2024 IEEE International Conference on Information Reuse and Integration for Data Science (IRI)

Prompirit: Automatic Prompt Engineering Assistance for Improving AI-Generated Art Reflecting User Emotion

Hannah Kim, Hyun Lee, Sunyu Pang, and Uran Oh*
Department of Computer Science and Engineering, *Ewha Womans University*
Seoul, Republic of Korea
{hannahk01, hyunini0408, syoub, uran.oh}@ewha.ac.kr

***Abstract*—Recently, text-to-image generative Artificial Intelligence (AI) models have demonstrated their ability to generate high-quality art with text prompts. However, generative AI is still incapable of creating images that precisely reflect emotion. We propose *Prompirit*, an automatic prompt engineering assistance for improving AI-generated art in terms of expressiveness of emotion and aesthetics. We explored various approaches to refine users' free-form text input by incorporating user emotion and style modifiers. Statistical analysis and user evaluation with 100 respondents showed that *Prompirit* significantly improved the alignment of image-emotion and the aesthetics of the generated image while precisely conveying the content of the original input text. Based on the results, we provide implications for creating affective images.**

***Index Terms*—Prompt Engineering, Text-to-Image Generative AI, Emotional and Art Technology**

Target Emotion: sadness			
Original	**LLaMA-2**	**RePrompt**	**Prompirit**
So I moved away from my parents not too long ago. It's been a big change. Out of state. I'm by myself now, so it's pretty quiet! I miss them a lot.	An illustration of a young adult sitting alone in an empty room, surrounded by boxes and unfamiliar furniture, looking wistful and nostalgic with a hint of sadness, longing for the warmth and comfort of home and family.	moved, parents, big, change, state, pretty, quiet, miss, sadness	sadness, moved away parents long ago, big change, state, pretty quiet, miss lot, away parents long, lugubriousness, sorrowfulness, unhappiness, Soft Pastels, Watercolor Brush Strokes, Hand-Drawn

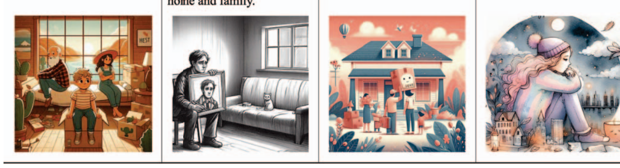

Fig. 1. Example images generated by DALL · E 3 given an original input text (*Original*) and three other prompt engineering methods: LLaMA-2 [9], RePrompt [8], and *Prompirit* (ours). The red and blue text refer to emotion labels and style modifiers, respectively.

I. Introduction

It is widely acknowledged that emotion is an important part of human life and its expression is crucial for one's health and well-being [1]. Unlike text-based literature, visual art can encapsulate nonverbal elements, making it a powerful tool for expressing emotion [2]. Recently, it has been explored in the domain of generative Artificial Intelligence (AI). By using text-to-image generative AI models (e.g., Stable Diffusion [3] and DALL · E [4]), anyone can now easily create images with simple text prompts.

Since emotion is one of the distinctions between humans and computers [5], the emotion in AI-generated artwork is considered as a significant aspect [6]. In prior works, for example, Galanos et al. [7] proposed a model applying generative adversarial networks to create artistic images expressing specific emotion. However, rather than receiving free-form text from a user, their images were created using a prompt with a semantic structure (i.e., A <emotion> <genre>) such as "a happy cityscape". On the other hand, Wang et al. [8] focused on reflecting accurate emotion expression in images by appending an emotion label to the adjusted number of parts of speech in the text prompt (e.g., "My best friend will be going to school in another country for 4 years." into "best, friend, going, school, country, current, advance, cold, sad"). Yet, it did not address visual elements such as styles and color themes. Moreover, it was not evaluated from an aesthetic point of view.

