A co-creative system to assist users who want to draw.

Tingyan Lu s4226992
Zhiwei Ma s4437586
Yuanyuan Rong s4508580
Xijie Cao s4245016

1 Abstract

Recent advances in generative AI, especially in text-to-image synthesis, have opened exciting opportunities for human-AI collaboration in creative projects. However, making these technologies truly collaborative and user-friendly remains a challenge. This paper presents a co-creative system to help users generate and refine creative sketches. It combines generative algorithms, sentiment analysis, and user preferences, and the system tailors its outputs to align with the user's intent while fostering creativity and engagement. The system adapts and evolves on requests based on user feedback to the generative algorithm, creating a dynamic and interactive experience. Early evaluations show that the system can generate diverse, high-quality outputs matching user inputs, highlighting its potential to enhance human-AI co-creative processes.

2 Introduction

Human-AI (AI) collaboration has emerged as an important and growing field, particularly in art, music, and literature. Recent advances in generative AI models, such as systems based on stable diffusion and GPT, have expanded the possibilities of human-AI interaction, allowing systems to not only generate outputs autonomously but also engage in meaningful creative collaboration with usersElgammal et al. [2017] and Goodfellow et al. [2014]. This paper explores the development of a co-creative drawing assistant that integrates user preferences, sentiment analysis, and generative AI to produce personalized visual outputs.

The main aim of the project is to help users solve common creative challenges, such as finding inspiration or turning vague ideas into clear artistic creationsKantosalo and Toivonen [2020]. Many existing tools generate results with very little input from users, limiting interaction. Co-creative systems focus on teamwork, allowing people and AI to work together. Users can share their preferences, describe ideas, and choose styles, making the process more engaging and enjoyable. Sentiment analysis is a key feature of this system, as it adjusts visual output based on the emotional tone of user inputHussain et al. [2019]. For example, the system can match the mood of the user's input. It also uses a flexible design to combine different inputs, such as themes (e.g., landscape, portrait), styles (e.g., impressionism, surrealism), and written descriptions. Create high-quality personalized images and encourage user and system teamwork.

Using Gradio provides an easy-to-use interface to build a prototype system where users can input their preferences and receive generated outputAbid et al. [2019]. The system also uses Genetic Algorithms to adjust prompts based on feedback, making the process more interactive and creative.

This system shows promise in improving creative workflows. It helps users explore new artistic ideas by working together with AI. The following sections will explain how the system was built and tested and its contributions to making creativity more collaborative and accessible.

3 Methods

3.1 System Architecture

The system comprises multiple interconnected modules which work together to achieve the desired output. The workflow shown in figure 1 begins with user input, such as theme preferences and descriptive text, which are processed and optimized through dynamic adjustments and a genetic algorithm. The results generate a text prompt and input to a generative model such as Stable Diffusion. The output module displays the generated image alongside the textual prompt for user feedback and saves preferences for future interactions.

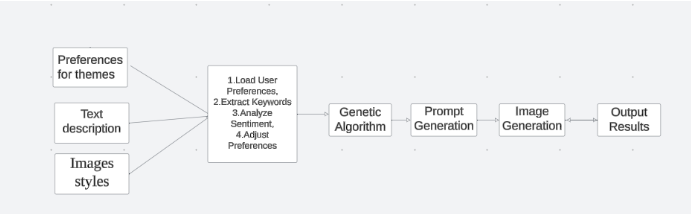


Figure 1: The Workflow Diagram of the System

3.2 Core Modules

Preference Storage and Loading This system dynamically manages user preferences and creates a personalized experience with multiple sessions. These preferences are different themes, such as "landscape," "portrait," "abstract," and so on. They play a key role in shaping the image generation process. By saving and adapting these preferences over time, the system ensures continuity and evolves to better align with the user's creative needs.

- Saving Preferences: This part is to store user-defined preferences for future sessions. We save the preferences as a Json file that can avoid the need for users to reset preferences each session, enhancing convenience and user satisfaction.

```
{
    "landscape": 3,
    "portrait": 5,
    "abstract": 2,
    "still life": 4,
    "fantasy": 1
}
```

- Loading Preferences: The purpose of this part is to retrieve saved preferences at the start of a session. The system attempts to load preferences from Json file. If it is unavailable, this code handles missing files by falling back to defaults. It can ensure the system that can operate, even on the first use or after file loss.

