Research on Innovative Methods and Visual Efficacy of Artistic Graphic Design Based on Multimodal Generative AI

Yong Yang
Shaanxi Open school, Cultural Communication and Art Design University, Xi'an 710119, Shaanxi, China
Corresponding author: Yong Yang (e-mail: artdesignyy@163.com)

ABSTRACT In the context of the deep integration of artificial intelligence generated content (AIGC) and digital creative industries, traditional art graphic design generally faces prominent problems such as long creative cycles, slow style iteration, visual expression homogenization, difficulty in multi-element collaboration, and difficulty in quantifying visual efficacy. Multimodal generative AI provides revolutionary technological support for the innovation of the entire process of artistic graphic design by leveraging the collaborative understanding and high-precision content generation capabilities of multisource information such as text, images, layout, color, and semantics. This article focuses on multimodal generative AI technology and systematically constructs an innovative method system for the entire process of artistic graphic design, which includes "requirement understanding multimodal guidance hierarchical generation precise control visual efficacy evaluation iterative optimization"; Deeply analyze multimodal collaborative generation mechanisms such as text guidance, image reference, sketch constraints, style transfer, layout alignment, etc; Propose innovative paradigms for poster design, brand vision, cultural and creative graphic design, packaging design, and other scenarios.

INDEX TERMS multimodal generative AI,art graphic design,visual efficacy,human-computer collaboration,engineering visualization

I. INTRODUCTION

A. RESEARCH BACKGROUND AND SIGNIFICANCE

ART graphic design is a core foundational link in visual communication, brand dissemination, cultural and creative industries, and digital media fields, undertaking important functions such as information expression, aesthetic presentation, value transmission, and brand building. Traditional graphic design heavily relies on designer experience, aesthetic literacy, and manual execution, and generally faces practical bottlenecks such as low creative output efficiency, high trial and error costs, difficulty in coordinating multi-element typesetting, lagging style iteration, and difficulty in quantitatively evaluating visual effects. Under the dual drive of new media communication and consumer upgrading, the market's demand for graphic design presents a trend of short cycle, high originality, strong visual, multi scene, and quantifiable. Traditional work models are difficult to adapt to the efficient, precise, and diversified needs of the modern design industry[1-3].

Multimodal generative AI, as the core technology of the new generation of artificial intelligence, can achieve unified encoding and deep fusion of multiple types of information such as text, images, sketches, layouts, and colors. It has the ability to generate high-precision images, understand and transfer styles, optimize layouts intelligently, align text visual accurately, and output editable layers. It provides key technical support for the full process reconstruction of artistic graphic design. The maturity of technology systems such as StableDiffusion, DALL · E3, Midjourney, and specialized models for graphic design such as CreatiDesign and CreatiPoster enables AI to upgrade from a creative assistant to a design collaborative subject, promoting the transformation of design from "experience driven" to "data-driven, model driven, and efficiency driven". Consequently, investigating innovative methodologies for AI-integrated graphic design is essential for the evolution of digital creative industries. This study establishes a theoretical framework for multimodal design innovation, improving the framework of hierarchical generation, precise regulation, and efficiency evaluation[4].

B. RESEARCH SIGNIFICANCE

1) Theoretical Significance

Build a multimodal generative AI driven innovation theoretical system for art graphic design, and improve the complete theoretical framework of multimodal guidance, hierarchical generation, precise regulation, and efficiency evaluation; Establish a multidimensional quantitative evaluation model for visual efficacy to fill the research gap in intelligent and quantifiable evaluation of artistic graphic design; Expand the interdisciplinary research boundary between design disciplines and artificial intelligence, providing theoretical support for human-machine collaborative design[5].

2) Practical Significance

Propose innovative methods that can be directly implemented throughout the entire process, significantly improving design efficiency, shortening creation cycles, and lowering professional barriers; Realize high-precision, highly controllable, and highly consistent professional graphic design output, alleviate pain points such as text distortion, layout imbalance, and style confusion; Establish a quantitative evaluation system for visual efficacy, achieve measurable, comparable, and optimized design effects, and provide efficient and intelligent solutions for brand design, advertising communication, cultural and creative products, new media vision, and other fields[6].

