
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202603.00022

A Survey of Research on Digital Art Design Based on World Models

Authors: Zeng Guokun, Chenxi Deng, Yufeng Huang, Guokun Zeng

Date: 2026-03-01T00:52:33+00:00

Abstract

Driven by the dual impetus of iterative advances in artificial intelligence technology and the rise of the digital art industry, the integration of AI world models and digital art has evolved from conceptual exploration to practical implementation, becoming a core driving force for the innovative development of the digital art industry. Through systematic research, this paper defines the core concepts and integration logic of AI world models and digital art, analyzes the current development status, core technical support, and mainstream application scenarios of their integration, examines the core challenges faced in the integration process at technical, artistic, copyright, talent, and industry levels, proposes targeted optimization paths based on industry development trends, and forecasts future development directions. The research indicates that the deep integration of AI world models and digital art requires coordinated efforts in technical adaptation, artistic integrity, copyright improvement, talent cultivation, and industry standardization. Such integrated development will engender new art forms and industrial ecosystems, injecting new momentum into the high-quality development of the digital art industry. This paper provides references for researchers, creators, and regulatory authorities in related fields, facilitating the benign integration and sustainable development of artificial intelligence technology and digital art.

Full Text

Preamble

Digital Media Art Design Based on World Models
Zeng Guokun¹, Deng Chenxi¹, Huang Yufeng²

Abstract

Driven by the dual forces of artificial intelligence technology iteration and the rise of the digital art industry, the integration of AI world models and digital art has evolved from conceptual exploration to practical implementation, becoming a core driver for innovative development in the digital art industry. Through systematic investigation, this paper defines the core concepts and fusion logic of AI world models and digital art, analyzes the current development status, core technological support, and mainstream application scenarios of their integration, examines the core challenges faced in terms of technology, art, copyright, talent, and industry, proposes targeted optimization paths based on industry development trends, and predicts future development directions. Research indicates that the deep integration of AI world models and digital art requires synergistic efforts in technological adaptation, artistic integrity, copyright improvement, talent cultivation, and industry standardization. This integration will catalyze new art forms and industrial ecosystems, injecting new momentum into the high-quality development of the digital art industry. This paper provides reference for researchers, creators, and regulatory authorities in related fields, facilitating the benign integration and sustainable development of AI technology and digital art.

Keywords: World Model; Generative Agent; Interactive Simulation; AIGC; Digital Art; Metaverse; NeRF

CLC Number: J0-05; TP183

As artificial intelligence technology continues to evolve, AI world models, serving as the core hub connecting AI perception, decision-making, and action, have not only broken through the development bottlenecks of general artificial intelligence with their core capability as “reality simulators,” but have also gradually penetrated into creative fields. Digital art, as a new art form utilizing digital technology for creation, display, and dissemination, has formed an industrial system encompassing diverse sectors such as digital painting, VR/AR art, and NFT art, becoming an important growth pole in the cultural and creative industries. According to relevant data, China’s digital art market scale exceeded 60 billion yuan in 2024, with a compound growth rate of 28% from 2020 to 2024, demonstrating rapid industrial development momentum [11].

Currently, the integration of AI world models and digital art has advanced from shallow combinations of AI-assisted artistic creation to deep synergy characterized by “world models constructing virtual art scenes and digital art enriching world model connotations.” In 2024, the AI-generated artwork *Neural Symphony* sold for \$1.2 million at Sotheby’s auction, becoming a typical case of this integration [16]; domestically, platforms such as Baidu’s “Wenxin Yige” and Alibaba’s “Tongyi Wanxiang” have successively launched AI digital art generation functions, promoting the localized implementation of fusion technologies [27]. However, the integration process still faces numerous issues including insufficient technological adaptation, controversies over artistic value, imperfect

copyright systems, and shortages of compound talents, constraining the depth and breadth of their integration [17][28].

Based on this context, this paper conducts systematic research to clarify the internal logic and current development status of the integration, analyze core challenges, and propose optimization paths, aiming to 梳理 the research and practical status of AI world model and digital art integration, provide reference for research and industrial implementation in related fields, and promote high-quality development of the digital art industry empowered by AI technology—this constitutes the purpose and significance of this study.

1. Research Objectives

1. Clarify the core concepts, development history, and core characteristics of AI world models and digital art, and 厘清 the internal logic and core value of their integration;
2. Investigate the current development status, mainstream application scenarios, and typical cases of AI world model and digital art integration, summarizing the advantages and shortcomings of existing fusion modes;
3. Analyze the core challenges faced in the integration process regarding technology, art, copyright, and talent, exploring the root causes of these problems;
4. Predict the future development directions of their integration, explore new fusion modes, and promote high-quality development of the digital art industry empowered by AI world model technology.

2.1 AI World Models

AI world models (World Models) serve as the core hub connecting artificial intelligence perception, decision-making, and action. Essentially, they construct an internal prediction system for agents, enabling them to predict future environmental evolution trends based on historical states and action inputs, thereby optimizing decision paths—making them the “reality simulators” of artificial intelligence [1]. The core mathematical expression is $s_{t+1} = f_{\theta}(s_t, a_t)$, where s_t represents the environment state set at time t , a_t represents the agent’s action input at time t , and f_{θ} represents the core function characterizing the world model [1].

A complete AI world model contains three core layers: state representation (transforming environmental information into structured representations processable by the model), action representation (adapting discrete or continuous action encoding for different scenarios), and dynamic architecture (modeling the mapping relationship between states and actions, with mainstream architectures including RNN, Transformer, and diffusion models) [1]. Its core characteristics manifest as sample efficiency (generating virtual training data to reduce

trial-and-error costs), strong generalization capabilities (learning general environmental laws for cross-scenario transfer), and decision autonomy (constructing a “perception-prediction-action” closed loop) [1]. Currently widely applied in game development, autonomous driving, and robotics, with technological iteration, they are gradually penetrating creative fields such as digital art [1].

