable Diffusion, a text-to-image deep learning model that generates images from text descriptions. It explains the model's architecture, which includes latent diffusion models, perceptual image compression, and conditioning mechanisms.

D. PKU-AIGI-500K: A Neural Compression Benchmark And Model for AI-Generated Images [4]

This paper introduces PKU-AIGI-500K, a large-scale dataset of AI-generated images (AIGI) for benchmarking and improving AIGI compression methods. The authors analyze the characteristics of AIGIs, benchmark existing compression methods, and propose a novel AIGI compression model called Cross-Attention Transformer Codec (CATC) that utilizes text prompts to improve compression efficiency.

E. Question Generation from PDF using Langchain [5]

This research paper focuses on generating questions from PDF documents using LangChain, a powerful natural language processing framework. The authors describe the development of a web application that leverages LangChain, along with Google PaLM and Streamlit, to extract information from PDFs and generate relevant questions based on user queries.

F. Neural Language Models in Natural Language Processing [6]

This paper examines natural language processing (NLP), specifically the use of neural networks and deep learning techniques for various NLP tasks.

The authors discuss the evolution of NLP, the core areas where deep learning has been applied, different types of neural language models, and the advantages and limitations of each model.

G. Exploring Natural Language Processing in Model-To-Model Transformations [7]

This paper explores the application of natural language processing (NLP) in model-to-model transformations. The authors evaluate several state-of-the-art NLP tools and techniques, including deep learning-based sequence tagging models and dependency parsing, for tasks such as verb/noun phrase extraction and acronym detection.

H. QueryMintAI: Multipurpose Multimodal Large Language Models for Personal Data [8]

This research paper presents QueryMintAI, a multimodal Large Language Model (LLM) capable of processing diverse inputs like text, images, videos, and documents, while generating outputs across various modalities. The paper highlights QueryMintAI's advantages such as its ability to handle multiple modalities and privacy-focused design through local data processing.

I. Realistic Image Generation from Text by Using BERT-Based Embedding [9]

This paper proposes a new text-to-image generation model that utilizes a pre-trained BERT model for text embedding and a StackGAN architecture for generating realistic images. The authors argue that fine-tuning the pre-trained BERT model on a target dataset results in richer text embeddings compared to using text encoders pre-trained on image-text pairs.

J. The Work of Art in the Age of AI Image Generation [10]

This essay examines art generated by AI image generators such as DALL-E and discusses if AI-generated images can be considered art and if AI can be considered an artist. The author proposes that AI image generation can be viewed as a poietic process, where artistic subjects, objects, and roles are not predetermined but emerge through processes and performances involving humans and non-humans.

III. METHODOLOGY

The primary objective of this project is to integrate natural language processing (NLP), sentiment analysis, and image generation to produce visuals that accurately represent the meaning and emotion of textual quotes input phase, where quotes are tokenized, and thematic and emotional content is analyzed through sentiment analysis. Based on this analysis, a suitable prompt is generated and passed to an image generation model to create visuals that resonate with the textual input. The system thus aims to produce images that closely align with the intent of the quotes, making them visually intuitive and high-quality representations of the original text. [3][4][17]

A. Architecture

The system architecture is designed to achieve specific objectives, focusing on creating an efficient flow from text input to image output. The system design includes Data Flow, Use Case, and Sequence diagrams, each providing a framework for organizing interactions and data processing within the architecture. Fig. 1 demonstrates the architecture below, it consists of two main components: the Front-End Component, which gathers the input quote from users, and the Processing Layer, which converts the quote into a suitable prompt for the image generation model. This model then uses the prompt to create an image.

