Journal of Humanities and Arts Perspectives

press

https://doi.org/10.70731/xn9v3p17

The Practice of AI-Generated Gongbi Painting: Comparative Analysis of Multi-Platform Tools and Stylistic Exploration

Mohan Kan [a,*]

[a] *Baekseok University, Cheonan 31065, South Korea*

KEYWORDS

Gongbi Painting;
AI-Generated Imagery;
Style Simulation;
Six Principles of Xie He

ABSTRACT

This study investigates the capabilities and limitations of artificial intelligence (AI) image generation platforms in simulating the stylistic features of traditional Chinese Gongbi painting. Focusing on four widely used platforms—leonardo.ai, Doubao, ChatGPT, and Artbreeder—the research conducts a comparative analysis across two key modalities: text-to-image and image-to-image generation. Using parameters such as composition, line quality, color treatment, and style fidelity, the study evaluates each platform's performance in reproducing the refined aesthetics of Gongbi painting. The classic Chinese art theory Six Principles of Xie He is employed as a critical framework, assessing the AI-generated images through six dimensions: spirit resonance (qiyun), structural brushwork (gufa), formal likeness, color application, compositional arrangement, and stylistic imitation. Findings reveal that while AI systems demonstrate competence in replicating visual structures and color harmonies—making them suitable for preliminary sketching and stylistic exploration—they struggle with the nuanced expression of brush rhythm, cultural semantics, and artistic intentionality that define traditional Chinese painting. The study concludes that although current AI platforms cannot replace human artists as primary creative agents, they serve as valuable tools for inspiration, stylistic experimentation, and augmenting traditional workflows. This research thus contributes to a deeper understanding of how AI can interact with and extend the practices of classical art in the contemporary technological context.

INTRODUCTION

In recent years, artificial intelligence (AI) technologies have witnessed rapid and continuous development, particularly achieving groundbreaking progress in fields such as deep learning, image recognition, and natural language processing. These advancements have brought unprecedented innovation to the field of image generation, significantly increasing AI's participation in artistic creation. From early applications in basic style transfer to current capabilities in generating highly aesthetic works based on textual prompts or visual references, AI image generation platforms are gradually becoming essential tools in contemporary visual practice.

* Corresponding author. E-mail address: 278245721@qq.com

Received 17 June 2025; Received in revised from 9 July 2025; Accepted 22 July 2025; Published online 30 August 2025.

Notably, in the area of stylistic simulation, AI is now capable of approximating classical Western styles such as Impressionism and Surrealism, and has also begun to engage with Eastern traditional painting styles that are deeply rooted in regional culture and specialized techniques—an emerging trend that has drawn increasing attention from both scholars and artists.

Among these, Chinese Gongbi painting stands as one of the most representative forms within the traditional Chinese painting system, with a long historical lineage and a highly rigorous technical framework. Since its maturation during the Tang Dynasty, Gongbi painting has been characterized by meticulously refined line work, multi-layered coloring techniques, and carefully structured compositions that emphasize symmetry and spatial rhythm. Unlike the expressive and interpretive qualities of Xieyi (freehand) painting, Gongbi demands both precise figurative representation and harmonious integration of brushwork, color, and compositional energy. It values the unity of form and spirit, requiring from the artist a high level of aesthetic judgment, observational acuity, and technical mastery. These qualities also present significant challenges for AI systems attempting to emulate this style.

At present, numerous AI image generation platforms have emerged, including DALL·E, Midjourney, Stable Diffusion, Leonardo.ai, Doubao (ByteDance), and Artbreeder, each employing distinct algorithmic architectures and generative models. However, they share a common goal: to convert abstract natural language descriptions into concrete visual imagery through deep generative models. Increasingly, these platforms demonstrate enhanced capabilities in text-image alignment and stylistic control, making them adept at reproducing specific visual styles and contextual content with notable vividness. Nevertheless, when faced with a highly stylized and culturally embedded art form like Gongbi painting, it remains unclear whether AI can transcend the limitations of semantic interpretation to accurately replicate technical nuances, reconstruct spatial logic, and convey the ink-based aesthetic spirit central to Chinese traditional art.