AI world models (World Models) serve as the core hub connecting artificial intelligence perception, decision-making, and action. Essentially, they construct an internal prediction system for agents, enabling them to predict future environmental evolution trends based on historical states and action inputs, thereby optimizing decision paths—making them the “reality simulators” of artificial intelligence [1]. The core mathematical expression is:

$$s_{t+1} = f_{\theta}(s_t, a_t)$$

where s_t is the environment state set at time t , a_t is the agent’s action input at time t , f_{θ} is the core function characterizing the world model, and θ is the set of learnable parameters of the function [1].

To precisely describe the dynamic evolution characteristics of environmental states, a state error correction term is introduced to improve the prediction accuracy of world models. The corrected dynamic evolution formula is:

$$s_{t+1} = f_{\theta}(s_t, a_t) + \varepsilon_t \cdot \mathcal{N}(0, \sigma^2)$$

where ε_t is the error adjustment coefficient at time t , and $\mathcal{N}(0, \sigma^2)$ is Gaussian noise with mean 0 and variance σ^2 , used to simulate random disturbances in real environments and enhance the model’s adaptability to complex scenarios.

A simple AI world model contains three core layers: state representation, action representation, and dynamic architecture. The state representation layer transforms continuous environmental information into high-dimensional structured feature vectors. If implemented using MLP and CNN, the transformation formula is:

$$z_t = \mathcal{E}(s_t) = \text{MLP}(\text{CNN}(s_t))$$

where z_t is the feature vector representation of the environment at time t , $\mathcal{E}(\cdot)$ is the feature extraction mapping function, $\text{CNN}(\cdot)$ is the convolutional neural network completing local feature extraction, and $\text{MLP}(\cdot)$ is the multi-layer perceptron implementing high-dimensional feature mapping.

The action representation layer completes the encoding and normalization of actions for discrete/continuous action requirements in digital art creation. The formula is:

$$a_t = \frac{\mathcal{A}(a_t) - \mu_A}{\sigma_A}$$

where a_t is the normalized action feature vector, $\mathcal{A}(\cdot)$ is the action encoding function, and μ_A and σ_A are the mean and standard deviation of the action dataset, respectively.

Figure 1

Figure 1: Figure 1

Furthermore, in more advanced world models, the dynamic architecture layer can achieve more complex non-linear mappings between states and actions. Taking the Transformer architecture as an example, its core computational formula is:

$$\begin{aligned} h_{t+1} &= \text{Transformer}(z_t \oplus a_t) \\ \hat{s}_{t+1} &= \mathcal{D}(h_{t+1}) \end{aligned}$$

where \oplus is the feature concatenation operation, h_{t+1} is the hidden layer output of the Transformer, $\mathcal{D}(\cdot)$ is the decoding function realizing the restoration of hidden layer features to environmental states, and \hat{s}_{t+1} is the environment state predicted by the model.

World model training aims to minimize the error between predicted states and real states, with the loss function defined as:

$$\mathcal{L}_{\text{WM}}(\theta) = \frac{1}{T} \sum_{t=1}^T \|\hat{s}_{t+1} - s_{t+1}\|_2^2 + \lambda \|\theta\|_2^2$$

where T is the training sequence length, $\|\cdot\|_2^2$ is the L2 norm, and λ is the L2 regularization coefficient used to prevent model overfitting.

Its core characteristics manifest as sample efficiency (generating virtual training data to reduce trial-and-error costs), strong generalization capabilities (learning general environmental laws for cross-scenario transfer), and decision autonomy (constructing a “perception-prediction-action” closed loop) [1]. Currently widely applied in game development, autonomous driving, and robotics, with technological iteration, they are gradually penetrating creative fields such as digital art [1].

Figure 1 illustrates the operating principle of a four-layer architecture general world model: the input layer receives environmental state data and action command inputs; the core layer transforms input information into structured representations understandable by the system through state representation modules, dynamic architecture modules, and action representation modules; the processing layer conducts trend prediction through prediction modules and achieves strategy optimization through feedback adjustment modules; the output layer generates virtual scenes, outputs decision suggestions, or generates virtual data based on processing results, forming a complete closed loop from perception, understanding, decision-making to interaction.

Figure 2

Figure 2: Figure 2

2.2 Development Stages of AI World Model and Digital Art Integration

Digital art refers to art forms utilizing digital technology as creative tools, display carriers, and dissemination media, integrating artistic aesthetics and technical logic to achieve creative expression and emotional transmission [2]. Compared with traditional art, it possesses core advantages including near-zero marginal replication costs, fast online dissemination speeds (over 90% faster than traditional art), strong interactivity, and high modifiability [2]. The quantitative model for digital art's artistic value can be expressed as:

$$V = \alpha \cdot C + \beta \cdot I + \gamma \cdot E - \delta \cdot H$$

where V is the comprehensive artistic value score of digital art works, C is the creativity uniqueness index, I is the interactive experience index, E is the emotional expression index, H is the homogenization degree index, and $\alpha, \beta, \gamma, \delta$ are weight coefficients satisfying $\alpha + \beta + \gamma + \delta = 1$ and $\alpha, \beta, \gamma, \delta \in [0, 1]$. Domestic scholars classify digital art into “technology-driven” and “creativity-dominated” categories, providing a localized research perspective for their integration with AI world models [27]. Technology-driven digital art satisfies $\alpha < \beta$, while creativity-dominated digital art satisfies $\alpha > \beta$.

Figure 2 outlines the complete context of AI and digital art integration development. The development history of AI digital art can be clearly divided into four stages. During the embryonic period (before 2016), the combination of early AI algorithms and digital art basic tools laid preliminary technical foundations for AI-art integration. Although explorations in this stage appeared immature, they planted important seeds for subsequent development. Entering the startup period (2016-2020), the iterative upgrading of GAN technology and the rise of digital art platforms became key turning points. AI first gained high-quality image generation capabilities, and the popularization of professional creation platforms accelerated the dissemination and exchange of digital art, pushing the field from concept to practice. During the rapid development period (2021-2024), breakthroughs in diffusion models, multimodal pre-training, and large model training and application completely revolutionized the quality and diversity of AI-generated content. Multimodal large models enabled AI to understand and generate various contents including text, images, and audio, ushering AI art creation into an explosive growth phase. Looking ahead (2025 and beyond), as large model training technology deepens, multimodal AIGC tools popularize, and “world models” are constructed, AI digital art will evolve toward deep AI-art integration, personalized creation, and virtual-real interaction. AI will no longer be merely a creation tool but deeply integrated into the entire creation process, achieving richer interactions with the real world and creating brand

Figure 3

Figure 3: Figure 3

new art experiences.