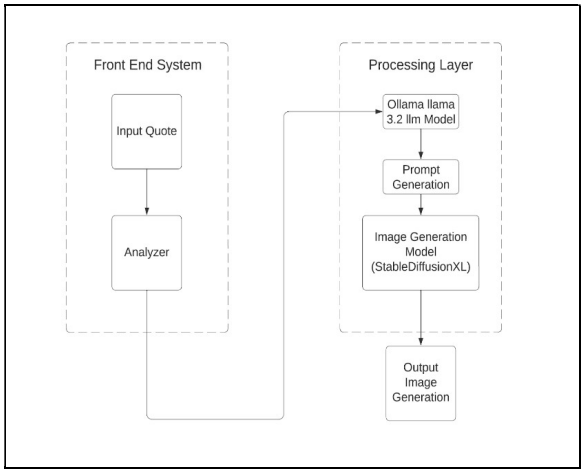

Fig 1. Proposed System Architecture

B. System Implementation

The technology stack supporting this project includes compatible operating systems, programming languages, libraries, and tools to handle NLP, prompt generation, and image synthesis tasks efficiently. Python 3.8 or later is selected as the core programming language for its extensive library support and suitability for NLP and deep learning tasks. Key libraries and frameworks include LangChain for natural language processing and sentiment analysis, the Transformers Library for LLama 3.2 model management, and both PyTorch and TensorFlow for deep learning and model deployment, ensuring a robust computational foundation [13][20]. Stable Diffusion Implementation, integrated with PyTorch, is used to generate high-quality images that align with the prompt's sentiment.

Fig. 2 displays the quote input field, where users enter a quote for which an image is to be generated. The Analyzer Module then processes the quote, generating a creative prompt with LLama 3.2 that aligns with the quote's tone and theme. For example, a quote like "Time and Tide wait for no man" emphasizes the continuous flow of time and the necessity of wise decisions. With this input, the Analyzer Module captures the serious, focused tone, generating a prompt like "Illustrate the ancient Greek phrase 'Time and Tide wait for no man' in a dynamic, abstract scene with swirling water currents and a lone figure standing resolute against the flow.". Over time, it generates increasingly realistic images based on the feedback received from the discriminator.

This prompt is passed to the Stable Diffusion XL model, which generates the image based on these instructions. Fig. 3 presents the final generated image, aligning with the quote's intent. The LLama 3.2-generated prompt is shown to the user, who can re-enter the input for a revised prompt if needed to ensure the image fully encapsulates the quote's depth.

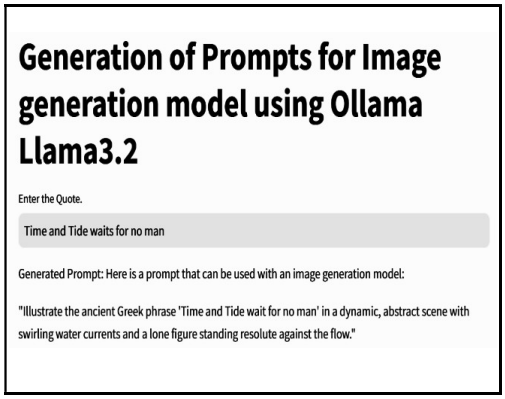

Fig 2. Quote Input Field

Fig 3. Output Image Generated by the System

IV. Results & Analysis

To assess the image quality produced by the model, two key metrics were calculated: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The PSNR value is expressed in decibels (dB) and generally reflects image quality with higher values indicating closer similarity.

The formula for PSNR is as follows:

$$PSNR = 10 \cdot \log_{10} \frac{MAX^2}{MSE}$$

where

- MAX is the Maximum pixel intensity of the image (typically 255 for 8-bit images).
- MSE is the Mean Squared Error between the original image and the generated image. It is calculated as follows:

$$MSE = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (I(i,j) - K(i,j))^2$$

where

- *M* and *N* are Image dimensions.
- *I(i, j)* is the pixel intensity of the original image at position *(i, j)*.
- *K(i, j)* is the pixel intensity of the generated image at position *(i, j)*. [19]

For Average PSNR, the PSNR values for all test images are summed and divided by the total number of images. With an Average PSNR of 30.5 dB, the generated images exhibit a high level of quality and maintain a close resemblance to the target images.