This study originates from a critical inquiry into this very issue, centering on the stylistic simulation of Gongbi painting through empirical testing and comparative analysis across multiple mainstream AI image generation platforms. Specifically, the research focuses on Leonardo.ai, Doubao, ChatGPT, and Artbreeder, evaluating their generative performance along two methodological axes: text-to-image (T2I) and image-to-image (I2I). The analysis emphasizes key evaluative criteria including compositional layout, brushstroke texture, color handling, and stylistic coherence.

Methodologically, the study involves the use of standardized prompts and visual references to examine each platform's semantic comprehension of "Gongbi painting" and its strategic approach to image generation. Furthermore, a selection of the author's own Gongbi-style artworks serves as comparative material to assess each platform's ability to preserve details and transfer stylistic attributes in I2I scenarios. To strengthen the cultural and theoretical foundation of the analysis, the study also incorporates the "Six Principles of Chinese Painting" (Xie He's Six Laws) as an aesthetic evaluation framework. This traditional art theory, formulated in the 6th century, allows for a nuanced examination of AI-generated images through the lenses of vitality (qi yun), brushwork (gufa), formal structure (ying wu xiang xing), coloration (sui lei fu cai), composition (jing ying wei zhi), and imitation (chuan yi mo xie), thereby enriching both the analytical rigor and cultural depth of the study.

The ultimate aim of this research is not to reject the value of AI in artistic production, but rather to use the highly codified paradigm of Gongbi painting as a lens to assess the boundaries and potentials of AI-generated imagery when confronted with intricate, codified, and culturally significant visual languages. On one hand, the findings of this research help identify current limitations and challenges in AI's capacity for stylistic reproduction and artistic simulation. On the other hand, they also provide theoretical and practical insights for developing future frameworks of "human-machine co-creation" in the arts. Through this exploration, the study seeks to expand the dialogue between traditional art and contemporary technology, ultimately contributing to the emergence of new pathways for art generation in the age of algorithmic intervention.

PLATFORM TYPOLOGIES AND EXPERIMENTAL PATHWAYS

The current landscape of AI-driven image generation is marked by increasing platform diversification and refined functional specialization. Different platforms exhibit distinct characteristics in terms of algorithmic architecture, training data sources, interaction models, and stylistic tendencies, contributing to a complex and evolving ecosystem of tools. In order to ensure the representativeness and generalizability of this study, four widely used and influential platforms—Leonardo.ai, Doubao (ByteDance), ChatGPT, and Artbreeder—were selected as comparative samples. These platforms were chosen based on four criteria: image generation quality, stylistic control capability, linguistic adaptability, and user operability. The core objective of this section is to systematically evaluate each platform's ability to simulate the stylistic features of Chinese Gongbi painting.

As shown in **Table 1,** the four selected platforms exhibit clear distinctions in terms of functional priorities and operational logic, making them representative of two primary technological pathways: text-driven generation and image-driven generation. Leonardo.ai and ChatGPT, with their dual capabilities, demonstrate strong adaptability to both text and image inputs.

Table 1 | Fundamental attributes and functional classifications of the selected platforms

Platform	Type	Image/Text Generation Capabilities
Leonardo.ai	Image Generation Platform	Supports both Image-to-Image and Text-to-Image generation
Doubao (ByteDance)	Chinese-language Intelligent Platform	Supports Text-to-Image (limited Image-to-Image capabilities)
Chat GPT	Multimodal AI System	Supports both Image-to-Image and Text-to-Image generation
Artbreeder	Image Blending Platform	Does not support text prompts; only Image-to-Image mixing

Doubao stands out for its advanced understanding of Chinese semantic structures, while Artbreeder emphasizes image fusion through genetic algorithms, supporting gradual style evolution rather than precise semantic matching.

In the text-to-image generation experiment, the study focuses on critical dimensions such as prompt comprehension, compositional execution, stylistic alignment, and color rendering accuracy. This pathway provides an insight into AI's "creative" capabilities, wherein the system generates an image based directly on user-provided language prompts. Given the abstract and highly compressed nature of textual input, a platform's ability to accurately interpret and reconstruct Gongbi-related terminology and aesthetic intent becomes a core metric of its stylistic simulation capacity.