Inspired by these pioneering works, we propose *Prompirit*, an automatic prompt engineering assistance for reflecting emotion and matching styles when generating images. In addition to tokenized free-form text, *Prompirit* uses emotion label with it's synonyms and applies style modifiers that take emotion into account (e.g., 'dynamic composition' for *excitement*, 'muted neutrals' for *sadness*) as shown in Fig. 1.

We assessed our pipeline through objective and subjective evaluation. Four different prompt editing methods were set as conditions: the original input text (baseline), regenerated prompt with LLaMA-2 [9], RePrompt [8], and *Prompirit* (ours). For the performance evaluation, we generated 1,000 images under each condition to assess Image-Emotion Alignment (IEA), Image-Text Alignment (ITA), and Aesthetic Score [10]. In the user evaluation, we generated 80 images under each condition and recruited 100 participants to score them under the same aspects.

In both evaluations, *Prompirit* demonstrated superior performance, especially in terms of the expressiveness of emotion.

*Corresponding author

2835-5776/24/$31.00 ©2024 IEEE
DOI 10.1109/IRI62200.2024.00038

In particular, *Prompirit* ranked first in all three aspects, IEA, ITA, and Aesthetics, with subjective evaluation. Moreover, Pearson correlation for each pairs of the three aspects was calculated to explore discrepancies between objective and subjective outcomes. The result of the user evaluation metrics' correlation was higher than the performance evaluation, which suggests that people assess the three indicators comprehensively. In other words, *Prompirit*'s approach, which considers the context, sentiment, and style together, is effective.

Our work contributes:

- The proposal of an automatic prompt engineering assistance that enhances AI-generated images with an emphasis on emotion and matching styles.
- The empirical evaluation of *Prompirit* in comparisons to other prompt engineering methods in terms of both objective and subjective metrics. [1]

II. Related Work

A. Emotion in Art and Artificial Intelligence

In the field of art, the expression of emotion holds considerable importance [2]. Especially, the visual elements in images significantly contribute to evoking emotions [11], [12]. Since the ability to feel and share emotion is a significant difference between AI and human [5], emotion-based art creation with generative AI has been conducted, focusing on the relation between emotion and art [7], [13]. Accordingly, we aimed to continue this trend by incorporating the users' emotion and aligned visual elements to enhance prompts and overcome limitations in AI art generation.

B. Prompt Engineering for Text-to-Image Generative AI

Due to the open-ended nature of text, prompt engineering is essential in order to achieve the desired output with minimum trial-and-error. This is especially important in text-to-image generative AI, where the machine handles the entire creative process [14]. Consequently, there have been efforts such as defining prompt modifiers (e.g., 'style modifier' to enhance aesthetics, and 'repeating term' to strengthen the association) [15], aiming the improvement of image quality by better aligning with user requests. Moreover, prompt recommendation models have been developed to generate more aesthetically pleasing images, by extracting style modifiers from similar captions or prompts [16], [17]. However, these approaches were mainly focused on semantics and aesthetics in generated images, neglecting emotional expression. From this perspective, Wang et al. [8] introduced a method that involves adding an emotion label and adjusting the number of nouns, adjectives, and verbs in a text prompt to enhance emotional reflection. Nevertheless, it had semantic losses and did not integrate visual elements (e.g., styles, colors, value, and texture), nor was evaluated in terms of image aesthetics. Thus, we introduce an automatic prompt engineering assistance with detailed experiments on applying emotion labels and style modifiers to enhance both emotional expression and aesthetics.

[1] This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by Ewha Womans University Institutional Review Board under Application No. ewha-202404-0010-01.

III. Prompirit

The process of prompt input refinement with *Prompirit* for AI-based image generation involves two parts as shown in Fig. 2: (A) preprocess to make generative AI understand users' emotional and semantic intention better and (B) main process to emphasize emotional expression using emotion labels and style modifiers.

A. Preprocess

Before the main process, *Prompirit* labels the emotion, tokenizes the input text, and then extracts keyphrase for precise emotional and semantic reflection.