Sentiment Analysis The purpose of the Sentiment Analysis component in the system is to determine the emotional tone of the user's input text. This emotional understanding is then used to influence and customize the tone and style of the generated image. By identifying the emotional sentiment like joyful, somber and netural form input text, it brings a lot of advantages. The first one is that the system can tailor the color scheme, mood, and overall tone of the generated image to match the user's emotional state or intention. Then it also can move beyond literal interpretations of text to create more nuanced and expressive artwork that resonates emotionally with user.

The function categorizes the sentiment into three classes based on p:

$$S = \begin{cases} \text{Joyful}, & p > 0.5 \\ \text{Somber}, & p < -0.5 \\ \text{Neutral}, & -0.5 \leq p \leq 0.5 \end{cases}$$

- Joyful: Promotes the use of vibrant and bright colors in the image.
- Somber: Encourages darker, muted tones.
- Neutral: Balances vibrant and subdued elements for a moderate visual tone.

This module adds a layer of emotional intelligence to the system, enabling it to generate images that align not only with the literal content but also with the underlying sentiment of the user's input. It enhances user satisfaction by making the outputs more expressive and emotionally relevant.

Keyword Extraction The Keyword Extraction Module interprets and maps specific keywords from the user's textual description to predefined themes. It ensures the system can align user input with relevant categories for personalized image generation.

The first step is input pre-processing, including converting the input description to lowercase for case-insensitive, matching, and splitting the text into individual words. Then, using a predefined dictionary to map specific words to broader themes:

- Portrait: Keywords like ”earrings”, ”jewelry”, ”necklace”.
- Landscape: Keywords like ”mountain”, ”forest”, ”ocean”.
- Abstract: Keyword ”abstract”.
- Still Life: Keywords like ”fruit”, ”flowers”.
- Fantasy: Keywords like ”dragon”, ”castle”.

Finally, it can match the input with a dictionary, maps them to their corresponding themes, and returns a list of mapped themes based on the input description.

Dynamic Preference Adjustment The Preference Adjustment Module dynamically adjusts user-defined preferences based on the extracted keywords from the user's description. This ensures that the system gives more weight to themes explicitly mentioned by the user, enhancing personalization and relevance in image generation.

1. Input: A user preference is a dictionary storing the user's predefined preferences for various themes and a list of keywords derived from the user's description.
2. Process: Firstly, a copy of the user preferences is created to preserve the original data. Secondly, the module iterates through the keywords. If a keyword matches a theme in the user preferences dictionary, the corresponding preference score is incremented by 5.
3. Output: The updated preference dictionary is returned, prioritizing themes that match the user's keywords.

Genetic Algorithm The Genetic Algorithm is a core part of this project. It is designed to optimize theme selection that balances user preferences with contextual relevance. The algorithm can also adapt dynamically so that the mutation and crossover steps ensure diversity and adaptability. Last but not least, it can provide personalized results. By connecting the final theme combination with user input and preferences, the system generates better image prompts.

How the algorithm works is displayed in the following steps:

- Population Initialization: Randomly generate an initial population of theme combinations, where each individual is a subset of the themes.
- Fitness Evaluation: Calculate a preference score by summing the user's preference values for the themes. Calculating a keyword score is important by giving additional weight, such as 10 points for themes matching the description keywords. The total fitness is the sum of these scores.

- Selection: Select individuals for the next generation based on their fitness scores, with higher scores being more likely to be selected.

- Crossover: Combine the themes of two-parent individuals to produce a child, ensuring diversity while maintaining reasonable length.

- Mutation: Randomly modify an individual to introduce diversity, including adding a random new theme not already in the individual, removing a random theme from the individual, and swapping two themes within the individual.

- Iteration and Tracking the Best Solution: Repeat the evaluation, crossover and mutation process for specified number of generations and track the best-performing individual and its fitness score across generations.

Prompt Generation The Prompt Generation module is responsible for crafting personalized and descriptive text prompts by integrating user preferences, contextual keywords, and emotional tone. This module dynamically tailors prompts to reflect the user's specified themes, stylistic nuances, and mood, serving as the foundation for the image generation model to produce visually coherent and meaningful outputs.