The stage maturity evaluation formula for integrated development is:

$$M = \frac{1}{n} \sum_{i=1}^n (\eta_T \cdot T_i + \eta_S \cdot S_i + \eta_C \cdot C_i)$$

where M is the stage maturity ($M \in [0, 1]$), n is the number of evaluation indicator dimensions, T_i is the technology landing degree, S_i is the scenario expansion degree, C_i is the commercialization degree, and η_T, η_S, η_C are weights for each indicator satisfying $\eta_T + \eta_S + \eta_C = 1$.

2.3 Core Logic of Integration

The integration of AI world models and digital art is essentially a bidirectional empowerment and symbiotic collaboration between “technical tools and creative expression.” On one hand, AI world models provide efficient, intelligent creation tools and scene carriers for digital art. By constructing logically consistent, dynamically evolving virtual worlds, they break through spatial, temporal, and technical limitations of traditional digital art creation, lowering creation thresholds and improving efficiency—for example, architectural designers can quickly generate 100 conceptual schemes using AI world models, extracting disruptive spatial structures from them [16]. On the other hand, digital art injects creative connotations and emotional value into AI world models, preventing them from falling into the dilemma of “technological hollowness,” making the virtual worlds they construct more humanistic and artistically infectious, achieving a balance between “technical rationality” and “artistic sensibility” [1][10][27].

The synergistic empowerment efficiency formula for their integration is:

$$\eta = \eta_{T \rightarrow A} \times \eta_{A \rightarrow T} = \frac{O'_A - O_A}{O'_T - O_T}$$

where η is the integration synergy efficiency, $\eta_{T \rightarrow A}$ is the empowerment efficiency of technology to art, $\eta_{A \rightarrow T}$ is the feedback efficiency of art to technology, O_A, O'_A are the creative output efficiencies of digital art before and after integration, and O_T, O'_T are the scene construction capabilities of world models before and after integration.

Figure 3 presents the application architecture diagram of world model and digital art integration, showing the complete technical chain from real perception to immersive experience. The overall performance evaluation formula for the fusion architecture is:

$$P = \left(1 - \frac{\tau}{\tau_{\max}}\right) \times \left(1 - \frac{\epsilon}{\epsilon_{\max}}\right) \times \frac{F}{F_{\text{std}}}$$

where P is the comprehensive architecture performance score ($P \in [0, 1]$), τ is the data processing delay, τ_{\max} is the maximum tolerable delay, ϵ is the error rate of generated content, ϵ_{\max} is the maximum tolerable error rate, F is the artistic feature similarity of integrated content, and F_{std} is the standard artistic feature similarity.

First, the perception layer serves as the system's information entrance, completing data collection and environmental perception through sensors, computer vision, and other technologies, providing real, multimodal real-world data for subsequent modeling. This data is input to the AI world model layer through "data interaction." This layer serves as the core "brain," relying on deep learning, generative adversarial networks, and other key technologies to learn from massive data and build models, constructing AI models capable of simulating and deducing virtual worlds, and providing core capabilities for upper-layer creation through "model driving."

Supported by the model, the digital art creation layer undertakes content generation and interaction design functions, transforming the abstract capabilities of AI models into perceivable digital art works through digital painting, virtual simulation, and other technologies, achieving the transformation from technical capability to artistic expression. Finally, these art contents enter the application layer through "content output," relying on AR/VR, digital twin, and other technologies, applied to art display, education, entertainment, and other scenarios in immersive, interactive ways, allowing users to experience the charm of AI-empowered digital art firsthand.

Overall, this architecture forms a closed loop of "perception—modeling—creation—application," both improving the creation efficiency and expressiveness of digital art and expanding its application boundaries, making artificial intelligence the core driving force for digital art innovation and development.

3.1 Overall Development Status

Currently, the integration of AI world models and digital art has entered a rapid development period, characterized by "technology iteration driving, continuous scenario expansion, and accelerated commercialization landing." From the industrial scale perspective, the rapid growth of the digital art industry provides broad space for integrated development. In 2025, the industry's total revenue is expected to hit 85 billion yuan, with the proportion of AI-related digital art output value continuously increasing [2][11][27]. From the technical level, breakthroughs in generative models, computer vision, and natural language processing have promoted the integration from shallow assistance to deep synergy [1][4][26]. From the participant perspective, multiple parties including internet technology enterprises, digital art studios, university research institutions, and independent creators participate, forming a diversified integration ecology—leading technology enterprises focus on R&D, while small and medium institutions and creators focus on scenario landing and creative expression [2][11][27].

Figure 4

Figure 4: Figure 4

Figure 4

Figure 5: Figure 4

Domestically, leading enterprises such as Tencent, Baidu, and Alibaba have successively laid out plans, forming a localized integration ecology of “technology R&D—scenario landing—commercialization monetization” [27].

3.2 Core Integration Technology Support

Their integration relies on collaborative support from multi-domain technologies, corely concentrated on the adaptation and integration of AI world model-related technologies and digital art creation technologies, mainly including the following four categories as shown in

:

3.2.1 Generative Model Technology

As the core engine of their integration, it learns from massive artistic data, masters the internal laws of style, composition, and color, and realizes autonomous generation or assisted creation of artistic works [1]. Mainstream technical paths include diffusion models (Diffusion Model), generative adversarial networks (GAN), and multimodal large models. Among them, the core forward diffusion process formula of diffusion models is:

$$x_t = \sqrt{1 - \beta_t} \cdot x_{t-1} + \sqrt{\beta_t} \cdot \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, I)$$

where x_0 is the original artistic image, x_t is the noisy image after t steps of diffusion, $\beta_t \in [0, 1]$ is the diffusion coefficient, ϵ_t is Gaussian noise, and I is the identity matrix.