1) Datasets: The Empathetic Dialogues dataset [18] was chosen for its reliance on actual conversations. We used 'utterance', 'prompt', and 'context' in the dataset. The 32 emotion labels were classified into 8 labels, according to the knowledge graphs from SenticNet [19] and the vector similarity of word embeddings using BERT [20], spaCy [21], and Word2vec [22]. The eight labels include four negative emotions (i.e., *anger, disgust, fear, sadness*), which are considered universal and basic [23], and four positive emotions (i.e., *amusement, awe, contentment, excitement*) that are finer grained versions of happiness [24]. Lastly, GoEmotions dataset [25] was additionally used to address the data imbalance, with the same label classification as mentioned above.

2) Text Emotion Recognition: We used RoBERTa [26] to recognize emotion from the input text, which showed high performance [27]. With the datasets described in Section III-A1, our text emotion recognition model achieved the accuracy of 0.8973, and the hyper-parameters used are as follows: learning rate 2e-5, batch size 16, maximum length 128 with a total of 2 epochs.

3) Tokenization and Keyphrase Extraction: To deal with both colloquial and written forms as input, we explored the best text tokenization method with minimum semantic loss from the initial input. We tested (1) removing Natural Language Toolkit (NLTK) [28] stopwords, (2) extracting only nouns, verbs, and adjectives, and (3) appending the top-ranking keyphrase from KeyBERT [29], each and combined. We adopted ITA and BERTScore [30] to measure the semantic

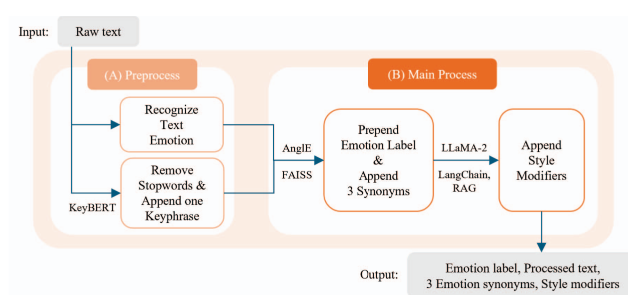

Fig. 2. Prompt engineering pipeline of *Prompirit*

loss and context reflection in the images. Among the presented methods, stopwords removal with keyphrase appending achieved the highest score.

B. Main Process

This part is designed to enhance the emotional expression and aesthetic quality of images consists of applying *Emotion Labels* and *Style Modifiers*.

1) Emotion Labels: We devised two main approaches to emphasize emotion in the input text based on the classified emotion label: prepending intensifiers [31] and appending synonyms [15].

- **Prepending Intensifiers (Intensifier).** We utilized intensifiers inspired by Carrillo et al. [31], such as "extremely" and "very". Each assigned a weight indicating its impact on emotion intensity. For instance, "very" amplifies the emotion by 75 percent. These intensifiers were added to prompts based on the probability distribution by the model in Section III-A2.
- **Appending synonyms (Synonym).** We collected synonyms from WordNet [32] and SenticNet, with duplicates and context-specific words (e.g., revolt, repel, four-star, first class) removed. The top three synonyms are selected from the corpus considering cosine similarity between the original text and synonyms (i.e., AnglE [33] and FAISS [34]) to avoid the performance decline as in [8].

To identify the best way to attach emotion labels, we compared four main approaches listed below and the examples of each condition are shown in Fig. 3a:

- *E1 (Baseline)* : Prepend the emotion label to the prompt.
- *E2 (Intensifier)*: *Baseline* + prepending an emotion intensifier corresponding to the emotion label probability distribution.
- *E3 (Synonym)* : *Baseline* + appending the top three synonyms that are most similar to the "original text".
- *E4 (Intensifier + Synonym)*: *Baseline* + prepending an emotion intensifier and appending the top three synonyms.