C. RESEARCH CONTENT AND METHODS

1) Research Content

1. The adaptation mechanism and theoretical basis of multimodal generative AI and artistic graphic design;

2. Multimodal guidance, controllable generation, layered decoupling, and precise optimization core technologies for graphic design;

3. Construction of an innovative methodology system for the entire process of "demand guidance generation regulation evaluation iteration";

4. Construction of a multidimensional visual efficacy quantitative evaluation system and indicator model;

5. Empirical design, comparative experiments, and effectiveness verification of typical scenarios;

6. Optimize strategies and establish a long-term mechanism for human-machine collaboration.

2) Research Methods

Literature research method, technical analysis method, model construction method, comparative experiment method, eye tracking experiment method, questionnaire survey method, quantitative statistics method, empirical research method[7].

II. CORE CONCEPTS AND THEORETICAL FOUNDATIONS

A. DEFINITION OF CORE CONCEPTS

Art graphic design is a creative activity that uses visual symbols as a carrier to organize elements such as graphics, text, color, layout, texture, etc., to achieve information communication, aesthetic expression, and value dissemination. Emphasizing composition order, semantic accuracy, unified style, visual aesthetics, and communication effectiveness, it is widely used in posters, brand VI, packaging, advertising, cultural and creative fields, and combines artistic, normative, functional, and commercial aspects[8-9].

B. THEORETICAL BASIS

Emphasize the division of labor and collaboration between designers and AI: AI undertakes tasks such as efficient generation, batch iteration, quantitative calculation, and rule optimization; Designers are responsible for requirement definition, creative decision-making, cultural expression, value judgment, and professional regulation to achieve complementary advantages[10].

III. INNOVATIVE METHOD FOR THE ENTIRE PROCESS OF ARTISTIC GRAPHIC DESIGN BASED ON MULTIMODAL GENERATIVE AI

A. OVERALL METHODOLOGICAL FRAMEWORK

The innovative method of artistic graphic design based on multimodal generative AI is not a partial improvement of the traditional design process, but a comprehensive reconstruction of design objectives, input conditions, generation mechanisms, regulation methods, and evaluation rules based on the underlying support of artificial intelligence's multimodal understanding ability, controllable generation ability, and iterative optimization ability. Based on a comprehensive analysis of the creative laws of artistic graphic design, visual communication logic, and the operation mechanism of multimodal models, this study proposes a six-in-one innovative method system. The six core components are: (1) Requirement Analysis, (2) Multimodal Constraint Construction, (3) Hierarchical Generation, (4) Human-Machine Collaborative Regulation, (5) Visual Efficacy Evaluation, and (6) Closed-loop Iterative Optimization. This system is based on the unified representation of multimodal information, with hierarchical decoupling to ensure design professionalism, human-machine collaboration to ensure cultural and semantic accuracy, quantitative evaluation to achieve traceable effects, closed-loop iteration to continuously improve design quality, and ultimately form a stable, efficient, replicable, and scalable intelligent design paradigm. Compared with traditional design patterns, this method no longer relies on individual designer experience and long-term training. Instead, it transforms professional knowledge, composition rules, color patterns, layout specifications, and communication logic into structured conditions that can be learned, executed, and optimized by the model. This upgrades AI from an auxiliary tool to a collaborative creative subject, and transforms the design process from experience driven to data-driven, rule-based, and efficiency driven, achieving systematic breakthroughs in efficiency, quality, diversity, controllability, and other aspects[11-12].

Multimodal Input Forms and Design Control Dimensions in Artistic Graphic Design

As shown in Table 1:

TABLE 1. Multimodal Input Forms and Design Control Dimensions in Artistic Graphic Design

Modal Type	Input Form	Design Control Dimension	Application Scenario
Text Modal	Prompt, description	Theme, style, layout, semantics, tone	Poster, brand design, advertising
Image Modal	Reference, style map	Style migration, color reference, element reference	Cultural creation, VI design
Sketch Modal	Hand-drawn layout	Composition, position, proportion, structure	Layout design, rapid prototyping
Layout Modal	Grid, text area	Alignment, hierarchy, spacing, specification	Standardized commercial design
Color Modal	Palette	Hue, saturation, contrast, psychological matching	Visual system, packaging design

B. REQUIREMENT ANALYSIS AND MULTIMODAL CONSTRAINT DEFINITION

Firstly, the requirements are hierarchically decomposed into core information layer, style constraint layer, layout structure layer, color system layer, application scenario layer, and detail specification layer. The core information layer includes key content that must be presented, such as titles, subtitles, explanatory text, and identification graphics; The style constraint layer includes China-Chic, minimalism, retro, cyberpunk, ink texture, flat vector, artistic texture and other visual tendencies; The layout structure layer includes vertical composition, horizontal composition, center symmetry, grid system, visual flow line, main body highlighting method, etc; The color system layer includes main color, auxiliary color, contrast color, brightness relationship, saturation range, and color psychological orientation; The application scenario layer includes different output forms such as posters, brand VI, packaging, cultural and creative industries, new media covers, outdoor advertising, etc. Through hierarchical decomposition, all fuzzy descriptions are transformed into clear design constraints, providing stable inputs for the subsequent construction of multimodal prompts, as shown in Table 2.