The reverse generation process (denoising process) formula is:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2)$$

where $\mu_\theta(x_t, t)$ is the denoising mean predicted by the model, obtained through neural network fitting, and σ_t^2 is the variance of the denoising process.

The loss function of diffusion models is:

$$\mathcal{L}_{\text{DM}}(\theta) = \mathbb{E}_{x_0, \epsilon_t, t} [\|\epsilon_\theta(x_t, t) - \epsilon_t\|^2]$$

where $\epsilon_\theta(x_t, t)$ is the noise predicted by the model, and T is the total number of diffusion steps.

The objective function of Generative Adversarial Networks (GAN) is:

$$\mathcal{L}_{\text{GAN}}(G, D) = \mathbb{E}_{x \sim p_{\text{data}}}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]$$

where G is the generator responsible for generating artistic works, D is the discriminator responsible for distinguishing between real artistic works and generated works, p_{data} is the distribution of real artistic data, p_z is the random noise distribution, and $\mathbb{E}[\cdot]$ is the expectation operation.

For digital art style transfer, the cycle consistency loss formula of CycleGAN is:

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_X}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_Y}[\|G(F(y)) - y\|_1]$$

where $G : X \rightarrow Y$ is the generator from style X to style Y , $F : Y \rightarrow X$ is the inverse mapping generator, and $\|\cdot\|_1$ is the L1 norm, ensuring the reversibility and consistency of style transfer.

The feature fusion formula for cross-modal generation realized by multimodal large models is:

$$\mathcal{F}_{\text{fusion}} = \text{Attention}(\mathcal{F}_T, \mathcal{F}_I) = \text{softmax}\left(\frac{\mathcal{F}_T \mathcal{F}_I^T}{\sqrt{d_k}}\right) \mathcal{V}$$

where $\mathcal{F}_{\text{fusion}}$ is the fused feature, \mathcal{F}_T is the text feature vector, \mathcal{F}_I is the image feature vector, d_k is the feature dimension, and \mathcal{V} is the value matrix, achieving attention fusion of text and image features.

Domestic scholars have achieved staged progress in localized optimization of diffusion models, with generative models adapted to Chinese creative instructions gradually landing in applications [26].

3.2.2 Computer Vision Technology

Computer vision technology provides core support for artificial intelligence to “understand art,” with key applications covering style extraction, image segmentation, composition analysis, and color optimization. The feature extraction formula for CNN and ViT fusion is:

$$\mathcal{F}_{\text{total}} = \alpha \cdot \mathcal{F}_{\text{CNN}} + (1 - \alpha) \cdot \mathcal{F}_{\text{ViT}}$$

where $\mathcal{F}_{\text{total}}$ is the fused artistic feature vector, \mathcal{F}_{CNN} is the local detail features extracted by CNN, \mathcal{F}_{ViT} is the global layout features extracted by ViT, and α is the fusion weight ($\alpha \in [0, 1]$), which can be adaptively adjusted according to artistic work types.

The color similarity calculation formula for artistic works is:

$$\text{Sim}_C(A, B) = 1 - \frac{1}{3} \sum_{c \in \{R, G, B\}} |\mu_{A,c} - \mu_{B,c}|$$

where $\text{Sim}_C(A, B)$ is the color similarity between works A and B ($\text{Sim}_C \in [0, 1]$), and $\mu_{A,c}, \mu_{B,c}$ are the pixel means of the two works in the $R/G/B$ channels, respectively.

The composition balance calculation formula is:

$$\text{Bal} = 1 - \frac{|S_L - S_R|}{S_L + S_R}$$

where Bal is the composition balance degree ($\text{Bal} \in [0, 1]$), and S_L, S_R are the pixel feature sums on the left and right sides of the artwork, respectively.

Related research indicates that photography AI tools built based on computer vision (integrating CNN and ViT technologies) can enhance the artistic sense of ordinary photos by 35% [20]. The artistic sense improvement rate formula is:

$$\rho = \frac{V' - V}{V} \times 100\%$$

where ρ is the artistic sense improvement rate, and V, V' are the artistic value scores of the work before and after optimization, respectively.

3.2.3 Interaction Technology

Deep learning large models provide core support for the experience upgrade of digital art integration works. Their interaction technology innovation drives works from “static display” to “dynamic experience,” corely covering somatosensory interaction, real-time feedback, and VR/AR fusion [1]. The action matching degree formula for somatosensory interaction is:

$$\text{Match} = \frac{1}{k} \sum_{i=1}^k \cos \theta_i = \frac{1}{k} \sum_{i=1}^k \frac{m_i \cdot n_i}{\|m_i\| \|n_i\|}$$

where Match is the action matching degree ($\text{Match} \in [-1, 1]$), k is the number of joint points, m_i is the captured human body joint point vector, n_i is the action joint point vector in the artistic work, and θ_i is the angle between the two vectors.

The dynamic adjustment formula for real-time feedback is:

$$\mathcal{O}' = \mathcal{O} + \gamma \cdot \mathcal{F}(\mathcal{Q})$$

where \mathcal{O}' is the adjusted work output feature, \mathcal{O} is the original output feature, γ is the feedback adjustment coefficient, and $\mathcal{F}(\mathcal{Q})$ is the adjustment function based on user feedback features \mathcal{Q} .

With the image recognition and motion capture capabilities of models, human actions captured by devices such as Kinect can be transformed into artistic works in real-time; models can also recognize audience expressions and voice feedback, dynamically adjusting work presentation effects to achieve personalized expression [1][27].

