Open Access Library Journal
2026, Volume 13, e16135
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Visual Semiotics-Based Emotional Expression and Cross-Cultural Visual Narrative Construction in AI Text-to-Image Generation

—A Case Study of "Loneliness" and "Warmth"

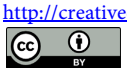
Yipu Kong, Zhengjie Guo

School of Foreign Studies, Hebei University, Baoding, China
Email: zhengjieguo086@126.com

How to cite this paper: Kong, Y.P. and Guo, Z.J. (2026) Visual Semiotics-Based Emotional Expression and Cross-Cultural Visual Narrative Construction in AI Text-to-Image Generation. *Open Access Library Journal*, **13**: e16135
https://doi.org/10.4236/oalib.1116135

Received: October 9, 2026
Accepted: November 8, 2026
Published: November 11, 2026

Copyright © 2026 by author(s) and Open Access Library Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



Open Access

Abstract

With the rapid advancement of multimodal generative artificial intelligence, Text-to-Image technology is transitioning from simple pixel imitation to advanced semantic and emotional generation. This study selects two contrasting emotional themes, "loneliness" and "warmth", as experimental samples to compare the performance of international models (Leonardo.Ai) and Chinese domestic counterparts (Doubao, Kling) in visualizing abstract emotions. Grounded in Roland Barthes' theory of visual semiotics, this paper establishes an evaluation framework encompassing Signifier-Signified Matching Degree and Cultural Adaptability and Innovation to systematically analyze how different models translate abstract emotions into concrete visual symbols (color, lighting, and composition). The findings reveal that domestic large models exhibit superior performance in high-context decoding and localized symbol construction within Chinese cultural settings, though they still face challenges in balancing photorealistic perfection with feature diversity. Finally, this paper discusses the shift in aesthetic intent of human creators and the pathways of human-AI co-creation in the AIGC era, aiming to provide theoretical and empirical references for understanding human-AI emotional interaction and digital visual narration.

Subject Areas

Linguistics, Multimodal Communication

Keywords

AIGC, Text-to-Image, Visual Narrative, Visual Semiotics, Affective Expression

1. Introduction

In recent years, Artificial Intelligence Generated Content (AIGC) technology has experienced explosive growth. Early image generation technologies, such as Generative Adversarial Networks (GANs), primarily focused on image inpainting or simple style transfer. Their core objective was to make generated images physically realistic, addressing the issue of whether a drawing resembles the real object [1]. Liu [2] (2024) points out that cutting-edge painting tools (such as Midjourney), video production models (such as Sora), and even brain-computer interfaces can now acquire skills such as playing Go, composing music, calligraphy, dancing, and piloting airplanes. This will profoundly reshape people's concepts of learning and knowledge. However, with the emergence of new-generation diffusion models represented by Stable Diffusion, AI has begun to develop robust semantic understanding capabilities. By simply inputting a text prompt, users can enable AI to generate highly complex images.

This technological advancement marks a transition of AIGC from simple pixel imitation to a higher-level semantic generation. Through continuous dialogue with users, AI incessantly identifies, learns, and integrates personalized user elements, and structurally processes the output elements to present them organically in a user-centric manner. This essentially accomplishes the establishment of a relationship with the user, serving as a deep simulation of human communication patterns [3]. Despite the exquisite quality of AI-generated artwork, a critical question persists: Does AI truly "understand" what loneliness and warmth are? Or does it merely calculate through massive datasets that the word "loneliness" has a high probability of co-occurring with black and silhouette? Gan and Wang [4] (2024) argue that emotion, as a vital adhesive, plays an essential role in driving human-computer interaction and constructing tightly woven social formations. Bender *et al.* [5] (2021) once compared large models to stochastic parrots, suggesting that they merely mimic the probabilistic combinations of human language without generating genuine consciousness. Emotion, as something unique to humans, was once considered the final barrier preventing artificial intelligence from replacing humans. However, on the one hand, driven by the trend of AI shifting toward emotional generation, humans seek barrier-free interaction when communicating with machines and anticipate establishing human-like feelings and intimate relationships with them. On the other hand, a growing body of scientific research indicates that artificial intelligence is capable of mimicking human emotions. Driven by this dual dynamic, human-computer emotional interaction is gradually becoming routinized [6].

The significance of this study lies in testing AI's reactions to abstract emotional words such as "loneliness" and "warmth" to observe the boundaries of its comprehension. This not only allows us to develop a more objective understanding of AI's capabilities but also holds practical value for future applications. If AI can accurately translate emotions into images, it will become a more empathetic human assistant in future artistic creations; additionally, in the field of psychological

healing, it can help individuals visualize inexpressible emotions, thereby assisting in psychological counseling [7].