On this basis, further mapping the constraint conditions into five modal guidance signals: text, image, sketch, layout, and color. The text modality undertakes the functions of semantic description, style definition, composition instructions, and logical relationship explanation; The image modality serves as a style reference, element reference, tone reference, and layout reference function; The sketch mode assumes the functions of position constraint, scale constraint, and structural constraint; The layout mode undertakes the functions of text area, graphic area, alignment method, and spacing rule; Color modes provide precise control over color values, color ratios, and temperature preferences. The construction of multimodal constraints makes AI generation no longer a random creation, but a precise output under the joint action of multiple professional conditions, fundamentally improving design reliability and professionalism, as shown in Table 3 below.

The weights in Table 3 were determined through the Analytic Hierarchy Process (AHP) based on consultations with 10 industry experts. The updated weights are: Visual Appeal (18%), Information Communication (20%), Aesthetic Integrity (17%), Style Consistency (15%), Recognition & Memory (15%), and Communication Adaptability (15%), totaling 100%.

IV. EMPIRICAL RESEARCH AND COMPARATIVE EXPERIMENTS

A. EXPERIMENTAL DESIGN AND GROUPING

To verify the practical effectiveness of the multimodal generative AI art graphic design innovation method proposed in this study, a rigorous controlled experiment was conducted. The experiment selects three most representative art graphic design scenarios: commercial poster design, brand visual aids, and cultural and creative graphic design. Each scenario is set with unified design requirements, time constraints, and evaluation standards. The experiment is divided into an experimental group and a control group: the experimental group adopts the full-process innovation method proposed in this study. Specifically, the system utilizes Stable Diffusion XL (SDXL) as the core generative engine, fine-tuned via Low-Rank Adaptation (LoRA) on a proprietary dataset of 5,000 high-quality professional posters and brand identities. The text-to-image process is further guided by ControlNet for sketch and layout constraints to ensure structural precision. This includes multimodal constraints, hierarchical generation, human-machine collaborative regulation, visual efficacy evaluation, and closed-loop iteration; The control group adopted traditional design methods. To eliminate bias, 20 professional designers with 5–8 years of experience were recruited and randomly assigned to the two groups. The control group utilized conventional design software, specifically Adobe Photoshop (v25.0) and Adobe Illustrator (v28.0), to complete the creation. The two groups are completely consistent in task difficulty, delivery standards, and evaluation system, ensuring comparability and credibility of experimental results.

B. EVALUATION INDICATORS AND DATA SOURCES

The experiment sets comparative indicators from four dimensions: design efficiency, generation quality, visual efficacy, and professional recognition. Design efficiency indicators include average completion time, output per unit time, and number of iterations; The generated quality indicators include first pass rate, text accuracy rate, composition standardization rate, and style consistency; The 'Qualified Rate' is defined as the percentage of designs receiving a comprehensive score of 80 or higher based on the six-dimensional rubric. To ensure objectivity, a triple-blind evaluation was conducted by five independent senior designers (each with >10 years of experience), who were unaware of whether a design was AI-generated or human-made. The visual efficacy index adopts the six dimensional evaluation system constructed in this study, including attractiveness, infor-

TABLE 2. Full-process Innovative Design Method Based on Multimodal AI

Process Link	Core Task	Technical Support	Output Result
Demand Analysis	Structured constraint definition	Demand modeling, classification	Multimodal constraint template table
Multimodal Prompt	Text+image+sketch+layout fusion	Prompt engineering, conditional	Standardized prompt fusion
Layered Generation	Background, graphic, text, decoration	Diffusion model, LoRA, layered decoupling	Initial design draft
Collaborative Adjustment	Layout, text, color, details	Human-computer interaction, fine-tuning	Revised draft
Effect Evaluation	6-dimensional quantitative scoring	Aesthetic computing, eye-tracking	Visual efficacy report
Closed-loop Iteration	Continuous design refinement	Feedback-driven optimization	Final optimized output