This study positions the four platforms within the two dimensions of text-to-image and image-to-image generation. Leonardo.ai, Doubao, and ChatGPT are examined primarily in the text-to-image pathway, focusing on their performance in linguistic interpretation, compositional arrangement, and stylistic fidelity. ChatGPT and Artbreeder are tested within the image-to-image framework, with attention paid to style transfer, detail preservation, and chromatic consistency.

By conducting a cross-pathway comparative analysis, this paper aims to reveal the platforms' adaptability and underlying mechanisms when confronted with the stylistic demands of Gongbi painting. This approach provides a technical and conceptual foundation for subsequent sections addressing semantic prompting and stylistic migration. Additionally, it offers actionable insights for artists seeking effective AI tools, enabling more efficient integration of AI into traditional artistic workflows.

SEMANTIC-DRIVEN VISUAL GENERATION: PROMPT ANALYSIS AND STRATEGY

In AI-based image generation systems, textual prompts function as the primary gateway to visual creation. This is especially critical in the simulation of highly codified and aesthetically rich traditional styles like Chinese Gongbi painting, where prompts serve not only to define content and composition but also to convey stylistic cues and activate cultural semantics. Therefore, the construction of informative and culturally coherent prompts is essential to facilitating the platform's ability to replicate a specific visual language.

Figure 1 | Author's original Gongbi painting.

This section introduces a set of manually designed semantic prompts aligned with the conceptual sketch illustrated in **Figure 1**. These prompts are tested across Leonardo.ai, Doubao, and ChatGPT to compare how each platform processes Gongbi-related semantics and responds in terms of stylistic fidelity.

Prompt Structure and Semantic Framework

The hypothetical artwork shown in Figure 1 is defined as follows:

This is a traditional Chinese Gongbi floral painting with a pale beige background. The composition is gentle and serene, characterized by soft lines and refined coloration. Broad blank spaces are left in the upper left third and lower right area of the canvas, forming a natural fan-shaped negative space to highlight the rhythm of vine growth and spatial density contrast. The main subject features two clusters of Campsis flowers. The primary cluster, in red-orange tones, is located at the upper right intersection of the nine-grid composition, with slightly faded edges. The secondary cluster, in lighter tones, is situated at the lower left intersection. Three symmetrical feather-like leaf groups are positioned above, surrounding the main flower. Another leaf group appears behind the secondary cluster, connected by fine stems and small leaves to create a layered effect. Overall, the composition exemplifies the ethereal beauty and natural structure typical of Gongbi painting.

The prompt follows a three-tier structure: **1)** primary compositional elements; **2)** stylistic emphasis; **3)** spatial rhythm and color guidance. This approach attempts to communicate key aspects of Gongbi practice, such as layout logic, visual hierarchy, coloring techniques, and the concept of intentional negative space, to the AI platform.

Leonardo.Ai: the "Illustrative" Deviation in Semantic Interpretation

Leonardo.ai, primarily trained on English-language data, exhibits certain limitations in prompt comprehension due to cultural and semantic gaps. Initial tests with prompts such as "Chinese Gongbi painting of flowers with elegant composition and fine lines..." produce images with discernible botanical layering, yet the overall style leans heavily toward digital illustration. The outputs display high color saturation and stark contrast, aligning more with concept art aesthetics than traditional painting.

This "illustrative" bias stems from the platform's misinterpretation of the term "Gongbi." Although "Gongbi" is increasingly recognized in global art discourse (Chen, 2022), Leonardo.ai often reduces it to a proxy for "delicate" or "floral illustration" in the absence of contextual cues. This undermines the painting's traditional stylistic features, such as linear precision, subdued transparency, and tranquil composition.

To address this issue, the prompt was revised to "Meticulous-style Chinese painting," explicitly detailing technical features rather than relying on ambiguous phonetic transcription. This modification led to some improvement in coloration, with occasional ink-wash-like gradients. However, the images retained a highly stylized digital texture, lacking the paper-air and ink-spirit qualities of hand-painted works. This highlights a broader issue of "cultural semantic rupture," where even syntactically sound prompts yield approximate visual equivalents when training data lacks sufficient cultural specificity.

Doubao: Semantic Precision and Structural Limitations in a Chinese Platform

In contrast, Doubao—trained primarily in Chinese—demonstrates superior comprehension of traditional art terminology. Upon receiving the same Chinese prompt, it successfully reproduces key compositional and chromatic elements such as vine movement, feathered foliage, and beige backgrounds. This suggests strong proficiency in parsing descriptive structures related to spatial orientation and color coordination.