2) Style Modifiers: Considering the fact that emotional response to an image is influenced by the elements (e.g., color, texture, saturation, and brightness) [11] and principles of art (e.g., emphasis, harmony, variety, and movement) [12], we adopted style modifiers. First, inspired by PromptMagician model [16], which clusters image-prompt pair dataset and searches crucial keywords, we clustered an affective image-captions dataset to extract style modifiers [17]. Second, based on the idea of improving outputs through lexical constraints to generate prompt [35], we aimed to optimize prompt by using Retrieval-Augmented Generation (RAG) to provide a categorized list of style modifiers to be selected. In both approaches, we instructed LLaMA-2-13B [9] through LangChain [36] framework to extract and select these style modifiers.

- **Extracting modifiers from captions dataset (Caption).** We leveraged the ArtEmis 2.0 dataset [37], which consists of captions written by humans connecting emotions and masterpieces. 58,186 out of total 237,998 entries were randomly sampled to achieve an even distribution of emotion labels. Next, a combined form of caption, emotion label, and style modifiers from the data is converted into vectors using BERT [20]. Then we applied t-SNE algorithm [38] for feature dimension reduction, and conducted hierarchical clustering. Cluster sizes were limited to a maximum of 30 child nodes, avoiding large clusters that could obscure style modifiers with general ones like stopwords [16]. Finally, we identified the nearest cluster to the input text and appended the top three style modifiers with the highest Term Frequency scores [8].
- **Selecting modifiers with RAG (RAG).** RAG enhances LLM by retrieving relevant document chunks from an external knowledge base [39], thereby alleviating hallucinations [40]. We gave a structured prompt to LLaMA-2: the input text and emotion label as query, a style modifiers list as context, and an instruction to select three style modifiers [2]. The style modifiers list was collected from MidJourney-Styles-and-Keywords-Reference [41] and categorized by color (e.g., 'Vivid Red', 'Vibrant Colors'), dimensionality (e.g., '3D'), style (e.g., 'Film noir', 'Painting By Van Gogh'), light (e.g., 'Dim Lighting'), and perspective (e.g., 'Top-View'). The embeddings to represent text chunks in vector space were downloaded using sentence-transformers/all-MiniLM-L6-v2 [42], and ChromaDB [43] was employed as vector stores.

To sum up, we devised two different methods in this section and the corresponding examples are shown in Fig. 3b:

- *S1 (Caption)*: Extract three important style modifiers from the nearest cluster utilizing the ArtEmis 2.0 captions.
- *S2 (RAG)* : Select three suitable style modifiers from the style modifiers list provided using RAG.

C. Data Analysis

The images were generated with the data in the Section III-A. 1,000 prompts (125 prompts per emotion label) in validation and test datasets were chosen, considering the maximum processable token length of BERT. We employed Stable Diffusion 2.1 [3] for generating images, which does not require any API key, making it suitable for generating images multiple times to derive statistical analysis.

As for the analysis, images were evaluated with three metrics. We adopted two CLIP Score [44] based metrics, Image-Emotion Alignment (IEA) and Image-Text Alignment (ITA), to assess emotional expression and semantic match in images. CLIP Score calculates the cosine similarity of text-image embedding pairs (i.e., the emotion label and the image for IEA; the input text and the image for ITA). In addition, we used Aesthetic Score [10] to evaluate the aesthetic quality of the image.

We then applied a Linear Mixed Effects Regression (LMER) model to compare the 1,000 data points across the editing

[2]The details and full results of the experiments can be found in the supplementary material: https://sites.google.com/view/prompirit/

methods on the original data. The fixed main effect in the model was the prompt editing method, and the random effect was prompt ID, with IEA, ITA, and Aesthetic Score as dependent variables. We performed ANOVA for fixed and interaction effects and conducted post hoc tests [2].

As shown in the Fig. 4, *E3 + S2* performed the highest IEA and normalized score of a comprehensive comparison [2]. Thus, we curated *Prompirit*'s pipeline rubric as Fig. 2.