2. Literature Review

In recent years, the rapid advancement of Text-to-Image generative AI models has profoundly transformed visual communication, prompting a burgeoning body of critical scholarship across media studies, discourse analysis, and human-computer interaction. Existing research on AI-generated images can be broadly categorized into three main areas. First, scholars conceptualize AI image generators as semiotic technologies that encode and reproduce social norms and ideologies. Rather than acting as neutral tools, models like Stable Diffusion and DALL-E function as cultural aggregators deeply inscribed with pre-existing societal values, meaning that their visual outputs inherently reflect the biases embedded within their training corpora [8]. Second, a significant strand of research focuses on the systematic amplification of stereotypes and social biases in AI visuals, extending from demographic identities to occupational stigma. While early audits revealed gender and racial disparities in professional representations [9], recent scholarship has begun to unpack how AI constructs specific occupational meanings. For instance, in an investigation of Chinese text-to-image models, Ye [10] (2026) introduced the concept of "visual prototype lock-in" to demonstrate how algorithms transform historical media discourses and social stigmas into rigid, recurring visual sign systems (such as default dark-uniformed, serious-expression middle-aged males for urban management officers). Similarly, studies in health and social domains show that AI models recycle reductionist cultural tropes, such as biomedical framings of "dementia" [11] or homogenized consumer-driven aesthetics of "teenagers" [8]. Third, recent methodological interventions treat AI image generation as a diagnostic tool for critical discourse analysis. Scholars like Hart [12] (2026) utilize AI-generated images to examine multimodal constructions, demonstrating how specific textual prompts systematically trigger recurring visual schemas—such as large, unbounded crowds of migrants—that evoke specific ideological framings and power relations.

Current research has successfully established that AI-generated imagery is far from politically or culturally neutral; it actively participates in the construction and naturalization of social realities. However, existing research exhibits two notable limitations. On the one hand, most studies predominantly concentrate on Western models (e.g., DALL-E, Midjourney), leaving the performance, cultural localization, and bias mechanisms of domestic or cross-lingual AI platforms underexplored. On the other hand, while extensive attention has been paid to demographic markers (race, gender) and macro-social topics (migration, health), there remains a scarcity of research investigating how generative AI translates abstract emotional concepts into visual narratives. Addressing this gap is crucial for understanding the boundaries of AI's semantic comprehension and its capacity for nuanced, cross-cultural visual storytelling.

3. Research Design

To investigate the semantic comprehension boundaries and visual narrative capabilities of Text-to-Image models regarding abstract emotions, this study adopts a comparative experimental design combined with visual semiotic analysis. The research design is structured around three core dimensions: model selection, experimental variables, and analytical framework.

First, in terms of model selection, this study establishes a cross-cultural comparative sample by deploying one international leading model (Leonardo.Ai) and two Chinese domestic models (Doubao and Kling). This selection allows for a rigorous examination of how different cultural backgrounds and training datasets influence AI's decoding of abstract concepts. Second, the experimental procedure employs a controlled-variable approach centered on prompt engineering and iterative optimization. The study investigates two contrasting emotional themes—"loneliness" (representing alienation and psychological vacancy) and "warmth" (representing interpersonal connection and intimacy). For each theme, a two-stage iteration is conducted:

Stage 1 (Baseline Generation, V1): Basic prompts containing abstract emotional keywords (e.g., extreme loneliness or family reunion dinner) are inputted to observe the models' baseline cultural decoding and default visual signifiers.

Stage 2 (Optimized Generation, V2): Advanced prompts incorporating technical parameters (e.g., cinematic lighting, shallow depth of field, negative space ratio), specific cultural signifiers (e.g., round tables, bamboo steamers), and negative constraints (e.g., no turkey, no fireplace) are introduced to test the models' capacity for psychological atmosphere conveyance and cross-cultural correction.

Finally, for data analysis, the study draws upon Roland Barthes [13]' theory of visual semiotics to decode the generated images beyond mere aesthetic evaluations. A comprehensive evaluation framework is established based on two core dimensions: (1) The Signifier-Signified Matching Degree, which quantifies the precision of emotional mapping by evaluating the number of effective signifiers (e.g., color tones, lighting, spatial composition) versus interfering noise; (2) Cultural Adaptability and Innovation, which assesses the models' ability to avoid cultural bias (such as Western-centric symbols in Chinese contexts) and break free from rigid emotion-signifier lock-ins. Through this multi-tiered design, the study systematically maps the transition of models from basic physical scene presentation to sophisticated, context-aware emotional storytelling.