Doubao's outputs resemble the "Mogu" style of Chinese painting, wherein linework is minimized and shapes are formed through layered shading. This aesthetic aligns well with neural networks' preference for blurred forms, and the results show decent atmospheric coherence. Particularly, Doubao exhibits superior handling of tonal harmony and spatial fusion, closely approximating the visual ethos of Gongbi painting.

However, Doubao currently lacks multi-turn interaction capability, meaning users cannot iteratively refine prompts to direct the generation process. Each image is treated as an isolated operation, limiting creative flexibility. This "single-turn decision model" poses a bottleneck in complex creative workflows that rely on iterative adjustment, thus constraining its application in advanced art generation scenarios.

ChatGPT: Potential in Multi-Turn Interaction and the Paradox of Semantic Decay

Unlike the above platforms, ChatGPT enables dynamic multi-turn interactions, maintaining user intent across multiple prompt iterations. For instance, after an initial prompt, users can follow up with refinements such as "enhance the curvature of the vines," "adjust background to a softer beige," or "increase leaf symmetry." ChatGPT accommodates these adjustments based on prior outputs, enhancing both control and stylistic granularity.

This cooperative capability allows for a more dialogic creative process, wherein users and system engage in iterative negotiation. However, a key drawback emerges in the form of "semantic decay": the platform's response to early prompts weakens over time, with newer instructions potentially overriding prior directives. This is especially problematic in intricate compositions or when balancing multiple stylistic objectives. It reveals that while the model possesses memory structures, its output prioritization still leans on prompt recency rather than conceptual importance.

Summary: Prompt Design Strategies and Platform-Specific Optimization

In summary, text-based image generation relies heavily on prompts as both technical inputs and semantic bridges. Each platform showcases distinct strengths and weaknesses: Leonardo.ai excels in fine detailing but struggles with cultural alignment; Doubao offers robust semantic parsing in Chinese but lacks iterative logic; ChatGPT supports interactive refinement but requires improvements in semantic consistency.

Effective prompt design must therefore be tailored to platform-specific traits. For Leonardo.ai, a segmented keyword structure (e.g., "delicate vines, symmetrical leaves, pale yellow background") proves useful. Doubao responds best to poetic and impressionistic phrasing (e.g., "藤蔓交织，空灵留白"). ChatGPT benefits from layered prompts combining initial layout, stylistic guidance, and iterative commands.

Future integration of AI and traditional art will hinge on more sophisticated language understanding systems capable of modeling not only vocabulary but also context, emotion, and cultural depth.

IMAGE REGENERATION AND STYLISTIC MIGRATION

With advancements in image-to-image transformation, AI's capacity to deconstruct and reconstruct visual features based on image inputs has become an important benchmark for stylistic learning. This mode allows users to bypass the ambiguities of text prompts, providing more precise control over visual output. To evaluate this, the study introduces original Gongbi paintings into two platforms—ChatGPT and Artbreeder—to examine how each processes, retains, or reinterprets traditional stylistic elements across the Six Principles.

Theoretical Foundation: Xie He's Six Principles as Evaluation Criteria

Xie He's Six Principles—proposed during the Southern Qi Dynasty—constitute the earliest known theoretical framework for Chinese painting, encompassing the following dimensions: spirit resonance (氣韻生動), structural use of line (骨法用筆), formal likeness (應物象形), color application (隨類賦彩), compositional design (經營位置), and stylistic transmission (傳移模寫) (Xie He, ca. 6th century / 1962). These principles not only established the foundational standards for evaluating classical Chinese painting, but also embody the discipline's core artistic pursuits—including formal language, aesthetic ideals, and technical spirit.

In this study, Xie He's framework serves as the theoretical basis for evaluating the aesthetic performance of AI-generated Gongbi-style images. Each principle is mapped onto a corresponding evaluative dimension—spirit, linework, form, color, composition, and imitation—reflecting the essential characteristics of Gongbi painting.

Traditionally, stylistic transmission (傳移模寫) referred to the practice of copying and studying canonical artworks, emphasizing the inheritance of technique, style, and spiritual essence. However, since the reference images used in this research are original Gongbi paintings created by the author, this principle is reconceptualized as stylistic simulation—that is, the AI's ability to reproduce and transform the visual attributes of the input images.