IV. Performance Evaluation

To identify the best approach for refining the original input prompt, we compared *Prompirit* against three other prompt engineering methods: the original text, recreated prompt using the LLaMA-2 [9], and RePrompt [8]. For LLaMA-2, we revised the instruction from prior work [45] to further emphasize the expression of emotion: "Use your imagination to add relevant descriptions to improve the *emotional expression*, beauty, and authenticity of the final image." [2].

The evaluation was conducted under the same experimental settings as described Section III-C. Fig. 5 shows the result of fitting the LMER model.

In terms of IEA, an indication of how well an image and a corresponding emotion is aligned, *Prompirit* has the highest performance as shown in Table I. This suggests that our approach to generate affective image outperforms LLaMA-2 and RePrompt which were also designed to emphasize emotion. Regarding ITA, we observed that the original text scored the highest. Since the CLIP score reflects the semantic consistency between text and image, it is expected that prompts without modifications from the original text can naturally receive the high score in terms of ITA [45]. However, despite adding considerably more words than RePrompt, *Prompirit* exhibited a comparable ITA to RePrompt. This indicates that *Prompirit* maintained the semantic intention of the original text. For Aesthetic Score, while original text and RePrompt which neglected visual aspects received low scores, *Prompirit* and LLaMA-2 were considered highly aesthetic. This demonstrates that appending style modifiers is an important factor for improving the aesthetic appeal of images.

Fig. 3. (a) Examples of prepending and appending emotion labels to the prompt. The red and yellow text refer to the emotion label and its intensifier, respectively. The orange text indicates the top three emotion synonyms with high similarity to the original sentence. (b) Examples of appending style modifiers to the prompt. The blue text represents the style modifiers.

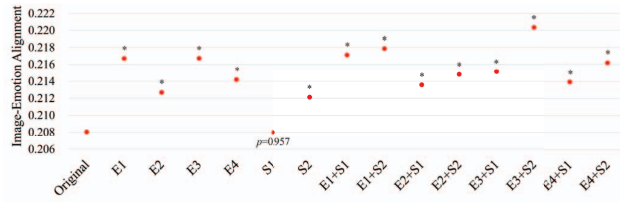

Fig. 4. Results of fitting the LMER model on IEA of proposed prompt engineering factors. '*' indicates significant difference at $p<.0001$ compared to the *Original*.

TABLE I
Results of IEA, ITA, and Aesthetic Score on four different prompt engineering methods in performance evaluation. The average score was calculated with all scores normalized into [0,1].

Method	IEA	ITA	Aesthetic Score	Avg. Score
Original	0.208	**0.245**	4.921	0.33
LLaMA-2	0.214	0.223	**6.053**	0.49
RePrompt	0.212	0.231	5.185	0.30
Prompirit	**0.220**	0.230	5.445	**0.59**

V. User Evaluation

In addition to the evaluation of objective performance, we conducted a subjective evaluation. We conducted an online survey using Google Forms where participants were asked to evaluate the images generated under four different prompt engineering methods as shown in Fig. 1 in terms of IEA, ITA, and Aesthetics on the 7-point scale. This task consisted of total eight sessions (one session per emotion label), which were randomly selected out of a total 80 session database (10 sessions per emotion label). In the case of the user study, we employed DALL · E 3 [4] to automatically block the deceptive and harmful content.

A. Participants

Participant recruitment was done through social networking services and online communities for university students. A total of 100 participants completed the survey, consisting of 62 females and 38 males. The age distribution was as follows: 54 participants were between 18 and 24 years old, 38 were between 25 and 34 years old, and 7 were 35 years or older.

B. Findings

Results were analyzed in the same way as in Section III-C with additional analysis on Pearson correlation coefficient. As shown in Table II and Fig. 6, *Prompirit* showed the highest mean and the smallest standard deviation in all three indicators, exhibiting a modest difference from the objective evaluation in Section IV. RePrompt's high p-value in Fig 6c suggested no significant difference from the *Original*, attributable to not considering visual features unlike *Prompirit* and LLaMA-2.