SSIM is another image quality metric that assesses structural similarity between images by comparing luminance, contrast, and structural aspects. SSIM ranges from -1 to 1, with values close to 1 indicating high structural similarity.

The SSIM formula is as follows:

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where

- μ_x and μ_y are the mean intensities of images x (target) and y (generated).
- σ_x^2 and σ_y^2 are the variances of images x and y.
- σ_{xy} is the covariance of images x and y.
- c_1 and c_2 are the constants to stabilize the division when the denominator is close to zero. [19]

With an Average SSIM of 0.87, the model achieves a high level of structural similarity, meaning the generated images maintain essential features and details that closely align with the target images.

These metrics indicate that the model retains a high degree of similarity to target images, confirming both high image quality and structural fidelity in the generated outputs. Fig. 4 demonstrates the different Image Generation models along with their PSNR and SSIM values for comparision.

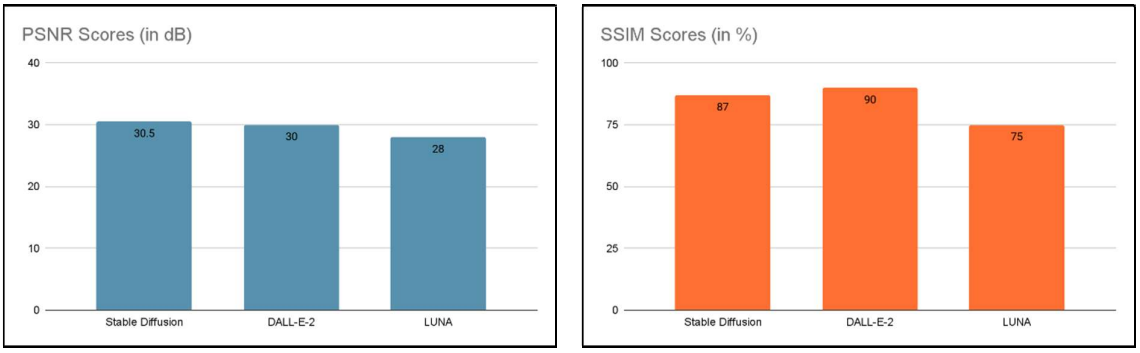

Fig 4. Bar Charts comparing the PSNR and SSIM scores for various models

V. Conclusion

This project marks a significant advancement in creative artificial intelligence, demonstrating a novel approach to generating visually inspired content from textual quotes.

The system harnesses the capabilities of LLama 3.2 for nuanced prompt generation, which interprets the semantic and emotional depth of quotes, in combination with Stable Diffusion's powerful image synthesis engine.

Together, these components enable the generation of images that not only capture the literal content of the quotes but also reflect an artistic interpretation that aligns with the text's essence.

Quantitative evaluation supports the effectiveness of the model. High-quality output is reflected in structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) metrics, where values of 0.87 and 30.5 dB, respectively, indicate clear, detailed images that closely match the intended design and visual fidelity. Additionally, the system achieves impressive performance in terms of image diversity and generation speed, allowing for a range of distinct styles and formats suited to various creative applications.

A comparative analysis highlights the model's competitiveness within the generative AI landscape.

When benchmarked against leading tools such as DALL-E, this system not only demonstrates comparable quality and variety but, in certain respects, surpasses them, particularly in areas of prompt interpretation and thematic coherence.

This robustness and versatility position it as a valuable asset in the suite of generative AI tools, particularly for applications in web design, digital storytelling, and other creative domains that require efficient and diverse visual outputs.