4. Data Collection and Analysis

4.1. Data Collection

The empirical dataset comprises a total of 48 primary image samples generated across two distinct experimental phases (V1 and V2) for the two contrasting emotional themes. The data collection procedure was systematically organized as follows:

First, three mainstream generative AI platforms were selected for parallel test-

ing: Leonardo.Ai (utilizing its flagship Leonardo Diffusion XL model representing international platforms), Doubao (utilizing its multimodal native image generation model), and Kling (utilizing the Kling-V1.5 image generation engine representing Chinese domestic platforms). All prompts were submitted in English. Image dimensions were standardized at 1024 × 1024 pixels (1:1 aspect ratio) across all platforms. Random seed initialization was applied with fixed seeds where supported to account for generation variance, and platform-level "automatic prompt enhancement" (or prompt magic features) was strictly disabled to prevent confounding interventions from built-in language models.

Second, regarding the sampling procedure and generation workflow, exactly one generation call was executed for each platform, for each emotional theme, and at each experimental stage. Each single generation call inherently produced a grid of four sub-images originating simultaneously from the same random seed batch and the identical prompt input. In Phase 1 (baseline generation, V1), minimalist baseline prompts containing abstract emotional keywords were inputted into all three models to test their zero-shot comprehension. Specifically, the prompt "Extreme loneliness, a person sitting by the raining window, dim light, blue tone" was used for the loneliness theme, and "Family reunion dinner, warmth" was applied for the warmth theme. This phase yielded a baseline dataset of 24 images in total (3 models × 2 themes × 1 generation call × 4 sub-images per generation grid), capturing the models' default visual signifiers and initial cultural decoding. To maintain empirical objectivity and avoid selection bias, no generated outputs were excluded, cherry-picked, or regenerated; all four sub-images from each initial batch were strictly retained for subsequent semiotic analysis to accurately reflect the models' default probabilistic distribution.

Following the diagnostic evaluation of limitations in Phase 1, a second round of data (V2) was collected using iteratively refined prompts through a single, identical execution call per model and theme. For the loneliness theme, technical parameters and compositional constraints were added. For the warmth theme, culturally specific signifiers and negative constraints were introduced. This phase generated an additional 24 optimized images (3 models × 2 themes × 1 generation call × 4 sub-images) under the exact same unfiltered, batch-inclusive protocol. Consequently, the entire dataset comprises a total of 48 primary images, capturing the complete, unfiltered trajectory of the models' visual output across both experimental phases.

4.2. Data Analysis

The first is the visual narrative and iteration of "loneliness". The initial text prompt was formulated as follows: "Extreme loneliness, a person sitting by the raining window, dim light, blue tone" (极致孤独，一个人坐在雨夜的窗边，昏暗光线，蓝色调). The pictures generated by the three large models are as follows:

As shown in the results (**Figures 1-3**), while all three models successfully generated scenes matching the literal semantics of the prompt, they exposed notable

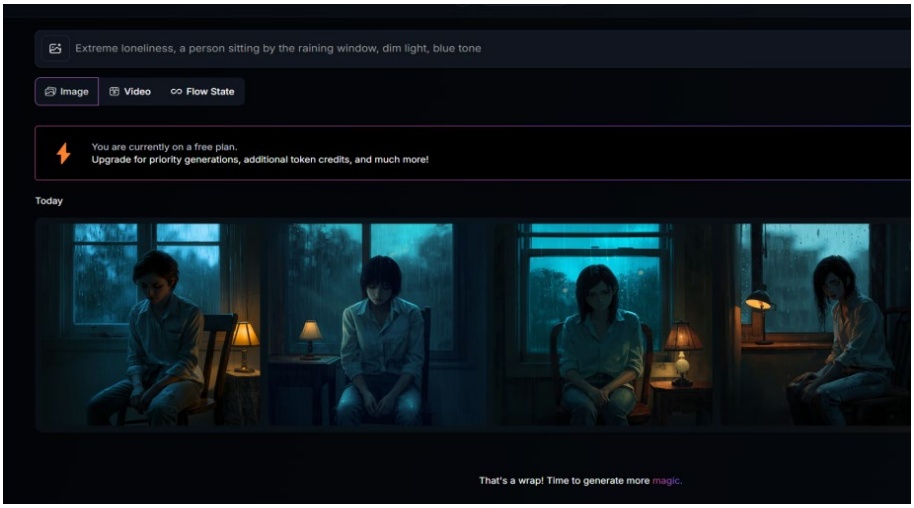


Figure 1. The first image on loneliness generated by Leonardo.Ai.