This methodological adjustment honors the internal value system of Chinese painting while establishing a culturally grounded framework for assessing AI-generated imagery. It enables the systematic analysis of how AI replicates or deviates from traditional aesthetics, revealing the creative potential and expressive gaps in technical simulation.

By using this model, the study facilitates a shift from purely technical replication toward aesthetic judgment, transforming traditional subjective standards into analytical dimensions applicable to AI-generated visual art.

ChatGPT Img2Img Analysis: Strength in Structure, Weakness in Nuance

When provided with original Gongbi reference images, ChatGPT generates images that show relative structural clarity and compositional integrity. The botanical forms—such as vines and leaves—are generally accurate in shape and positioning, reflecting strong image recognition and spatial mapping capabilities.

However, limitations emerge in the areas of spirit resonance and linework. The visual atmosphere often feels flat, lacking the poetic elegance and spiritual rhythm intrinsic to Gongbi painting. Line treatment tends to rely on pixel blending rather than deliberate brush-like articulation. As a result, the outputs fall short in replicating the nuanced control of brush pressure, rhythm, and tonal modulation characteristic of traditional hand-drawn techniques.

Color rendering also reveals a tendency toward simplification. Although base tones are retained, layering effects—such as the fine glazing techniques used in Gongbi—are underdeveloped, resulting in either over-blended areas or uniform color patches. This indicates that while ChatGPT is capable of replicating structure, it struggles with expressive detailing and tonal complexity.

Artbreeder Img2Img Analysis: Success in Atmosphere, Shortcomings in Form

Artbreeder, by contrast, emphasizes genetic blending of images, enabling it to excel in stylistic migration. When input with images sharing tonal unity, the platform produces visually coherent results with preserved atmosphere and stylistic consistency. The platform is particularly effective in translating the ethereal quality of

Gongbi through soft transitions, diffused light, and delicate color gradients.

In terms of spirit resonance and color rendering, Artbreeder often achieves a painterly softness and tonal luminosity akin to Gongbi's aesthetic. Background transitions are smooth, and the ambient tone aligns well with traditional Chinese concepts of balanced emptiness and poetic subtlety.

However, Artbreeder's performance in shape accuracy and linework is less consistent. Structural deformities—such as distorted leaves or misaligned petal arrangements—frequently occur. Fine details, especially in vines and stamens, are often simplified or blurred. The system favors soft diffusion over precise contouring, making it less suitable for works requiring formal fidelity.

Compositional balance is another mixed area. While Artbreeder maintains a general sense of layout, it lacks fine-tuned control over focal points and negative space management. The outputs tend to be densely packed, violating Gongbi's emphasis on compositional breathing room and spatial poetics.

Style Fusion Experiments: Comparative Results

To further test the platforms' capacity for stylistic synthesis, a batch of three author-created Gongbi works with unified color palettes was input into both platforms. The results diverged in significant ways: ChatGPT's outputs remained stable in tone but showed limited stylistic variation, indicating weak responsiveness to input diversity. In contrast, Artbreeder displayed greater flexibility, adjusting hue, rhythm, and atmospheric tone in accordance with the reference set. This demonstrates Artbreeder's superior aptitude for stylistic adaptation.

In terms of stylistic transmission (傳移模寫), Artbreeder outperforms ChatGPT by more effectively absorbing and reconfiguring aesthetic features. However, neither platform demonstrates adequate proficiency in brushwork simulation (骨法用筆) or compositional cadence (經營位置), both of which are foundational to Gongbi artistry.

Summary: Artistic Potential and Technical Constraints of Img2Img Generation

In summary, ChatGPT and Artbreeder each present unique strengths and limitations within the Img2Img domain. ChatGPT is better at structural reconstruction, making it suitable for representational image regeneration. Artbreeder is more adept at atmosphere and stylistic blending, offering promise for poetic or impressionistic visual outputs.

Yet both platforms fall short in replicating the calligraphic rhythm, nuanced detailing, and compositional logic required for high-fidelity Gongbi simulation. The core challenge lies in the AI's inability to internalize non-quantifiable elements such as brush energy, spatial intentionality, and aesthetic philosophy—facets deeply rooted in the Gongbi tradition.