REFERENCES

[1] S. M, G. S, M. Reni, M. M, and R. M. S, "Image generation with Stable Diffusion AI," IJARCCE, vol. 12, no. 5, Apr. 2023, doi: 10.17148/ijarcce.2023.125106. http://dx.doi.org/10.17148/IJARCCE.2023.125106

[2] J. Oppenlaender, "A taxonomy of prompt modifiers for text-to-image generation," Behaviour and Information Technology, pp. 1–14, Nov. 2023, doi: 10.1080/0144929x.2023.2286532. https://doi.org/10.1080/0144929X.2023.2286532

[3] S. S. AlamA, J. N, M. F. A. B, and Veerasundari. R, "Stable Diffusion Text-Image generation," International Journal of Scientific Research in Engineering and Management, vol. 07, no. 02, Feb. 2023, doi: 10.55041/ijsrem17744.https://doi.org/10.55041/ijsrem17744

[4] X. Duan, S. Ma, H. Liu, and C. Jia, "PKU-AIGI-500K: A Neural Compression Benchmark And Model for AI-Generated Images," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 14, no. 2, pp. 172–184, Apr. 2024, doi: 10.1109/jetcas.2024.3385629.https://doi.org/10.1109/jetcas.2024.3385629

[5] D. Madhav, S. Nijai, U. Patel, and K. Champanerkar, "Question Generation from PDF using LangChain," IEEE, Feb. 2024, doi: 10.23919/indiacom61295.2024.10499105.https://doi.org/10.23919/INDIACom61295.2024.10499105

[6] Z. Chen, "Neural Language Models in Natural Language Processing," IEEE, vol. 12, pp. 521–524, Oct. 2023, doi: 10.1109/icdacai59742.2023.00104. https://doi.org/10.1109/ICDACAI59742.2023.00104

[7] P. Danenas and T. Skersys, "Exploring Natural Language Processing in Model-To-Model Transformations," IEEE Access, vol. 10, pp. 116942–116958, Jan. 2022, doi: 10.1109/access.2022.3219455.https://doi.org/10.1109/access.2022.3219455

[8] A. Ghosh and K. Deepa, "QueryMintAI: Multipurpose Multimodal Large Language Models for Personal Data," IEEE Access, p. 1, Jan. 2024, doi: 10.1109/access.2024.3468996.https://doi.org/10.1109/access.2024.3468996

[9] S. Na, M. Do, K. Yu, and J. Kim, "Realistic Image Generation from Text by Using BERT-Based Embedding," Electronics, vol. 11, no. 5, p. 764, Mar. 2022, doi: 10.3390/electronics11050764.https://doi.org/10.3390/electronics11050764

[10] M. Coeckelbergh, "The Work of Art in the Age of AI Image Generation," Journal of Human-Technology Relations, vol. 1, Jun. 2023, doi: 10.59490/jhtr.2023.1.7025.https://doi.org/10.59490/jhtr.2023.1.7025

[11] T. Chen, Y. Yuan, and B. Yin, "Application of Prompt Engineering in AIGC — Taking Stable Diffusion as an Example," IEEE, pp. 465–469, Jun. 2024, doi: 10.1109/mlise62164.2024.10674412.https://doi.org/10.1109/MLISE62164.2024.10674412

[12] A. Dhar, A. Datta, and S. Das, "Analysis on Enhancing Financial Decision-making Through Prompt Engineering," IEEE, Dec. 2023, doi: 10.1109/iementech60402.2023.10423447. https://doi.org/10.1109/IEMENTech60402.2023.10423447

[13] M. Cahyadi, M. Rafi, W. Shan, H. Lucky, and J. V. Moniaga, "Accuracy and Fidelity Comparison of Luna and DALL-E 2 Diffusion-Based Image Generation Systems," IEEE, Jul. 2023, doi: 10.1109/iaict59002.2023.10205695.https://doi.org/10.1109/IAICT59002.2023.10205695

[14] L. V. Nguyen, T. T. Nguyen, O. A. Dobre, and T. Q. Duong, "Leveraging Stable Diffusion with Context-Aware Prompts for Semantic Communication," IEEE, pp. 610–615, Sep. 2024, doi: 10.1109/mass62177.2024.00097. https://doi.org/10.1109/MASS62177.2024.00097