The module prioritizes themes such as ”landscape” or ”portrait” based on user preferences. Detailed elements like ”mountain” for landscapes or ”earrings” for portraits are added if they match the user's description. Then, Adjectives like ”vivid,” ”elegant,” or ”realistic” are included to enhance the stylistic quality of the prompt. In addition, Sentiment analysis guides the tone of the prompt. For example, a ”joyful” sentiment produces vibrant, bright descriptions, while a ”somber” sentiment produces muted, melancholic tones. Finally, The generated prompt encapsulates user input, style preferences, and mood adjustments, providing a detailed and context-aware directive for the image generation model.

Image Generation The Image Generation step used the Stable Diffusion model to create a high-quality image based on the generated text prompt.

- Pipeline Initialization: The Stable Diffusion Pipeline is loaded using the ”runwayml/stable-diffusion-v1-5” model and is initialized with GPU acceleration (if available) to optimize processing speed and efficiency.

- Prompt Execution: The personalized prompt, which includes user-selected themes (e.g., landscape, portrait), specific details (e.g., mountains, jewelry), and emotional tone (e.g., joyful, somber), are input into the pipeline.

- Image Synthesis: The pipeline processes the prompt, analyzing its elements to produce a visual output that aligns closely with the described content, such as landscapes with towering mountains or portraits with intricate details.

- Output Delivery: The resulting image is extracted and returned, completing the process by providing a visual representation that reflects the user's preferences, extracted keywords, and emotional sentiment.

4 Results

The way for us to use this system is to select preferences scores in five aspects, input a description, and choose a style. Then the system generated a Generated Prompt that describes this picture and a suitable painting. The interactive page and some generated images are shown in figure 2. We test this system in the following four aspects.

Quality Assessment The assessment of quality is to evaluate how well the theme, description, and emotions. The aim is to determine whether the images align with the user's description and preferences. As a robust way, we used CLIP model, a text description, and an image path are specified. This model outputs the similarity score between the image and the text. Evaluating 50 samples from the tasks, we found that the similarity score is between 30% to 60% compared to the prompt and generated image.

Diversity Assessment The system generates ten images based on the same description. For images with the same theme, the styles are different. For images with the same style, only one aspect of the theme differs slightly.We try to generate similar pictures to do the test.

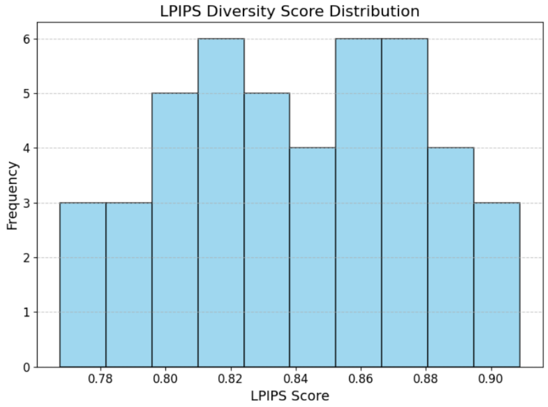
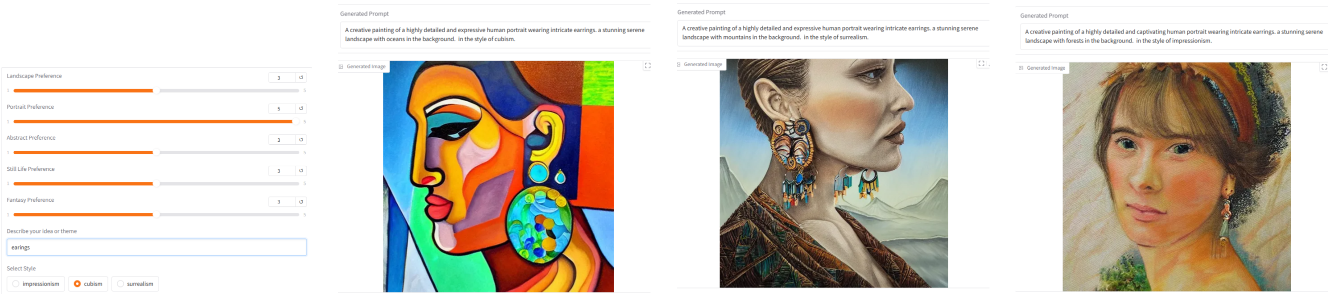


Figure 3: Evaluate diversity among generated images, Average diversity score (LPIPS): 0.8393

The distribution of LPIPS scores (ranging from 0.78 to 0.90) indicates a moderate level of perceptual diversity among the images analyzed. The scores are fairly evenly spread, with slight peaks around the ranges of 0.82 to 0.84 and 0.86 to 0.88, where the frequency of occurrence is highest. As shown in the figure3. It suggests that image pairs generally exhibit perceptual differences within a consistent range, without extreme outliers or clustering in the distribution. The results demonstrate a balance between diversity and similarity, reflecting adequate variability in the evaluated image set.