TABLE 3. Visual Efficacy Quantitative Evaluation Index System

Dimension	Specific Indicators	Weight	Evaluation Method
Visual Appeal	Contrast, color harmony, prominence	18%	Algorithm + Subjective score
Information Communication	Clarity, readability, speed	20%	Eye-tracking + Test
Aesthetic Integrity	Composition, balance, detail	17%	Aesthetic computing
Style Consistency	Unity, coordination, order	15%	Feature matching
Recognition & Memory	Uniqueness, identification	15%	Memory test
Communication Adaptability	Scene fit, audience matching	15%	Questionnaire

mation transmission efficiency, aesthetic integrity, memory, etc; The professional recognition is evaluated by a blind evaluation panel composed of industry designers, brand practitioners, and communication experts. The data sources include platform automatic recording, algorithm scoring, eye tracker collection, questionnaire survey, expert scoring, and user testing to ensure comprehensive, objective, and reliable data, as shown in Table 4 below.

TABLE 4. Comparison of Experimental Results between Experimental and Control Groups

Indicator	Experimental Group	Control Group	Improvement Rate
Average Design Cycle	2.1 h	6.6 h	-68.4%
Qualified Rate	92.7%	61.5%	+50.7%
Visual Efficacy Score	86.3	60.6	+42.5%
Information Efficiency	88.2%	64.3%	+37.2%
Style Consistency	94.3%	67.8%	+39.1%

V. OPTIMIZATION STRATEGY AND LONG-TERM MECHANISM OF HUMAN-MACHINE COLLABORATION

A. MULTIMODAL GENERATION ACCURACY OPTIMIZATION STRATEGY

Build a professional graphic design dataset and train LoRA models specifically for text, layout, and branding; Improve the engineering standards for multiple constraint prompts to reduce layout conflicts and text distortion; Adopting a combination of hierarchical generation and post refinement to ensure professional output.

B. STRATEGY FOR CONTINUOUS IMPROVEMENT OF VISUAL EFFICIENCY

Establish a performance parameter mapping library to provide precise optimization paths for low dimensional data; Strengthen the rules and constraints of composition, color, and information hierarchy; Combine feedback from dissemination data and iteratively update the evaluation model.

C. HUMAN MACHINE COLLABORATIVE DIVISION OF LABOR MECHANISM

AI: multimodal generation, batch iteration, layout optimization, performance calculation, quantitative evaluation;

Designer: requirement definition, cultural expression, creative decision-making, textual semantics, brand standards, and fine-tuning approval.

D. INDUSTRY LANDING GUARANTEE STRATEGY

Establish standardized process templates, prompt word libraries, style libraries, and control manuals; Conduct AI capability training for designers; Promote the construction of copyright norms and quality standards, and promote the healthy and orderly application of technology, as shown in Table 5 below.

VI. CONCLUSION AND PROSPECT

A. RESEARCH CONCLUSION

This article systematically constructs an innovative method for the entire process of artistic graphic design based on multimodal generative AI, clarifies the internal mechanisms of multimodal collaboration, hierarchical generation, controllable regulation, and efficacy evaluation, and establishes a six dimensional visual efficacy quantitative evaluation system. Empirical evidence shows that this method can significantly improve design efficiency, originality, and visual efficacy, effectively addressing pain points such as long design cycles, homogenization, and blurred efficacy in traditional design. It provides a complete theoretical framework and practical path for the intelligent transformation of artistic graphic design.

B. FUTURE OUTLOOK

In the future, we will combine technologies such as semantic understanding of large models, digital humans, multimodal

TABLE 5. Comprehensive Comparison Between Multimodal AI Design and Traditional Design