3.2.4 Virtual Scene Construction Technology

Based on the dynamic architecture of AI world models, combined with technologies such as Neural Radiance Fields (NeRF), high-fidelity, interactive, and dynamically evolving virtual art scenes are constructed [1]. The volume rendering formula of NeRF is:

$$C(r) = \int_{t_n}^{t_f} T(t) \sigma(r(t)) c(r(t), d) dt$$

$$T(t) = \exp \left(- \int_{t_n}^t \sigma(r(s)) ds \right)$$

where $C(r)$ is the rendering color of ray $r(t) = o + td$, t_n, t_f are the near and far clipping planes of the ray, $\sigma(r(t))$ is the volume density of 3D point $r(t)$, $c(r(t), d)$ is the view-dependent color of this point along direction d , and $T(t)$ is the transmittance of the ray from t_n to t .

This realizes the scene-based presentation and immersive experience of digital art works. For example, Monet's *Water Lilies* can be transformed into a dynamic virtual scene, allowing viewers to gain an immersive artistic experience [1][5][13]. The immersion evaluation formula for virtual scenes is:

$$\text{Imm} = \omega_1 \cdot \text{Res} + \omega_2 \cdot \text{Int} + \omega_3 \cdot (1 - \tau)$$

where Imm is the immersion score ($\text{Imm} \in [0, 1]$), Res is the scene resolution normalization value, Int is the scene interactivity index, τ is the interaction delay normalization value, and $\omega_1, \omega_2, \omega_3$ are weight coefficients.

3.3 Mainstream Application Scenarios and Typical Cases

As technology continues to mature, their integration has penetrated multiple 细分 scenarios of digital art, forming a batch of representative application cases covering the entire process of creation, display, dissemination, and experience. Among them, domestic cases have formed distinct localized characteristics.

3.3.1 Film and Animation Assisted Creation Scenario

This is currently the most technically mature and widely implemented AI-art creation integration scenario. AI world models rely on deep learning from massive artistic materials, classic art styles, lighting laws, and spatial structures to form understandings of aesthetic laws and creation logic. Combined with creators' creative instructions, style preferences, and theme requirements, they quickly output high-quality art drafts, scene settings, character concepts, color schemes, and composition schemes. Through human-machine collaboration, they significantly lower the technical threshold for creative landing, shorten the cycle from inspiration conception to visual presentation, and notably improve film and animation creation efficiency and iteration speed.

The efficiency improvement formula for film and animation creation is:

$$\eta_{\text{eff}} = \frac{T_{\text{trad}} - T_{\text{AI}}}{T_{\text{trad}}} \times 100\%$$

where η_{eff} is the creation efficiency improvement rate, T_{trad} is the cycle of traditional creation mode, and T_{AI} is the cycle of human-machine collaborative creation.

In the film and animation field, this technology has achieved benchmark-level applications. Disney officially introduced AI world models to assist scene design and visual development in the production of *Frozen 2*: based on plot requirements, the model quickly generated multiple versions of high-fidelity virtual scene schemes, covering key environments such as forests, canyons, and magical secret realms, providing diverse choices for lighting, materials, and atmosphere. The art team conducted aesthetic control, detail optimization, and artistic adjustments on this basis, greatly compressing the originally lengthy scene exploration process. Practice data shows that this mode improved overall creation efficiency by 3 times ($\eta_{\text{eff}} \approx 66.7\%$), both guaranteeing Disney's consistent artistic standards and providing a replicable path for industrialized efficient production of complex animation projects [1][16].

3.3.2 Virtual Art Scene and Immersive Experience Scenario

AI world models construct high-fidelity, dynamic virtual worlds, integrating digital art works into them, combined with VR/AR technology to provide users with immersive art experiences, breaking through the spatial limitations of traditional art display.

The user experience scoring formula for virtual exhibition halls is:

$$U = 0.4 \cdot \text{Imm} + 0.3 \cdot \text{Fre} + 0.3 \cdot \text{Nov}$$

where U is the comprehensive user experience score ($U \in [0, 1]$), Imm is immersion, Fre is operation freedom, and Nov is experience novelty.

Typical case: A certain digital art exhibition uses AI world models to construct virtual art exhibition halls, integrating AI-generated art, digital paintings, and digital sculptures into virtual scenes. Users can “walk into” the exhibition hall through VR devices, freely browse and touch virtual works, and works can produce dynamic changes based on user interaction behaviors. Meanwhile, AI world models can simulate different lighting and atmospheres, allowing the same work to present diverse artistic effects. Furthermore, the application of 5G+8K cloud rendering technology has compressed virtual scene delays to within 15ms, further enhancing the immersive experience [2][13][24]. Domestic case: The Palace Museum's “Digital Cultural Relics Library” uses AI world models to construct virtual exhibition halls, transforming ancient paintings and calligraphy into dynamic virtual scenes, combined with VR devices allowing viewers to “travel through” to ancient courts and experience the unique charm of traditional art [27].

3.3.3 NFT Art and Digital Collectibles Scenario

AI world models generate unique digital art works, combined with blockchain technology to create NFT art collectibles with uniqueness and immutability, promoting the commercialization landing of digital art while solving the problems of low creation efficiency and serious homogenization in NFT art creation.

The scarcity quantification formula for NFT art is:

$$S = \frac{1}{N} \sum_{j=1}^N \left(\frac{1}{m} \sum_{i=1}^m (1 - \text{Sim}_i) \right)$$

where S is the scarcity index ($S \in [0, 1]$), N is the issuance quantity of the same NFT, m is the feature comparison dimension, and Sim_i is the similarity between the i -dimensional feature and similar works.

Currently, leading platforms such as Alibaba Whale Discovery, JD Lingxi, and Baidu Super Chain occupy 66% of NFT transaction volume, among which many NFT collectibles are unique digital art works generated based on AI world models [2][11][27]. In 2024, the unit price of AI-generated art in the auction market has risen to 58,000 yuan per piece, 2.4 times that of 2022 [2]. The price growth model is:

$$P(t) = P_0 \cdot (1 + g)^{t-t_0}$$

where $P(t)$ is the work unit price in year t , P_0 is the unit price in the base year t_0 , g is the average annual growth rate, and $g \approx 65\%$ for 2022-2024. This data fully demonstrates market recognition of the value of AI-generated digital art [10]. Domestic case: Alibaba Whale Discovery cooperated with intangible cultural heritage inheritors, using AI world models to generate intangible cultural heritage-themed NFT collectibles, both preserving the core characteristics of intangible cultural heritage techniques and achieving the youthful dissemination of traditional culture [27].