[15] P. Esser, S. Kulal, A. Blattmann, R. Entezari, J. Müller, H. Saini, Y. Levi, D. Lorenz, A. Sauer, F. Boesel, D. Podell, T. Dockhorn, Z. English, K. Lacey, A. Goodwin, Y. Marek, and R. Rombach, "Scaling Rectified Flow Transformers for High-Resolution Image Synthesis," arXiv preprint arXiv:2403.03206, Mar. 2024. [Online]. Available: [Accessed: Jul. 13, 2024].https://doi.org/10.48550/arXiv.2403.03206

[16] M. Elasri, O. Elharrouss, S. Al-Maadeed, and A. S. Al-Maadeed, "Image Generation: A Review," Neural Processing Letters, vol. 54, no. 5, pp. 3887–3938, Mar. 2022. [Online]. Available: [Accessed: Jul. 13, 2024] https://doi.org/10.1007/s11063-022-10777-x

[17] L. Yin, "A Review of Text-to-Image Synthesis Methods," IEEE, Apr. 2024, doi: 10.1109/cvidl62147.2024.10603609.https://doi.org/10.1109/CVIDL62147.2024.10603609

[18] S. Khan, "Image Generation Using Diffusion and Stable Diffusion Models," ResearchGate, May 2024, [Online]. https://www.researchgate.net/publication/380577836_Image_Generation_Using_Diffusion_and_Stable_Diffusion_Models

[19] "Introducing Llama 3.2." https://www.llama.com/ (accessed Oct. 29, 2024).

[20] "Introduction | LangChain." https://python.langchain.com/docs/introduction/

[21] "Get started with LangSmith | LangSmith." https://docs.smith.langchain.com/

[22] "Streamlit • A faster way to build and share data apps." https://streamlit.io/

[23] "Stable Diffusion 2-1 - a Hugging Face Space by stabilityai." https://huggingface.co/spaces/stabilityai/stable-diffusion

[24] L. Naimi, E. M. Bouziane, M. Manaouch, and A. Jakimi, "A new approach for automatic test case generation from use case diagram using LLMs and prompt engineering," IEEE, vol. 238, pp. 1–5, Jun. 2024, doi: 10.1109/iccsc62074.2024.10616548.

[25] B. Puranik, R. Inamdar, P. Ghule, S. Awlankar, and P. Ovhal, "Insights and Performance Evaluation of Stable Diffusion Models with OpenVINO on Intel Processor," IEEE, pp. 1408–1411, Jul. 2024, doi: 10.1109/aic61668.2024.10731090.

[26] S. Dehnie, T. Sencar, and N. Memon, “Digital Image Forensics for Identifying Computer Generated and Digital Camera Images,” IEEE, Oct. 2006, doi: 10.1109/icip.2006.312849.
[27] Z. Ji, W. Wang, B. Chen, and X. Han, “Text-to-Image Generation via Semi-Supervised Training,” IEEE, vol. 70, pp. 265–268, Dec. 2020, doi: 10.1109/vcip49819.2020.9301888.
[28] A. Kotiyal, P. G. J, G. P. M. S, R. M. Devadas, V. Hiremani, and P. Tangade, “Chat with PDF using LangChain model,” IEEE, pp. 1–4, Jul. 2024, doi:10.1109/icait61638.2024.10690817.
[29] Y. Sasaki, H. Washizaki, J. Li, D. Sander, N. Yoshioka, and Y. Fukazawa, “Systematic Literature review of prompt engineering Patterns in Software Engineering,” IEEE, pp. 670–675, Jul. 2024, doi: 10.1109/compsac61105.2024.00096.
[30] D. Park, G.-T. An, C. Kamyod, and C. G. Kim, “A Study on Performance Improvement of Prompt Engineering for Generative AI with a Large Language Model,” Journal of Web Engineering, pp. 1187–1206, Feb. 2024, doi: 10.13052/jwe1540-9589.2285.
[31] K. Fujihira and N. Horibe, “Multilingual sentiment analysis for web text based on word to word translation,” IEEE, vol. 15, pp. 74–79, Sep. 2020, doi: 10.1109/iiai-aai50415.2020.00025.