Comparison Items	Multimodal Generative AI Design Method	Traditional Manual Design Method	Advancement of the Proposed Method
Design Cycle	Short, average 1–3 hours	Long, average 6–24 hours	Shortened by more than 68%
Scheme Output Quantity	8–15 sets per task	2–3 sets per task	Increased by 4–6 times
Layout Standardization	High, automatic alignment & hierarchy	Unstable, depends on designer experience	More standardized & consistent
Text Accuracy & Quality	High (stratified generation + text enhancement)	High, but time-consuming	More efficient under same accuracy
Style Consistency	Up to 94.3%	About 67.8%	Improved by 39.1%
Visual Attraction	86.3 (average score)	60.6 (average score)	Improved by 42.5%
Information Communication Efficiency	88.2%	64.3%	Improved by 37.2%
Degree of Innovation & Diversity	High, cross-style fusion & rapid iteration	Limited by thinking & experience	Significantly enhanced
Human Intervention Intensity	Low, mainly for decision-making & fine-tuning	High, full-process manual production	Reduced labor workload >60%
Application & Promotion Value	High, standardized & replicable	Low, highly personalized & non-standard	Convenient for large-scale application

interaction, and 3D extended design to build a more intelligent, accurate, and professional AI design system; Deepen research on copyright, ethics, and standards; Promote the large-scale implementation of AI graphic design in cultural and creative, branding, advertising, new media and other fields, and assist in the high-quality development of the digital creative industry.

AUTHOR CONTRIBUTIONS

Yang Yong designed, collected and analyzed the data, and drafted the manuscript. Yang Yong conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Yang Yong participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] D. Koundal, M. Khan, A. R. Mahajan, et al., "Intelligent Precision Healthcare: Multimodal Generative AI for Neurodegenerative Disease Management," CRC Press; 2026.
[2] T. Yin, J. Chen, T. Yao, et al., "Development and initial validation of a video-mediated listening test based on multimodal generative AI," Acta psychologica, vol. 264, 106521, 2026, doi: 10.1016/j.actpsy.2026.106521.
[3] M. Dehen and N. Aharony, "Multimodality Usage of Generative AI (GenAI) Chatbots: An Exploratory Study of User Characteristics," TechTrends, vol. 70, no. 2, pp. 1–14, 2026, doi: 10.1007/s11528-026-01169-2.
[4] J. McGreevy, E. Klang, and G. Nadkarni, "Ethics Guides Innovation: Reimagining Sepsis Care with Multimodal Generative AI," The American Journal of Bioethics, vol. 26, no. 2, pp. 110–112, 2026, doi: 10.1080/15265161.2025.2608652.
[5] I. Poggi, T. Scaramella, S. Violini, et al., "The Ten Minutes That Shocked the World-Teaching Generative AI to Analyze the Trump-Zelensky Multimodal Debate," Information, vol. 17, no. 2, pp. 136–136, 2026, doi: 10.3390/info17020136.
[6] A. Rofena, L. C. Piccolo, B. B. Zobel, et al., "Augmented intelligence for multimodal virtual biopsy in breast cancer using generative artificial intelligence," Journal of biomedical informatics, vol. 174, 104971, 2025, doi: 10.1016/j.jbi.2025.104971.
[7] K. A. Saxena, R. Prasad, and S. Laha, "Early detection of mental health conditions using multimodal generative AI with MI-GBF and fusion-three branch network," International Journal of Information Technology, vol. (prepublish), pp. 1–8, 2025, doi: 10.1007/s41870-025-03042-6.
[8] M. Pang, K. T. Roy, X. Wu, et al., "Query-driven generative AI synthesizes multi-modal spatial omics from histology," bioRxiv : the preprint server for biology, 2025, doi: 10.64898/2025.12.11.693669.
[9] N. H. Mai, H. D. Lee, J. Kaenploy, et al., "Performance of Multimodal Generative AI Models in Addressing Complex Dental Inquiries With Text, Images, and Analytical Data," Journal of esthetic and restorative dentistry : official publication of the American Academy of Esthetic Dentistry, vol. 38, no. 1, pp. 166–172, 2025, doi: 10.1111/jerd.70064.
[10] P. C. Míguez, C. M. Suárez, D. Crucitti, et al., "Code red (Código Rojo): Multimodal generative AI for specialized hematology education," Blood, vol. 146, no. S1, pp. 4331–4331, 2025, doi: 10.1182/blood-2025-4331.
[11] G. Revathy, K. A. Natarajan, and N. Kshetri, "Fusion of Multimodal Generative AI and Blockchain Technology in Digital Media," IGI Global Scientific Publishing; 2025, doi: 10.4018/979-8-3373-1504-1.
[12] H. Hang and Z. Li, "Research on the construction of intelligent art design system based on multimodal perception and generative AI," Discover Applied Sciences, vol. 7, no. 9, pp. 945–945, 2025, doi: 10.1007/s42452-025-07513-0.