3.3.4 Dynamic Visual Design Scenario

AI world models can generate dynamic visual effects, character actions, scene evolution, and interaction logic in real-time, deeply supporting the creation and operation of interactive games, open worlds, and real-time rendering digital content, significantly improving interactive experience, shortening R&D cycles, and reducing development costs. The probability formula for game scene dynamic evolution is:

$$P(s_{t+1}|s_t, a_p) = \frac{\exp \phi(s_t, a_p, s_{t+1})}{\sum_{s \in \mathcal{S}} \exp \phi(s_t, a_p, s)}$$

where $P(s_{t+1}|s_t, a_p)$ is the probability of the scene evolving from state s_t to s_{t+1} after the player executes action a_p , $\phi(\cdot)$ is the scoring function for scene evolution, and \mathcal{S} is the set of all possible scene states.

Typical case: Game manufacturers rely on AI world models to build high-freedom open worlds. Models can autonomously drive scene dynamic evolution

(such as real-time weather systems, day-night light alternation, and NPC autonomous behavior logic), and dynamically adjust level difficulty, scene states, NPC interaction feedback, and even plot branch directions based on player real-time operations, choices, and behavior data, constructing highly immersive, high-freedom interactive gaming experiences. Meanwhile, models can generate and drive character actions, expressions, and demeanor changes in real-time, greatly reducing the manual production volume of traditional keyframe animation [1][18][27].

3.4 Summary of Existing Integration Modes

Combining current application status, the integration of AI world models and digital art has mainly formed three core modes, each with advantages and applicable scenarios, as shown in the following table:

Integration Mode	Core Characteristics	Applicable Scenarios	Advantages	Disadvantages
AI-Assisted	AI world model serves as creation tool, assisting creators in completing drafts, scenes, details, etc. Creator dominates creativity	Digital painting, film and animation, visual design	Lowers creation threshold, improves efficiency, preserves creator's artistic creativity	Limited AI autonomy, insufficient understanding of complex creativity, prone to homogenization [1][23][29]
AI-Autonomous	AI world model generates complete digital art works or virtual worlds based on preset instructions and training data	NFT art, virtual exhibition halls, short video materials	Extremely high creation efficiency, can generate large quantities of unique works, suitable for commercial batch production	Lacks human emotion and creative depth, artistic value controversial, prone to copyright disputes [1][10][14][28]

Integration Mode	Core Characteristics	Applicable Scenarios	Advantages	Disadvantages
Human-Machine Symbiosis	Creator and AI world model collaborate, AI handles technical implementation and dynamic optimization, creator handles creative guidance and artistic control	Immersive exhibitions, game virtual worlds, high-end digital art	Balances efficiency and creativity, achieves balance between technology and art, high work quality	High technical literacy requirements for creators, difficult technology-art adaptation [1][17][29]

4.1 Technical Level: Insufficient Adaptability and Stability

- Insufficient Technical Adaptability:** The core advantage of AI world models lies in virtual world construction and dynamic evolution, while digital art emphasizes creative expression and emotional transmission. There are obvious differences in technical logic between the two, causing some AI world model technologies to be difficult to adapt to the personalized creation needs of digital art [1]. For example, AI world models' learning of abstract art and niche art styles is not deep enough, making it difficult to generate works meeting creator expectations. Meanwhile, generative models still have considerable deficiencies in detail control and emotional expression, making it difficult to reach the creation level of professional artists [1][4][18][26]. Domestic research points out that localized AI models still have deviations in understanding Chinese creative instructions, and their ability to adapt to local art styles such as Guochao (China-chic) and intangible cultural heritage needs further improvement [26].
- Insufficient Stability and Controllability:** Currently, AI world model-generated digital art works still have strong uncertainty, possibly presenting logical confusion, style deviation, and detail distortion, making it difficult for creators to precisely control the final presentation effect of works [1]. Furthermore, technical bottlenecks are prominent: GPU rendering efficiency is insufficient, 4K scene frame rates are often below 30fps, and device compatibility presents a fragmented state, seriously affecting the experience effect of integrated works [2][24][26].
- High Technical Threshold:** The R&D and deployment costs of high-end AI world models are high, making them difficult for small and medium digital art studios and independent creators to afford, resulting in insufficient popularization of integration technology [2]. Meanwhile, the human-machine collaboration mode has high technical literacy requirements for

creators, requiring creators to proficiently master AI tool usage methods and understand the technical logic of world models. Currently, most art creators lack relevant technical capabilities, constraining the widespread promotion of integration modes [2][9][17][29]. Domestic small and medium studios face more prominent technical cost pressures, and the supply of lightweight localized AI tools is relatively insufficient [29].

4.2 Artistic Level: Value Controversies and Homogenization Issues Prominent

- **Artistic Value Controversies:** Some viewpoints argue that AI world model-generated digital art works lack creators' emotional investment, ideological expression, and artistic accumulation, being merely products of technical algorithms without true artistic value [1]. Meanwhile, AI-generated works are mostly based on learning and imitation of massive existing artistic materials, making it difficult to form unique artistic styles, resulting in low artistic recognition of works, thereby triggering widespread controversy over “whether AI creation belongs to true art” [1][10][29]. Domestic scholars believe that the core of the value controversy over AI-generated art lies in the boundary definition between “technical tool” and “artistic subject,” requiring the construction of scientific value evaluation systems combined with local artistic contexts [29].
- **Prominent Homogenization Phenomenon:** Because AI world model training data mostly concentrates on mainstream art styles and classic works, generated digital art works are prone to homogenization issues, lacking innovation and uniqueness [1]. Especially in commercial scenarios such as NFT art and short video materials, batch-generated works mostly have similar compositions, colors, and styles, making it difficult to meet market demands for personalized artistic works [1][23][29]. The homogenization problem in the domestic NFT market is more prominent. Some platforms, in pursuit of traffic, batch-generate low-quality works, seriously disturbing market order [27].
- **Ambiguous Creative Subjectivity:** In AI autonomous generation mode, AI becomes the core subject of creation, weakening the creator's role and causing ambiguous creative subjectivity of digital art works [1]. In human-machine collaboration mode, the balance between technology and art is difficult, prone to the tendency of “emphasizing technology, neglecting art,” making works fall into the 误区 of “technical showing off,” lacking emotional connotations and humanistic value [1][10][29].

