Proceedings of the 3rd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/138/2025.21387

Judging the Reasons for Generating Artificial Intelligence Paintings in Human-Computer Interaction: Topic Analysis Based on Natural Language Processing

Jinhua Yang[1,a], Tianyue Niu[2,b,*], Huali Long[3,c]

[1]School of Humanities, Tongji University, Shanghai, 200092, China
[2]Xiamen Academy of Arts and Design, Fuzhou University, Xiamen, 361021, China
[3]Faculty of Applied Sciences, Macao Polytechnic University, Macao, 999078, Macao
a. yeohgenghwa@outlook.com, b. t24095@fzu.edu.cn, c. longhuali.mpu@gmail.com

Abstract: Studying the interaction between people and generative AI art has become a popular direction in interdisciplinary studies: this study explores the key mechanisms in people's cognitive responses when identifying generative AI paintings through natural language processing (NLP) techniques. By analyzing the textual feedback of 17 Chinese college students with art appreciation backgrounds, we adopt topic modeling and sentiment analysis to explain their perceptual criteria and emotional responses. This study reveals two key contributions: first, it finds the "paradoxical authenticity" of AI-generated artworks in the eyes of students. Second, it reveals that people tend to reinterpret human imperfections ,such as brushstroke tremors and color imbalances, as signs of artistic authenticity. These findings promote the discussion of creative coexistence between humans and AI and provide empirical insights into human-computer interaction and dynamic evaluation systems for generative art.

Keywords: Generating AI Paintings, Topic Analysis, Human-Computer Interaction, Natural Language Processing

1. Introduction

Artificial Intelligence (AI) has demonstrated constructive capabilities in a variety of fields[1-3] and has proven its ability to produce outputs that are traditionally considered creative.It is possible to enhance the final prediction or generation results by using multiple modalities or models[4, 5], especially with the evolution of large language models[6, 7], It has made great contributions to the development of generative artificial intelligence art. . Text-to-image generation AI including Midjourney, Stable Diffusion[8], Dall-E, is an AI system that can automates human tasks to generate high-quality digital artworks. While reshaping the creative paradigm, generative AI art has also triggered in-depth public debate on the nature of art, authorship, and aesthetic value. The emergence of AI-generated art, especially AI-generated paintings, has sparked many discussions, including on copyright ethics[9]. As generative AI art has advanced, scholars have begun to study how factors such as text quality and the use of AI influence audience perceptions[8].

However, few studies have discussed people's immediate reactions to this new technology, especially from the perspective of directly reflecting people's creation of discriminative spanning tree AI paintings. Although some studies have explored people’s attitudes toward generative AI paintings

DOI: 10.54254/2755-2721/138/2025.21387

on specific platforms, the progress in evaluating and analyzing professional attitudes toward generative AI paintings is still lagging behind[10]. In this paper, we explore the interview process of identifying generative AI paintings using natural language processing techniques [11] to understand the public's thought process and experience with generative AI identification. Our findings are based primarily on primary interview data. This approach provides us with a deeper, less filtered understanding of the topics of focus.

2. Methodology

2.1. Experiment Overview

This study aims to explore the public's judgment process regarding generative AI paintings. In the generation phase, the research utilized two advanced generative AI tools, Midjourney and DALL-E, to create paintings with similar styles based on the prompts "Post-Impressionism" and "Van Gogh." During the tuning process, the research team optimized the generation parameters multiple times to ensure that the AI-generated works closely resemble authentic Post-Impressionist paintings in terms of color usage, brushstroke expression, and overall composition. To evaluate the discernment effectiveness between AI-generated works and real paintings, the study invited 17 university students from various Chinese higher education institutions to participate as subjects. These participants possessed a certain level of art appreciation but had not received professional art training, ensuring their ability to engage in logical reasoning and critical thinking. During the judgment process, participants were required to thoroughly describe the key criteria upon which they based their judgments. These criteria specifically included, but were not limited to, the naturalness of color usage, the intricacy of brushstrokes, the harmony of overall composition, and the depth of emotional expression. These text-based feedbacks were systematically recorded and subjected to content analysis for coding and categorization, aiming to identify the visual elements and perceptual standards that participants most focused on during the discernment process. Some data examples of the article are shown in Table 1.