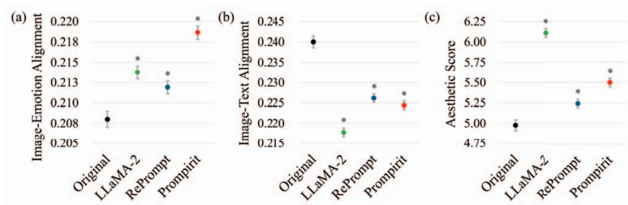

Fig. 5. Results of fitting the LMER model on IEA, ITA, and Aesthetic Score of four different prompt engineering methods in performance evaluation. '*' indicates significant difference at p<.0001 compared to the *Original*.

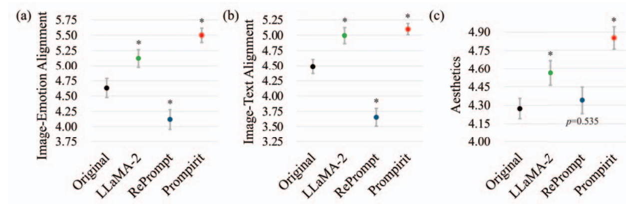

Fig. 6. Results of fitting the LMER model on IEA, ITA, and Aesthetics of four different prompt engineering methods in user evaluation. '*' indicates significant difference at p<.0001 compared to the *Original*.

TABLE II
AVERAGE AND STANDARD DEVIATION OF USER EVALUATION ON FOUR DIFFERENT PROMPT ENGINEERING METHODS (p<.0001).

Method	IEA	ITA	Aesthetics
Original	μ=4.11, σ=1.91	μ=4.51, σ=1.91	μ=4.72, σ=1.69
LLaMA-2	μ=4.90, σ=1.85	μ=5.00, σ=1.88	μ=4.76, σ=1.68
RePrompt	μ=4.04, σ=1.91	μ=3.66, σ=1.90	μ=4.40, σ=1.79
Prompirit	**μ=5.13, σ=1.77**	**μ=5.11, σ=1.76**	**μ=5.17, σ=1.61**

TABLE III
PEARSON CORRELATION COEFFICIENT RESULTS BETWEEN THREE INDICATORS FOR EACH EVALUATION METHOD. P-VALUES INDICATE SIGNIFICANCE IF NOT SPECIFIED. '*' STATES p<.0001.

Evaluation	IEA-ITA	IEA-Aesthetics	Aesthetics-ITA
Performance evaluation	0.070*	-0.039	0.048
User evaluation	**0.522***	**0.359***	**0.415***

In IEA, *Prompirit* received the highest score in both subjective and objective evaluations. On the other hand, ITA and Aesthetics exhibited different trends. To comprehend the cause for the discrepancy in results with Section IV, we examined the Pearson r correlation between each metric. In the Table III, the results suggested that the indicators in user evaluation were more strongly correlated with each other.

VI. DISCUSSION

A. Interpretation of Evaluation Results

Prompirit led IEA in both Section IV and V, proving its superiority in expressing emotions through images. Meanwhile, ITA showed the most opposite trend between the two sections as shown in Fig. 5b and Fig. 6b. Since *Prompirit* and LLaMA-2 have relatively large text variations, the CLIP-based ITA would have inevitably decreased as the input text was modified [45]. This indicates that CLIP does not fully represent human perception [8]. In Aesthetics, *Prompirit* received the highest ratings by a significant margin, particularly in user evaluation. The use of style modifiers from a verified database likely reduced hallucinations in LLM, and contributed to enhancing the visual preferences among humans [16], [17]. The improved correlation coefficient detailed in Table III indicates that it is difficult for humans to evaluate the three indicators completely separately. In other words, considering the emotional expression, contextual grasp, and aesthetics together can lead to overall satisfaction with the image. This further suggests that *Prompirit* received the highest scores in Section V because it comprehensively considered all three aspects, with particular emphasis on emotion.