4.3 Copyright Level: Imperfect Systems and Prominent Infringement Risks

- **Unclear Copyright Attribution:** Current laws have no clear definition of copyright attribution for AI world model-generated digital art works—

whether they belong to model developers, work generators, or belong to the public domain remains controversial [1]. Especially in scenarios where multiple people collaborate using AI world models, copyright attribution is more difficult to divide, easily triggering copyright disputes [1][6][14][28]. China's Copyright Law has not yet clarified the copyright subject of AI-generated works, and relevant judicial interpretations are still being further improved [28].

- **High Infringement Risks:** The training process of AI world models requires massive citation of existing digital art works and traditional art works as training data. If without copyright holder authorization, this easily constitutes infringement [1]. Meanwhile, AI-generated works may be highly similar to existing works, triggering plagiarism controversies. In 2024, copyright infringement cases in the digital art field reached 12,000, up 37% year-on-year. Infringement issues have become important factors constraining integrated development [2][6][14][28]. Domestic infringement cases mainly concentrate on training data infringement and work plagiarism, with cases of AI-generated works plagiarizing traditional intangible cultural heritage works increasing year by year [28].
- **Imperfect Copyright Protection Systems:** Digital art itself has characteristics of easy replication, easy dissemination, and difficult tracing. Combined with AI technology, copyright protection difficulty is further increased [1]. Although blockchain certification work volume increases 170% annually, due to lack of standards, cross-chain mutual recognition rates are below 30%, making it difficult to achieve effective protection of AI-generated digital art works [2]. Meanwhile, copyright monitoring and rights protection technical means are relatively lagging, making it difficult to quickly discover and investigate infringement acts [2][6][25][28]. Domestic blockchain copyright certification platforms lack unified standards, rights protection processes are complex and costly, which is not conducive to efficient resolution of copyright disputes [28].

4.4 Talent Level: Huge Gap in Compound Talents

The integration of AI world models and digital art requires compound talents who both master AI technology (world model R&D, AI tool usage) and possess solid artistic skills and creative capabilities [1]. However, such talents are currently severely scarce in the industry, forming an embarrassing dilemma of “technical talents don't understand art, artistic talents don't understand technology” [2][17][29].

From the talent supply perspective, university talent training systems are relatively lagging. Most AI-related majors focus on technology R&D while neglecting artistic cultivation. Digital art majors focus on artistic creation but lack AI technology-related curriculum settings, resulting in cultivated talents being difficult to adapt to integrated development needs [2][17][29]. Although domestic

universities have gradually opened interdisciplinary courses, there are still problems such as disjointed curriculum settings and insufficient practice links [29]. From the industry demand perspective, the compound talent gap is estimated to reach 120,000 people, especially for high-end creative technical talents and AI art directors, where the gap is more prominent. This status quo constrains the depth and breadth of their integration [2][17][29].

4.5 Industry Level: Lack of Standards and Imbalanced Commercialization

- **Lack of Industry Standards:** Currently, the field of AI world model and digital art integration lacks unified industry standards and norms, with no clear requirements for quality evaluation, art certification, and technical adaptation of AI-generated works [1]. Meanwhile, industry self-discipline is insufficient. Some enterprises and creators, in pursuit of commercial interests, batch-generate low-quality, homogeneous works, disturbing industry order [1][23][29]. China has not yet formed unified AI digital art quality evaluation standards, and industry self-discipline mechanisms also need improvement [29].
- **Imbalanced Commercialization Development:** Some integration scenarios (such as NFT art) excessively pursue commercialization, neglecting the value of art itself, resulting in inflated work prices and serious bubbles [2]. Some integrated works with artistic value, due to lack of effective commercialization channels, are difficult to achieve market value, thereby affecting creator enthusiasm. Furthermore, Chinese digital art enterprises' overseas revenue accounts for only 8%. Although the annual growth rate reaches 55%, the overall internationalization level is low and commercialization layout is unbalanced [2][11][19][27]. Domestic digital art commercialization mainly concentrates on the digital collectibles field, with commercialization monetization capabilities in other scenarios needing improvement [27].
- **Imperfect Regulatory Systems:** The integrated development speed of AI world models and digital art is rapid, while relevant regulatory policies and laws and regulations lag behind [1]. Especially in AI-generated content review and NFT art supervision, there is a lack of clear regulatory rules, making it easy for vulgar and 不良 content to appear, and even triggering financial risks, such as NFT complaint rates maintaining a high level of 15% for a long time [2][22][28]. Domestic supervision of NFT art mainly focuses on guidance and standardization, with specific regulatory details still needing further improvement [28].