2.2. Experimental Procedure

This article uses the method of topic analysis and sentiment analysis of comments to extract the hidden topics in the data to understand the main discussion points of people's comments. First, this article uses CountVectorizer (bag of words model) to convert text into numerical analysis. Filter out common invalid words (such as "the", "is", "and") and keep only the top 100 most common words, then train the LDA model to model the topic (LDA - Latent Dirichlet Allocation[12]), with a fixed number of topics of 5. LDA discovers hidden topics in the text through probability distribution, and records the top 10 representative words extracted from each topic. Then, by analyzing the LDA topic distribution, 5 groups of topic keywords are generated to see the parts that users are most concerned about. Sentiment analysis (Sentiment can calculate the sentiment score of each comment to understand the user's attitude towards generative AI painting. This article uses TextBlob for sentiment analysis: TextBlob is a natural language processing tool that can calculate the sentiment polarity of text. This article uses positive numbers (0 to 1) to represent positive emotions and negative numbers (-1 to 0) to represent negative emotions.

3. Results

In this LDA topic modeling result, this study conducted a statistical analysis of the word weight distribution across five topics. The topic results of LDA topic modeling are shown in Table 2. The mean word weight for each topic is approximately 0.003195, indicating that after TF-IDF

DOI: 10.54254/2755-2721/138/2025.21387

normalization, the average weight values remain consistent across different topics, suggesting that word importance is balanced overall.The standard deviation ranges from 0.00142 to 0.00151, meaning that the word weight distribution within each topic is relatively stable, and the differences between topics are minimal. The maximum word weight ranges from 0.00689 to 0.00997, indicating that each topic contains some key words with significantly higher weights than the mean, highlighting their importance within their respective topics.Topic 1 may refer to the "unnatural" defects of generated images. This unnatural defect includes some inherent characteristics of AIalgorithms when processing images.

The keywords of Topic 1 include lacks, noise, uniform, clear, result, directly. In semantic clustering, lacks and uniform refer to the "lack of natural randomness" of generated images, which is consistent with the pattern repetition problem caused by training data bias or quantization error in generative models (such as GAN and diffusion models)[13]. The contradiction between noise and clear (such as "uniform noise but overly clear details") may reflect the overfitting of high-frequency details by the generative model [14].This to some extent shows that the "mode collapse"[15] of the generative model will cause the output to lack diversity or be full of "AIflavor", which can be sensitively captured by people. Topic 2 may point to the controversy about texture and generative traces. Its keywords include words that emphasize details, such as texture, generative, pattern, tremor, and details. For example, tremor may be related to the irregular physical shaking of human brushstrokes[16]. The third category of topics includes keywords such as ai, human, perfect, thickness, etc. This may be about the comparison between AI and human creation. For example, there is a comment in the data that says "The shape of the clouds is too textbook-like and perfect. Real artists will have more opportunities next time." This sentence contains words that reflect the relationship between clouds, human, perfect, etc. Topic 4 has obvious algorithm-driven stylization features. Its keywords include mathematics, fractal, idealized, and related. This indicates that some critics with professional computer graphics are rethinking the generation logic of related generative AI paintings from the perspective of algorithms .Topic 5 takes another perspective: it judges the subjective imbalance of color from the perspective of artistic perception rather than from the perspective of an algorithm. Keywords include art, error, heavier, imbalance, sky, etc. Heavier and imbalance describe the subjective perception of composition and are related to artistic aesthetic standards. Error (such as light and shadow errors) is regarded by some commentators as "evidence" of human creation because AI may be more "accurate".

Table 1: Topic modeling results

Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**Topic 5**
lacks	texture	clouds	effects	art
painting	like	ai	makes	human
noise	ve	definitely	ai	sky
uniform	color	artist	contain	left
clear	way	just	mathematics	error
pressure	generative	perfect	follows	doesn
overall	pattern	human	fractal	like
reminds	hard	created	fractals	heavier
result	tremor	unexpected	idealized	appearing
directly	details	thickness	graffiti	imbalance