B. Potential Applications

Our approach accepts a wide range of text as input and expresses emotion via visual elements, demonstrating its potential for application across various fields. For instance, since *Prompirit* takes colloquial inputs, it can assist psychological counseling by supporting clients to externalize and visually interpret their emotional experiences [46]. Moreover, our approach can be leveraged in marketing to capture the target audience's attention with enhanced message delivery that evokes specific emotion [47].

C. Limitation

While our pipeline focuses on eight basic emotions in text, real emotions are more complex. Expanding the range of supported emotions could lead to richer and more nuanced image outputs, better reflecting the complications of human emotion. In addition, future research is required to assess whether it is more beneficial to reflect emotions directly or to relieve negative emotions with positive images.

VII. CONCLUSION

We presented *Prompirit*, an automatic prompt refinement assistance that prepends an emotion label and appends its synonyms with additional style modifiers to improve emotional intensity and aesthetics of the generated images. The evaluation with 100 participants showed that *Prompirit* was preferred the most compared to other prompting approaches in terms of emotional expression, context consistency, and visual satisfaction. This work bridges between image generative AI and human through emotion and art technology, offering guidance for future efforts.

ACKNOWLEDGMENTS

This work was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-2020-0-01460) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

REFERENCES

[1] A. Scarantino, "How to do things with emotional expressions: The theory of affective pragmatics," *Psychological Inquiry*, vol. 28, no. 2-3, pp. 165–185, 2017.

[2] P. Achlioptas, M. Ovsjanikov, K. Haydarov, M. Elhoseiny, and L. J. Guibas, "Artemis: Affective language for visual art," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 569–11 579.

[3] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.

[4] OpenAI. DALL · E 3. https://openai.com/dall-e-3.

[5] S. Turkle, *The second self: Computers and the human spirit*. Mit Press, 2005.

[6] T. R. Demmer, C. Kühnapfel, J. Fingerhut, and M. Pelowski, "Does an emotional connection to art really require a human artist? Emotion and intentionality responses to AI-versus human-created art and impact on aesthetic experience," *Computers in Human Behavior*, vol. 148, p. 107875, 2023.

[7] T. Galanos, A. Liapis, and G. N. Yannakakis, "AffectGAN: affect-based generative art driven by semantics," in *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2021, pp. 01–07.

[8] Y. Wang, S. Shen, and B. Y. Lim, "RePrompt: Automatic prompt editing to refine ai-generative art towards precise expressions," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–29.

[9] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale *et al.*, "Llama 2: Open foundation and fine-tuned chat models," 2023, arXiv:2307.09288.

[10] LAION. Aesthetic Predictor. https://github.com/LAION-AI/aesthetic-predictor.

[11] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 83–92.

[12] S. Zhao, Y. Gao, X. Jiang, H. Yao, T.-S. Chua, and X. Sun, "Exploring principles-of-art features for image emotion recognition," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 47–56.

[13] D. Alvarez-Melis and J. Amores, "The emotional GAN: Priming adversarial generation of art with emotion," in *2017 NeurIPS Machine Learning for Creativity and Design Workshop*, 2017.

[14] S. Brade, B. Wang, M. Sousa, S. Oore, and T. Grossman, "Promptify: Text-to-image generation through interactive prompt exploration with large language models," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023, pp. 1–14.

[15] J. Oppenlaender, "A taxonomy of prompt modifiers for text-to-image generation," *Behaviour & Information Technology*, pp. 1–14, 2023.

[16] Y. Feng, X. Wang, K. K. Wong, S. Wang, Y. Lu, M. Zhu, B. Wang, and W. Chen, "Promptmagician: Interactive prompt engineering for text-to-image creation," *IEEE Transactions on Visualization and Computer Graphics*, 2023.