Landscape Preference 3

Portrait Preference 5

Abstract Preference 3

Still Life Preference 3

Fantasy Preference 3

Describe your idea or theme: earings

Select Style: ○ impressionism ● cubism ○ surrealism

(a) input

Generated Prompt: A creative painting of a highly detailed and expressive human portrait wearing intricate earrings. a stunning serene landscape with oceans in the background. in the style of cubism.

(b) result1

Generated Prompt: A creative painting of a highly detailed and expressive human portrait wearing intricate earrings. a stunning serene landscape with mountains in the background. in the style of surrealism.

(c) result2

Generated Prompt: A creative painting of a highly detailed and captivating human portrait wearing intricate earrings. a stunning serene landscape with forests in the background. in the style of impressionism.

(d) result3

Figure 2: Different styles of portraits. A is the input with kinds of functions like people's preferences, ideas, and styles. B is a Cubism portrait with earrings and a generated prompt. C is a surrealism portrait with earrings. The last one is an impressionism portrait with earrings and a generated prompt.

User feedback We invited 10 students of different nationalities and specializations to test our system. Their ages range from 24 to 31 years. Some learned arts-related hobbies when they were young. All of them have used AI-creative tools before. Six participants gave a kind of satisfaction rating of 4, three gave a 3, and one gave a 2. Some want more creative and interactive, but not templated. Seven participants reported that the generated images matched their input emotions and themes, two found them generally aligned, and one felt that the match was insufficient.

Efficiency Performance of program execution varies significantly on different computing platforms. When running in a CPU environment, the program takes at least 30 minutes to generate results; under GPU acceleration, the execution time is greatly reduced, and the same results can be generated in 10 seconds.

5 Discussion

The system generates diverse and creative output that aligns with user inputs. Although the CLIP evaluation result is not good enough, this may be caused by the creativity of the generated images. It is easier to describe the content for normal ones, but for the novel ones, it always has some unique points; as a creative tool, this is what we want. Feedback indicates that most of the users felt that the images generally captured their emotions and themes, with 60% rating their experience as good. However, occasional mismatches suggest improvements in sentiment analysis and keyword extraction to better interpret user intent and offer greater flexibility.

Diversity metrics reflect balanced variability, avoiding repetitive outputs. However, exploring more significant stylistic variations could further improve the results. While the system's efficiency is commendable, its heavy reliance on GPU resources poses a challenge for users with less powerful hardware, limiting accessibility. In addition, while the Stable Diffusion model generates highquality images, the coherence between the generated images and the input descriptions might vary depending on the complexity of the prompt.

The system effectively supports users experiencing creative blockages, offering meaningful inspiration. Our future efforts will focus on refining emotion analysis, expanding stylistic diversity, and optimizing resource requirements. We aim to make the system more accessible and adaptable, ultimately serving as a valuable tool for everyone who likes to try new staff.

6 Conclusion

This project demonstrates a co-creative system that integrates user preferences, sentiment analysis, and generative AI to create personalized images. By combining genetic algorithms and Stable Diffusion, the system effectively adapts to user input and generates high-quality emotionally aligned visuals.

Bibliography

A. Abid, M. Farooqi, and J. Zou. Gradio: Hassle-free sharing and testing of machine learning models in the wild. *arXiv preprint arXiv:1906.02569*, 2019.

A. Elgammal, B. Liu, M. Elhoseiny, and M. Mazzone. Can: Creative adversarial networks, generating "art" by learning about styles and deviating from style norms. *arXiv preprint arXiv:1706.07068*, 2017.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, volume 27, 2014.

S. Hussain, A. Athar, and M. Arif. Sentiment analysis of social media content using machine learning. *International Journal of Advanced Computer Science and Applications*, 10(3):1–10, 2019.

A. Kantosalo and H. Toivonen. Human–computer collaboration in creative tasks: A study of the co-creative drawing agent. *Journal of Computational Creativity*, 3(1):1–21, 2020.