DOI: 10.54254/2755-2721/138/2025.21387

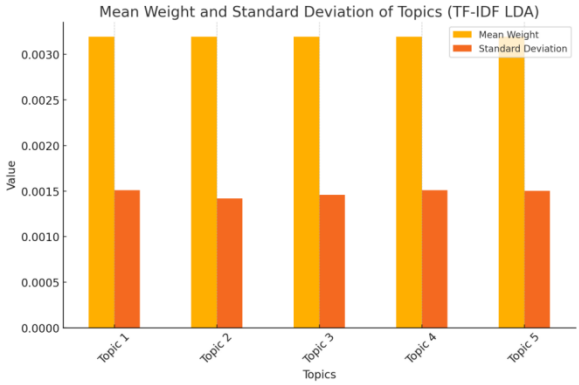

Figure 2: Figure of mean weight and standard deviation of topics

Meanwhile, this study also conducted sentiment analysis on 50 comments and calculated their basic statistical characteristics to reveal the overall sentiment tendency of the commentators towards the text content. The results of sentiment analysis are shown in Figure 3. The results show that the mean sentiment score of the comments is 0.095, which indicates that the overall sentiment tends to be neutral to positive. In addition, the standard deviation is 0.181, indicating that there is a certain volatility in the sentiment score. From the maximum value (Max = 0.40) and the minimum value (Min = -0.20), the most negative comment score is -0.20, while the most positive comment score reaches 0.40, indicating that some comments have a more obvious sentiment tendency. However, most comments are concentrated in a more neutral range, which is further verified in the quartile analysis (Quantiles). Specifically, the first quartile (Q1 = -0.028) shows that 25% of the comments have a slightly negative sentiment score, while the third quartile (Q3 = 0.20) shows that 75% of the comments have a sentiment score of no more than 0.20, further confirming the overall neutral sentiment distribution of the comments.

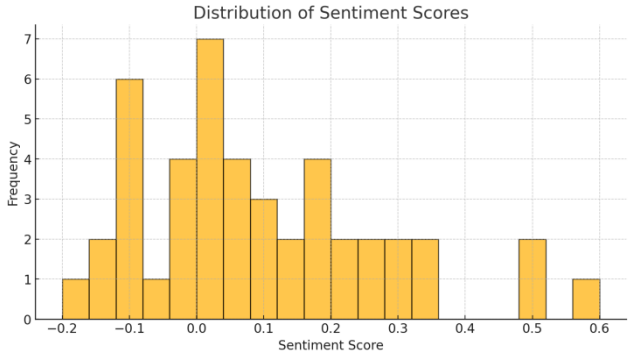

Figure 3: Sentiment analysis histogram

4. Discussion

Our study tried to reveal the cognitive relationship between humans and generative AI in the field of art through topic modeling and sentiment analysis,. Unlike traditional studies on technology acceptance, which focus on functional evaluation, our findings present some paradoxes of the "artistic authenticity crisis." First, on a technical level, certain details in AI-generated images such as well-executed lighting, ironically become the very fissures that expose their non-human nature. Second, on an aesthetic level, participants' persistent pursuit of "imperfection" (e.g., brushstroke tremors, color imbalance) has constructed a new paradigm of artistic appreciation, which elevating human creative flaws into markers of artistic authenticity. On the other hand, the study's discovery of "emotional

DOI: 10.54254/2755-2721/138/2025.21387

ambivalence" (as reflected in the average sentiment score of 0.095) suggests a deeper cultural anxiety. When commentators dismiss AI artworks as being "too textbook-perfect," they may, in fact, be defending the value of humanity's "necessary imperfection"—those creative traces shaped by physiological limitations and emotional fluctuations, which have already become digital surrogates for "proof of presence" in postmodern art philosophy. This collective unconscious cognitive resistance may constitute an important standard by which humans perceive and evaluate creative AI. From the perspective of art education, this study highlights new directions for cognitive training. The current art education system is built on a binary division between "creation" and "appreciation." However, in the AI era, artistic literacy may require the development of "three-dimensional perspective" skills: the ability to understand the psychological schemas behind human creativity, to grasp the technical logic of algorithmic generation, and to cultivate meta-cognitive skills for critical reflection at the intersection of both domains. Notably, the cross-modal perception exhibited by non-professional participants (such as associating color imbalance with dataset bias in algorithms) suggests that general art education should strengthen its modules on technological aesthetics cognition.