Future advancements in AI art generation must therefore move beyond stylistic mimicry toward deeper modeling of traditional artistic consciousness, integrating algorithmic precision with culturally attuned visual intelligence.

CONCLUSION

This study systematically compared four mainstream AI image generation platforms—Leonardo.ai, Doubao, ChatGPT, and Artbreeder—from two perspectives: text-to-image generation and image-to-image transformation. It aimed to evaluate the capability, mechanism, and limitations of these platforms in simulating the stylistic essence of traditional Chinese Gongbi painting. The findings reveal that although current AI platforms have made considerable progress in areas like color blending, visual style approximation, and shape recognition, they still face significant limitations in achieving higher-level artistic expression and cultural representation.

From the text-to-image generation perspective, most platforms produce results that resemble a visual collage based on extracted keywords rather than a deep understanding of the aesthetic logic underlying Gongbi painting. Critical features such as line quality, brush structure, and compositional rhythm remain elusive. Despite efforts to refine prompts for better output, existing models struggle to grasp and reproduce traditional Chinese art terms, aesthetic metaphors, and cultural subtleties. Even Doubao, which operates within a native Chinese semantic context, exhibits limitations in multi-round semantic linkage and compositional control, revealing broader issues in constructing multi-dimensional aesthetic frameworks through AI.

In contrast, the image-to-image pathway offers a more promising route for AI to simulate Gongbi styles. Notably, Artbreeder demonstrates a degree of proficiency in capturing stylistic consistency and atmospheric coherence, while ChatGPT shows strength in structural reconstruction and formal likeness. However, both platforms fall short in replicating core aesthetic principles such as brushwork vitality (骨法用筆) and spiritual resonance (氣韻生動), which are fundamental to Gongbi artistry. These gaps largely stem from the AI's reliance on visual statistics and algorithmic inference, lacking the embodied artistic intelligence that human creators employ through a synthesis of hand, eye, and mind.

Overall, AI has shown considerable capability in mimicking surface-level visual elements, including color schemes, compositional structure, and general stylistic tone, suggesting its potential as an auxiliary tool. By inputting reference images or refining prompt descriptions, creators can quickly generate visually appealing drafts, which is particularly useful in early-stage concep-

tualization, stylistic exploration, and visual brainstorming. For students, illustrators, and curators needing rapid visual ideation, AI serves as an effective "inspiration trigger."

However, this generative ability remains superficial, limited to the manipulation of formal features such as shapes, hues, and spatial arrangements. It lacks the depth required for cultural interpretation, aesthetic reasoning, and emotional transmission—dimensions that define the artistic spirit of Gongbi painting. Gongbi is not merely about representation but about intention. It embodies sensitivity to nature, rational structuring, meticulous brushwork, and an overarching spiritual coherence. Current AI systems are not yet capable of modeling or expressing these non-verbal, non-formalized dimensions of artistic experience.

Moreover, while AI can mimic the appearance of traditional paintings, it bypasses the procedural rigor of artistic training, omitting the multi-layered development path of brushwork, modeling, composition, coloration, and spiritual refinement. This form of "shortcut generation" may satisfy visual output demands but weakens the value of skill cultivation and personal artistic growth. In other words, AI has yet to fulfill the role of art-making as a practice of self-cultivation. The resulting images, while visually polished, often lack emotional depth and a sense of lived vitality.

To fully integrate AI into the domain of traditional artistic practice, future research must move beyond algorithmic imitation and engage with the deeper cognitive, emotional, and cultural layers of visual intelligence. Only through such integration can AI evolve from a tool of visual synthesis into a true collaborator in aesthetic creation.

References

1. Chen, F.-f. (Curator). (2022). Transcending the Gongbi tradition: Contemporary Hong Kong fine-brush painters [Exhibition catalogue]. Hong Kong: Hongkong Land.
2. Xie, H. (1962). Guhua Pinlu [古画品录] (Vol. 11). Mao's Jiguge Studio.
3. Zhang, Z., Wang, T., Guo, X., Long, Z., & Xu, K. (2024). A review of AI painting research. Journal of Frontiers of Computer Science & Technology, 18(6). https://doi.org/10.3778/j.issn.1673-9418.2401075