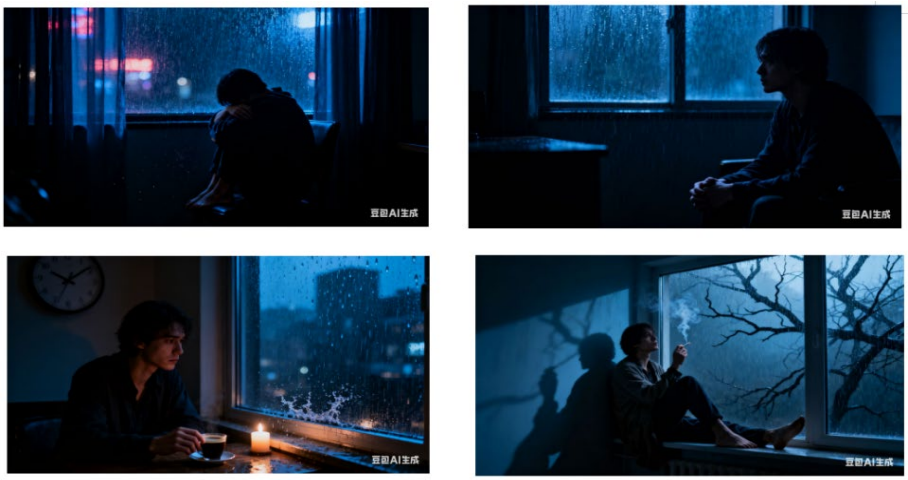


Figure 2. The first image on loneliness generated by Doubao.

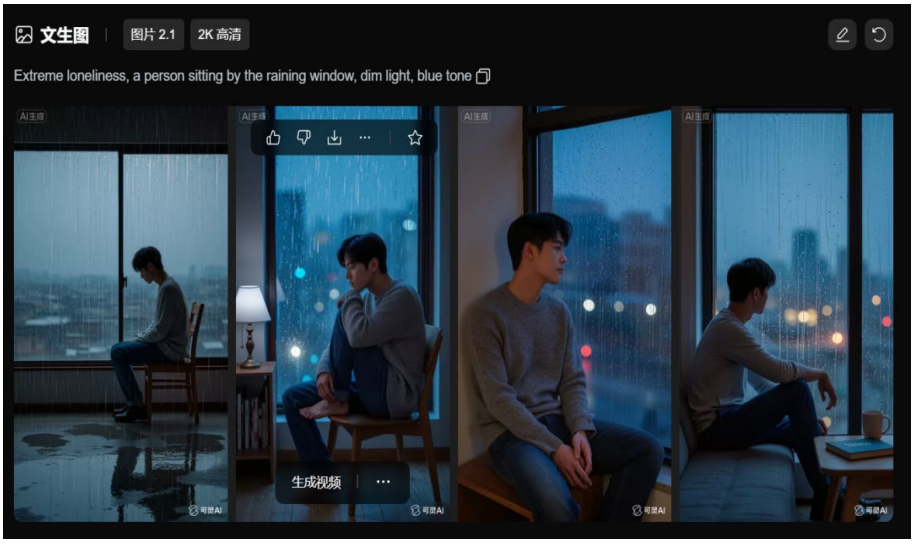


Figure 3. The first image on loneliness generated by Kling.

limitations in the depth and stability of emotional transmission: Leonardo.Ai demonstrated high element matching and delicate image textures, yet its emotional expression remained at a shallow level. The character's posture (hands clasped) appeared plain and facial expressions were vague, presenting merely the physical solitude of a single person in an empty scene rather than the introspective and wistful psychological state implied by extreme loneliness. The intensity modifier extreme failed to manifest visually. Doubao exhibited concrete emotional symbols and strong narrative qualities; however, it suffered from a lack of stylistic consistency. Significant variations in color tones (dark blue vs. dark yellow) and props (cigarettes vs. candles) across the four sub-images broke the unified dim atmosphere mandated by the prompt, resulting in fragmented emotional delivery. Kling provided a strong sense of immersion in everyday settings and clean composition, but presented an obvious conflict in affective atmosphere. Excessive warm light elements in the background—such as city neon lights and table lamps—clashed sharply with the cold, desolate baseline of blue tone and dim light, thereby diluting the intense chill required for extreme loneliness.

Then, optimize prompt: "Extreme loneliness, a person sitting by the raining window, cinematic lighting, blue and teal tone, soft rain streaks on glass, negative space accounting for 2/3 of the frame, shallow depth of field, moody and introspective". (极致孤独，一个人坐在雨夜的窗边，电影级灯光，蓝青色调，玻璃上的细雨痕，留白占画面 2/3，浅景深，忧郁且内省). The pictures generated by the three large models are as follows:

The output across all three models achieved a crucial transition from mere scene presentation to emotional narration (**Figures 4-6**). The optimized prompt—incorporating cinematic lighting, negative space (compositional breathing room), shallow depth of field, and introspective emotional cues—triggered key enhancements across all three platforms, elevating the imagery from a simple single-person scene to a psychologically introspective sense of loneliness.