[17] Y. Hao, Z. Chi, L. Dong, and F. Wei, "Optimizing prompts for text-to-image generation," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[18] H. Rashkin, E. M. Smith, M. Li, and Y.-L. Boureau, "Towards empathetic open-domain conversation models: A new benchmark and dataset," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2019, pp. 5370–5381.

[19] SenticNet: A semantic and affective resource for commonsense reasoning. https://sentic.net/.

[20] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, 2019, pp. 4171–4186.

[21] Explosion. spaCy: Industrial-strength natural language processing in python. https://spacy.io/.

[22] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, arXiv:1301.3781.

[23] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[24] E. Diener, C. Napa Scollon, and R. E. Lucas, "The evolving concept of subjective well-being: The multifaceted nature of happiness," *Assessing well-being: The collected works of Ed Diener*, pp. 67–100, 2009.

[25] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A Dataset of Fine-Grained Emotions," in *58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

[26] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," 2019, arXiv:1907.11692.

[27] A. F. Adoma, N.-M. Henry, and W. Chen, "Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition," in *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. IEEE, 2020, pp. 117–121.

[28] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc., 2009.

[29] M. Grootendorst. (2020) KeyBERT: Minimal keyword extraction with BERT. https://doi.org/10.5281/zenodo.4461265.

[30] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTscore: Evaluating text generation with BERT," in *International Conference on Learning Representations*, 2020.

[31] J. Carrillo-de Albornoz and L. Plaza, "An emotion-based model of negation, intensifiers, and modality for polarity and intensity classification," *Journal of the American Society for Information Science and Technology*, vol. 64, no. 8, pp. 1618–1633, 2013.

[32] Princeton University. WordNet. https://wordnet.princeton.edu/.

[33] X. Li and J. Li, "Angle-optimized text embeddings," 2023, arXiv:2309.12871.

[34] M. Douze, A. Guzhva, C. Deng, J. Johnson, G. Szilvasy, P.-E. Mazaré, M. Lomeli, L. Hosseini, and H. Jégou, "The faiss library," 2024.

[35] S. Rosenman, V. Lal, and P. Howard, "NeuroPrompts: An adaptive framework to optimize prompts for text-to-image generation," in *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations.*, 2024.

[36] LangChain. https://www.langchain.com/langchain.

[37] Y. Mohamed, F. F. Khan, K. Haydarov, and M. Elhoseiny, "It is Okay to Not Be Okay: Overcoming Emotional Bias in Affective Image Captioning by Contrastive Data Collection," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 21 231–21 240.

[38] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[39] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, and H. Wang, "Retrieval-augmented generation for large language models: A survey," 2023, arXiv:2312.10997.

[40] O. Topsakal and T. C. Akinci, "Creating large language model applications utilizing langchain: A primer on developing llm apps fast," in *International Conference on Applied Engineering and Natural Sciences*, vol. 1, no. 1, 2023, pp. 1050–1056.

[41] https://github.com/willwulfken/MidJourney-Styles-and-Keywords-Reference.

[42] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2.

[43] Chroma. https://www.trychroma.com/.

[44] J. Hessel, A. Holtzman, M. Forbes, R. L. Bras, and Y. Choi, "Clipscore: A reference-free evaluation metric for image captioning," *arXiv preprint arXiv:2104.08718*, 2021.

[45] T. Cao, C. Wang, B. Liu, Z. Wu, J. Zhu, and J. Huang, "BeautifulPrompt: Towards automatic prompt engineering for text-to-image synthesis," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*. Association for Computational Linguistics, 2023, pp. 1–11.

[46] M. M. Ginicola, C. Smith, and J. Trzaska, "Counseling through images: Using photography to guide the counseling process and achieve treatment goals," *Journal of Creativity in Mental Health*, vol. 7, no. 4, pp. 310–329, 2012.

[47] C. Achar, J. So, N. Agrawal, and A. Duhachek, "What we feel and why we buy: the influence of emotions on consumer decision-making," *Current Opinion in Psychology*, vol. 10, pp. 166–170, 2016.