This study also has some limitations. The cultural differences may influence aesthetic perception. For instance, audiences accustomed to the "semblance and non-semblance" aesthetic principles in traditional Eastern art may develop cognitive patterns toward AI paintings that differ from Western samples. Second, our experiment lacks a temporal dimension, as AI generation technology iterates and public exposure increases, the current cognitive models identified may undergo a paradigm shift over time. Lastly, the neural mechanism remains a black box. Our current linguistic analysis does not reveal subconscious cognitive conflicts. Future research should integrate eye-tracking and neuroimaging techniques to construct a multimodal analysis framework.

5. Conclusion

In this work, we propose a hybrid recommendation model based on the MOE framework combined with a large language model, aiming to improve the accuracy and practicality of personalized recommendation systems in the health field. Experimental results show that the hybrid model outperforms the baseline model using only MOE or LLM in multiple key indicators, showing its potential in processing multimodal data, improving recommendation accuracy and personalization effects. Some issues that need to be further addressed were also found in the experiment. The improvement effect of image data in personalized recommendation systems is relatively limited, especially in the cold start problem. In addition, the image reclassification problem also affects the accuracy of the recommendation results when the image is low-quality or the appearance of the item changes, resulting in the system performance failing to reach the optimal level. This study provides valuable insights for the development of efficient and scalable recommendation systems and opens up new directions for the promotion of personalized recommendation technology in real-world applications, especially in the field of health management. Future research can further optimize the processing methods of image data and explore how to effectively integrate different types of data to improve the application effect of the system in various scenarios.

References

[1] *Luo, Y. and Z. Wang. Feature Mining Algorithm for Student Academic Prediction Based on Interpretable Deep Neural Network. in 2024 12th International Conference on Information and Education Technology (ICIET). 2024. IEEE.*

[2] *Li, S. Harnessing multimodal data and mult-recall strategies for enhanced product recommendation in e-commerce. in 2024 4th International Conference on Computer Systems (ICCS). 2024. IEEE.*

DOI: 10.54254/2755-2721/138/2025.21387

[3] *Lu, J. Enhancing Chatbot User Satisfaction: A Machine Learning Approach Integrating Decision Tree, TF-IDF, and BERTopic. in 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS). 2024. IEEE.*

[4] *Lu, J. Optimizing e-commerce with multi-objective recommendations using ensemble learning. in 2024 4th International Conference on Computer Systems (ICCS). 2024. IEEE.*

[5] *Sun, Y., et al. A Multi-Objective Recommender System for Enhanced Consumer Behavior Prediction in E-Commerce. in 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS). 2024. IEEE.*

[6] *Li, S., et al. Strategic deductive reasoning in large language models: A dual-agent approach. in 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS). 2024. IEEE.*

[7] *Xu, J. and Y. Wang, Enhancing Healthcare Recommendation Systems with a Multimodal LLMs-based MOE Architecture. arXiv preprint arXiv:2412.11557, 2024.*

[8] *Bhatt, H., et al. An Extensive Analyis of AI-Generative Art Algorithms. in 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS). 2024. IEEE.*

[9] *Ünal, R. and A. Taylan, Who Owns This Artwork?: AI-Based Art Creation and Copyright Issues, in Understanding Generative AI in a Cultural Context: Artificial Myths and Human Realities. 2025, IGI Global Scientific Publishing. p. 383-398.*

[10] *Wang, J., et al., AI paintings vs. Human Paintings? Deciphering Public Interactions and Perceptions towards AI-Generated Paintings on TikTok. arXiv preprint arXiv:2409.11911, 2024.*

[11] *Xu, J., et al. Automated Scoring of Clinical Patient Notes using Advanced NLP and Pseudo Labeling. in 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA). 2023. IEEE.*

[12] *Luo, Y., P.C.-I. Pang, and S. Chang, Enhancing Exploratory Learning through Exploratory Search with the Emergence of Large Language Models. 2025.*

[13] *Rombach, R., et al. High-resolution image synthesis with latent diffusion models. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.*

[14] *Yang, P., Y. Lao, and P. Li. Robust watermarking for deep neural networks via bi-level optimization. in Proceedings of the IEEE/CVF international conference on computer vision. 2021.*

[15] *Zhang, K. On mode collapse in generative adversarial networks. in Artificial Neural Networks and Machine Learning–ICANN 2021: 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part II 30. 2021. Springer.*

[16] *Ma, C., et al., Automatic diagnosis of multi-task in essential tremor: Dynamic handwriting analysis using multi-modal fusion neural network. Future Generation Computer Systems, 2023. 145: p. 429-441.*