Specifically, character postures (such as knees hugged to the chest and sideways contemplation) closely aligned with the moody and introspective emotional setting. Rather than conveying a plain physical solitude, the figures radiated a sense of desolation and wistfulness, significantly amplifying the emotional intensity of extreme loneliness. Furthermore, the constraints of blue and teal tone and cinematic lighting unified the desolate cooling baseline across all models, thoroughly resolving previous issues such as warm light conflicting with cold ambiance and stylistic fragmentation, thereby enhancing the immersive isolation and oppression of the rainy night by the window. Additionally, negative space accounting for 2/3 of the frame effectively resolved the prior issue of visual crowding, using large-area dark-toned negative space to reinforce the metaphor of a psychological vacancy. Finally, soft rain streaks and shallow depth of field refined the scene details (rendering soft and orderly rain traces) while blurring the background to focus on the human subject, successfully highlighting the core relational dynamic of human alienation from the outside world.



Figure 4. The second image on loneliness generated by Leonardo.Ai.



Figure 5. The second image on loneliness generated by Doubao.

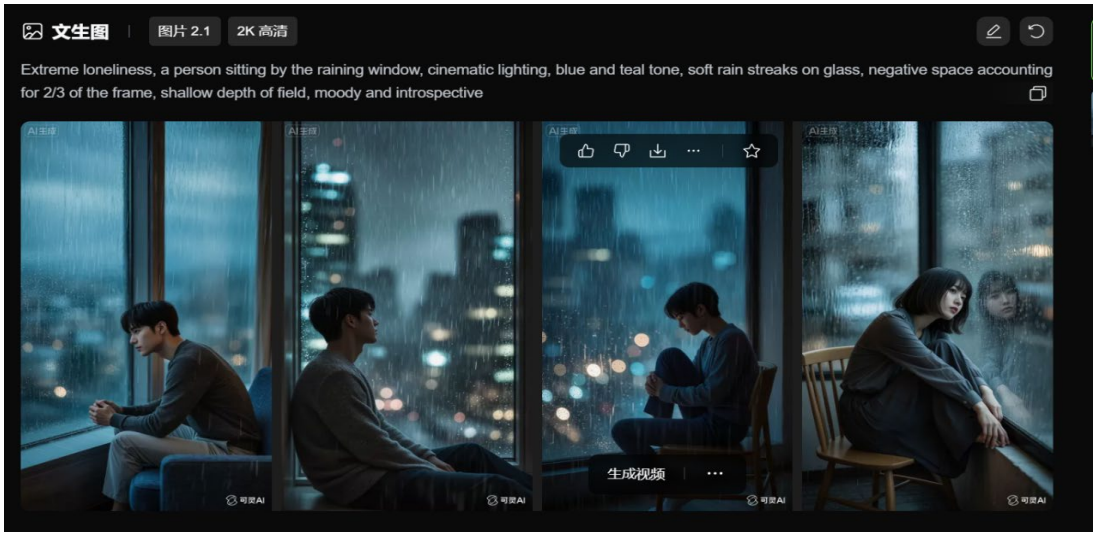


Figure 6. The second image on loneliness generated by Kling.

The second is the cross-cultural reconstruction and optimization of "warmth". To investigate the cross-cultural expression of the abstract emotion "warmth" by AI, this phase (V1) introduced a baseline prompt for a comparative experiment. The input prompt is set as: "Family reunion dinner, warmth" (家庭团圆饭，温暖). This foundational instruction was designed to strip away complex modifier parameters and purely examine the underlying training data encoding and cross-cultural decoding discrepancies among large language models from different backgrounds when confronted with highly culturally laden concepts, thereby establishing a comparative baseline for subsequent cultural symbol reconstruction and precise calibration. The AI-generated image results are as follows:

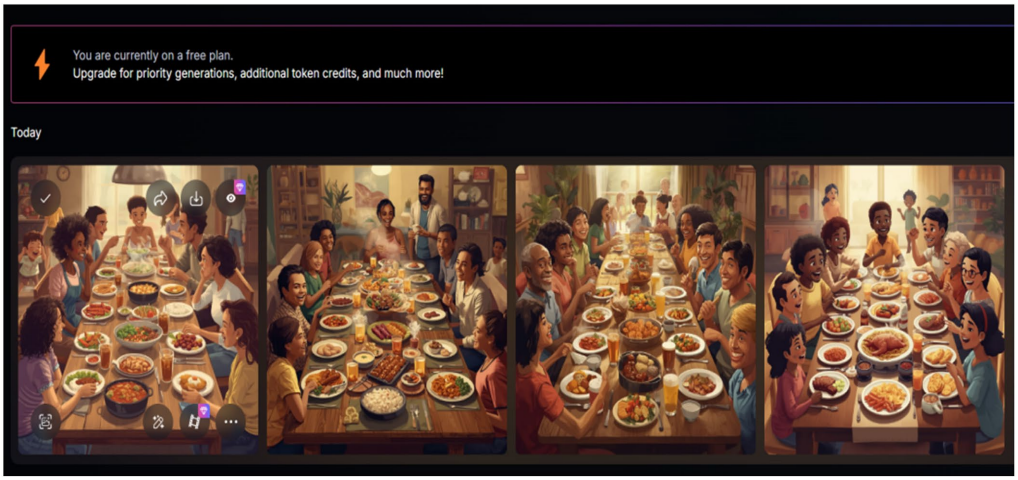


Figure 7. The first image on warmth generated by Leonardo.Ai.



Figure 8. The first image on warmth generated by Doubao.



Figure 9. The first image on warmth generated by Kling.

As evidenced by the baseline generation results in Figures 7-9, inputting the fundamental prompt "Family reunion dinner, warmth" led to distinct cross-cultural decoding divergences across models. Specifically: The international model Leonardo.Ai produced imagery dominated by large wooden rectangular tables, multi-ethnic family members, and Western-style or mixed meals (such as roasted chicken and salads) set within an open, brightly lit kitchen. These visual features reflect typical Western cultural paradigms, where the long table corresponds to

egalitarian interaction in Western households, multi-ethnic representation mirrors societal diversity, and Western-style food serves as the standard for domestic gatherings. In contrast, the domestic models Doubao and Kling manifested salient native cultural characteristics. Doubao generated round tables, traditional Chinese dishes (hot pot, dumplings, and whole fish), chopsticks, Fu-character lanterns, red tablecloths, and warm ambient lighting. These symbols collectively articulate Chinese traditional culture, where the round table signifies reunion and perfection, hot pot and dumplings represent traditional Spring Festival cuisine, and lanterns reinforce festive joviality. Similarly, Kling incorporated round tables, home-style Chinese dishes, and chopsticks, while integrating them with contemporary domestic settings and auspicious couplets, thereby highlighting the celebratory atmosphere of modern Chinese family reunions.

The root cause of this divergence lies in the culturally distinct encodings embedded within the training corpora of the respective models. Leonardo.Ai's training data strongly associates family reunions with Western symbols such as turkeys and fireplaces, whereas Doubao and Kling, operating within a Chinese linguistic and cultural context, deeply bind reunion dinners with indigenous symbols like round tables, dumplings, and hot pots. From a semiotic perspective, this variance fundamentally demonstrates that the signified emotion of "warmth" corresponds to entirely different visual signifiers across diverse cultural contexts.

Then, refine the prompt: "Chinese family reunion dinner, round table covered with red tablecloth, bamboo steamer full of dumplings, hot pot with rising steam, relatives chatting and smiling, warm yellow lighting, traditional Chinese New Year decorations, no turkey, no fireplace." (中式家庭团圆饭，铺着红桌布的圆桌，装满饺子的竹蒸笼，冒着热气的火锅，亲属谈笑风生，暖黄色灯光，中国传统新年装饰，无火鸡，无壁炉). The AI-generated image results are Figures 10-12:



Figure 10. The second image on warmth generated by Leonardo.Ai.



Figure 11. The second image on warmth generated by Doubao.

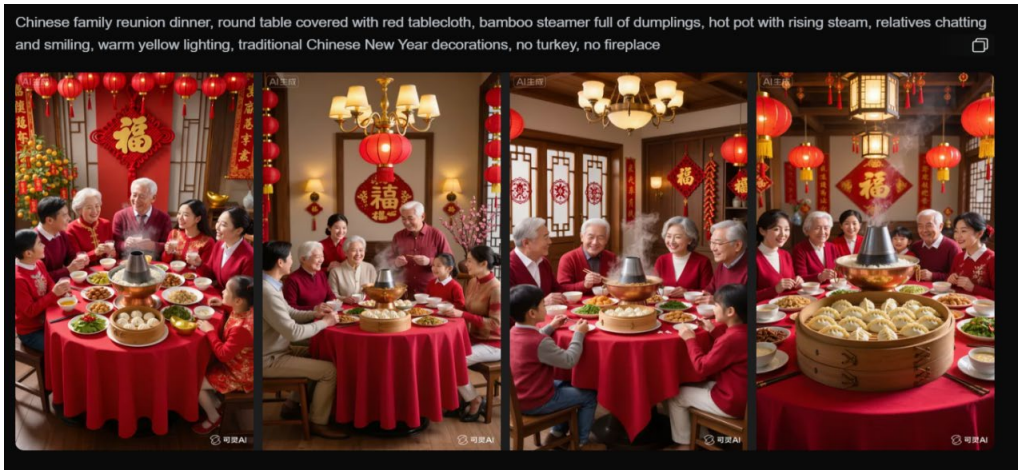


Figure 12. The second image on warmth generated by Kling.

The optimized prompt—achieved through the incorporation of core Chinese cultural symbols, traditional decorations, and negative constraints—accomplished a triple enhancement in cultural precision, elemental completeness, and atmospheric coherence. First, cultural bias was thoroughly rectified. By integrating indigenous symbols such as round tables, bamboo steamers, and Spring Festival decorations, alongside negative constraints that filtered out Western elements like turkeys and long tables, all evaluated models accurately anchored the cultural attributes of a Chinese family reunion dinner. Second, the matching degree of core elements achieved comprehensive coverage. Key components specified in the prompt—such as red tablecloths, rising hot pot steam, and steamed dumplings—were flawlessly rendered across the output results of all three models, successfully achieving precise mapping from textual instructions to visual imagery. Finally, the affective atmosphere was accurately established. The combination of yellow lighting and relatives chatting and smiling ensured that the images not only retained the festive vibrancy inherent to Chinese reunions but also effectively conveyed the tender emotion of "warmth", thereby avoiding any mechanical stiffness caused by the mere stacking of elements and successfully constructing a culturally resonant visual narrative of warmth.

5. Performance Evaluation Based on Visual Semiotics

Following the preliminary comparative experiments and iterative optimizations, this study integrates perspectives from semiotics and cognitive psychology to conduct an in-depth decoding of the visual outputs generated by the international model (Leonardo.Ai) and the domestic models (Doubao and Kling). Moving beyond superficial evaluations centered solely on aesthetic appeal, this research establishes a systematic evaluation framework to measure the emotional expression capabilities of artificial intelligence. To ensure the empirical rigor and objectivity of the evaluation, the assessment was conducted using a standardized quantitative semiotic coding procedure administered by two independent evaluators (two

postgraduates in visual semiotics). The coding scheme operated under explicit decision rules: effective signifiers were defined as concrete visual elements (e.g., color tones, lighting, negative space, cultural props, and postures) that directly map to the core emotional signified without contradiction, whereas interfering signifiers were elements that disrupted the target emotional atmosphere. Inter-Coder Reliability (ICR) was calculated using Cohen's Kappa and intraclass correlation coefficients (ICC), yielding high consistency ($\kappa = 0.86$, ICC = 0.91), with discrepancies resolved through consensus. Based on this methodology, the evaluation framework encompasses two core dimensions and subsequent comparative (empirical) analyses:

Core Dimension II: Cultural Adaptability and Innovation in Emotional Expression. Established to address the complexities of cross-cultural emotional expression, this dimension contains two sub-indicators: cultural symbol adaptability and symbol innovation freedom. Cultural adaptability examines whether artificial intelligence can select appropriate symbols based on the specific cultural context of the emotion, such as avoiding Western symbols like turkeys or fireplaces within a Chinese contextual framework for the theme of warmth. Symbol innovation freedom measures whether models can transcend rigid emotion-symbol linkages, such as attempting to portray loneliness through a static individual amidst a bustling crowd rather than merely relying on a monochromatic cold-toned scene. This evaluative dimension avoids technological determinism while highlighting the humanistic attributes inherent in AI emotional expression.

In testing the loneliness theme, international models such as Leonardo.Ai tended to generate a theatricalized sense of isolation. They often relied on exaggerated Tyndall lighting effects to forcibly create visual impact, imparting a deliberate sense of performative detachment. In contrast, Kling demonstrated superior immersive narrative capabilities. Regarding emotional symbol matching, Kling performed most prominently: its generated images constructed a complete semiotic chain, linking a blue-teal color tone (cold signifier) to alienation (loneliness signified), cinematic side-lighting outlining the human silhouette (lighting signifier) to individual presence, and two-thirds dark negative space (compositional signifier) to psychological vacancy. Kling achieved an effective signifier count of eight with zero invalid symbol interference. Conversely, while Leonardo.Ai and Doubao incorporated the core symbol of a rainy window, the additionally introduced burning candles (warm light signifier) conflicted with the signified loneliness, resulting in higher invalid interference. Regarding cultural adaptability and innovation, none of the three models exhibited apparent cultural bias, though they all fell into the stereotypical association that "cold colors equal loneliness". Nevertheless, Kling attempted to replace static cold tones with the dynamic blur of rain streaks on glass, showing slightly higher innovation.

In terms of warmth matching, Doubao and Kling demonstrated significant advantages. Doubao constructed a complete and precise semiotic chain that closely aligned with the emotional cognition of Chinese family reunions, linking a red

tablecloth (color signifier) to festiveness (warmth signified), dynamic hot-pot steam (detail signifier) to the vibrancy of daily life (*yanhuoqi* 烟火气), and an elder serving food (action signifier) to interpersonal bonding. Meanwhile, Leonardo.Ai, having corrected Western symbols such as turkeys and fireplaces, retained the long-table layout as a signifier while accurately conveying the signified Chinese-style barrier-free reunion. Regarding cultural adaptability and innovation, after prompt calibration, none of the three models displayed noticeable cultural bias. However, the domestic models Doubao and Kling generated more exquisite imagery that better resonated with domestic festive atmospheres, exhibiting higher innovation than the international model. These visual symbols are inherently high-context symbols deeply rooted in native cultural memory. Domestic models did not merely render physical warmth; rather, through the meticulous restoration of everyday details, they awakened the collective subconscious of the audience, thereby achieving instantaneous comprehension and emotional resonance in transmission efficiency.

6. Conclusions

Through visual narrative experiments focusing on the themes of loneliness and warmth, this study observes that within the limited scope of the three tested text-to-image systems (Leonardo.Ai, Doubao, and Kling) and the two prompted scenarios, Kling demonstrated high alignment between the input prompts and the generated visual symbols.

The empirical observations indicate that the tested domestic models effectively translated the specified Chinese cultural contexts into structured visual signifiers, exhibiting competitive performance in foundational rendering quality, lighting logic, and semantic-visual alignment under these specific conditions. These results suggest that advanced multimodal architectures can successfully instantiate abstract emotional concepts and cultural nuances through probabilistic pattern matching based on their training data, rather than indicating genuine emotional comprehension or subjective awareness.

In conclusion, while acknowledging the constraints of a limited sample size, a restricted set of prompts, and the absence of audience validation, this study empirically documents how Text-to-Image technology can achieve nuanced visual expression when guided by structured prompt engineering. Future research should incorporate broader datasets, diverse emotional categories, and empirical audience evaluations to further explore the intersection of global technological frameworks and localized aesthetic representations in AI-driven visual storytelling.

Acknowledgements

The authors would like to thank the 2027 Key Cultivation Project of Philosophy and Social Sciences of Hebei University: "A Study of the Construction and Recontextualized Dissemination Mechanisms of Intangible Cultural Heritage Emotional

Memory Discourse in Digital Spaces Based on NLP" (HBU2026SS007) for the support in this research.

Author Contributions

Yipu Kong: conceptualization, methodology, data collection original draft.

Zhengjie Guo: data analysis, review & editing, funding acquisition, project administration

All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., *et al*. (2020) Generative Adversarial Networks. *Communications of the ACM*, **63**, 139-144. https://doi.org/10.1145/3422622

[2] Liu, H.L. (2024) Generative AI and Knowledge Production. *Editorial Friend*, **3**, 5-13. (In Chinese)

[3] Yu, G.M. and Su, J.W. (2023) Communication Revolution and Media Ecology in the Context of Generative AI—From ChatGPT to a Comprehensively Intelligent Future. *Journal of Xinjiang Normal University* (*Edition of Philosophy and Social Sciences*), **44**, 81-90. (In Chinese)

[4] Gan, L.H. and Wang, H. (2024) From Emotional Projection to Digital Emotion: The Emotional Evolution of Human-Computer Interaction in Digital Landscapes. *Modern Publishing*, **3**, 27-38. (In Chinese)

[5] Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the* 2021 *ACM Conference on Fairness*, *Accountability*, *and Transparency*, Virtual Event, 3-10 March 2021, 610-623. https://doi.org/10.1145/3442188.3445922

[6] Su, T. and Peng, L. (2025) Communication and the New Human Condition in the Generative AI Era—An Overview of New Media Research in 2024. *Chinese Journal of Journalism & Communication*, **47**, 53-70. (In Chinese)

[7] Malchiodi, C.A. (2011) Handbook of art Therapy. Guilford Press.

[8] Westberg, G. and Kvåle, G. (2025) The Generic Uniqueness of AI Imagery: A Critical Approach to Dall-E as Semiotic Technology. *Discourse & Society*, **36**, 575-598. https://doi.org/10.1177/09579265241274788

[9] Strikovic, E. and Cools, H. (2025) Reality Re-Imag(in)Ed. Mapping Publics' Perceptions and Evaluations of Ai-Generated Images in News Contexts. *Digital Journalism*, **14**, 1301-1323. https://doi.org/10.1080/21670811.2025.2573076

[10] Ye, H. (2026) Algorithmic Reproduction of Occupational Stigma: Visual Prototype Lock-In in Chinese Text-to-Image AI Models. *Social Semiotics*, **36**, 1-30. https://doi.org/10.1080/10350330.2026.2659880

[11] Putland, E., Chikodzore-Paterson, C. and Brookes, G. (2023) Artificial Intelligence and Visual Discourse: A Multimodal Critical Discourse Analysis of AI-Generated Images of "Dementia". *Social Semiotics*, **35**, 228-253. https://doi.org/10.1080/10350330.2023.2290555

[12] Hart, C. (2025) Using Generative AI to Investigate the Multimodal Representation of Refugees and Migrants. *Discourse & Society*, **37**, 470-502. https://doi.org/10.1177/09579265251378665

[13] Barthes, R. (1977) Rhetoric of the Image. In: Heath, S., Ed. and Trans., *Image*, *Music*, *Text*, Fontana, 32-51.