5.1 Technical Level: Promoting Technological Innovation and Adaptation, Lowering Technical Thresholds

- **Strengthening Core Technology R&D:** Encourage research institutions and technology enterprises to increase R&D investment in AI world model and digital art integration-related technologies, focusing on breakthroughs in model adaptability, stability, and controllability issues [1]. For example, optimize AI world model training algorithms, increase training data for abstract art and niche art styles, and improve models' understanding and presentation capabilities for personalized creativity. Develop lightweight AI integration tools to lower usage costs for small and medium studios and independent creators. Promote the landing of lightweight WebGPU rendering and cross-platform unified standards to effectively solve rendering efficiency and device compatibility issues [2][4][9][24][26]. Domestically, support for localized technology R&D should be increased, focusing on model adaptation for local art styles such as Guochao and intangible cultural heritage [26][29].
- **Promoting Technical Adaptation Optimization:** Establish adaptation standards between AI world models and digital art creation tools, promote technical synergy between the two [1]. Encourage deep cooperation between technology developers and art creators, optimize AI world model functions according to digital art creation needs, and create targeted integration tools (such as AI-assisted abstract art creation tools and virtual scene art design tools) [1][3][7][26].
- **Lowering Technology Popularization Thresholds:** Increase technical support for small and medium digital art studios and independent creators, providing free or low-cost AI integration tools and technical training services [2]. Carry out online and offline technical training courses to popularize AI world model usage methods, improve art creators' technical literacy, and promote widespread application of integration technology [2][9][17][29]. Domestically, relying on industry associations, targeted localized technical training can be carried out to promote lightweight AI tools [29].

5.2 Artistic Level: Adhering to Artistic Original Intent, Improving Work Quality and Innovation

- **Strengthening Creative Subjectivity:** Clarify creators' core position in integrated creation, promote the popularization of human-machine symbiosis mode [1], making AI world models tools assisting creative expression rather than replacing creators as subjects. Encourage creators to integrate their emotions, thoughts, and artistic concepts into AI-generated works, improving works' emotional connotations and artistic value [1][10][12][29]. Domestically, creators should be guided to deeply excavate local cultural elements to create AI digital art works with

Chinese characteristics [27][29].

- **Curbing Homogenization, Improving Innovation:** Guide creators and enterprises to excavate unique creative perspectives, combine AI world model technical advantages, and create personalized, innovative digital art works [1]. Encourage AI world model developers to enrich training data, include more niche art styles and traditional cultural elements, and improve the diversity of model-generated works. Establish scientific art quality evaluation systems to guide the industry to focus on works' artistic value rather than purely pursuing quantity and commercial interests [1][10][21][23][29].
- **Resolving Artistic Value Controversies:** Strengthen industry exchanges and guidance, hold AI digital art exhibitions, forums, and other activities to display excellent integrated works and convey the core concept of “technology empowering art, art enriching technology” [1]. Promote public correct cognition of AI digital art, recognize its artistic value, and form an industry consensus of “technology and art collaborative development” [1][10][29]. Domestically, local AI digital art competitions can be held to build exchange and display platforms, assisting healthy industry development [29].

5.3 Copyright Level: Perfecting Copyright Systems, Strengthening Copyright Protection

- **Clarifying Copyright Attribution:** Accelerate the improvement of relevant laws and regulations to clarify copyright attribution for AI world model-generated digital art works [1], clearly distinguishing the rights and obligations of model developers, work generators, and training data copyright holders. For special scenarios such as multi-person collaborative creation and batch generation, formulate specific copyright division rules to reduce copyright disputes [1][6][14][28]. Domestically, the improvement process of judicial interpretations related to the Copyright Law should be accelerated to clarify the copyright subject and rights boundaries of AI-generated works [28].
- **Standardizing Training Data Usage:** Establish authorization mechanisms for AI world model training data, requiring developers to use works authorized by copyright holders as training data [1]. Encourage the construction of legitimate digital art material libraries to provide legal and compliant material support for AI model training, effectively reducing infringement risks [1][6][14][28]. Domestically, localized legitimate material libraries can be built, focusing on collecting local materials such as traditional art and intangible cultural heritage works to assist local cultural inheritance and protection [28].
- **Strengthening Copyright Protection Technologies and Means:** Promote the deep application of blockchain, big data, and other technolo-

gies in copyright protection [2], establish copyright certification, tracing, and monitoring systems for digital art works to achieve full-process traceability of works. Improve copyright rights protection mechanisms, simplify rights protection processes, lower rights protection costs, and improve the efficiency of investigating infringement acts. Promote the unification of blockchain certification standards, improve cross-chain mutual recognition rates, and enhance the effectiveness of copyright protection [2][6][25][28]. Domestically, unified blockchain copyright certification standards should be established and low-cost rights protection channels built to protect creators' legitimate rights and interests [28].

5.4 Talent Level: Building Compound Talent Training Systems, Filling Talent Gaps

- **Optimizing University Talent Training Models:** Universities should adjust curriculum settings for AI-related majors and digital art-related majors, increasing interdisciplinary courses [2], such as AI world model foundations, digital art and AI integration creation, and AI copyright law, focusing on cultivating compound talents who understand both technology and art. Strengthen cooperation with enterprises and research institutions, carry out school-enterprise joint training, internships, and practical training projects to improve students' practical abilities [2][17][29]. Domestic universities should optimize interdisciplinary curriculum settings, increase teaching content on local culture and AI integration, strengthen practice links, and improve the targeting of talent training [29].
- **Strengthening Industry Talent Training:** Encourage enterprises and industry associations to carry out targeted talent training [2], focusing on cultivating art creators' AI technical capabilities and technology developers' artistic literacy. Actively introduce high-end compound talents, give policy support and incentives, and drive the overall improvement of industry talent teams [2][17][29].
- **Building Talent Exchange Platforms:** Hold compound talent exchange conferences, creation competitions, and other activities [1], providing opportunities for technical talents and art talents to exchange and cooperate, promoting collaborative innovation among talents, and building diversified, high-quality talent teams [1][17][29].

5.5 Industry Level: Perfecting Standard Systems, Promoting Healthy Commercialization Development

- **Establishing Industry Standards and Norms:** Led by industry associations, jointly formulate industry standards for AI world model and digital art integration with enterprises, research institutions, and creators [1], including work quality evaluation standards, technical adaptation standards, and copyright usage standards. Strengthen industry self-discipline,

guide enterprises and creators to operate normatively and create with integrity, 杜绝 low-quality and homogeneous works, and maintain good industry order [1][23][29]. Domestically, localized industry standards should be formulated faster to promote the standardized and high-quality development of AI digital art [29].

- **Promoting Balanced Commercialization Development:** Guide the industry to view commercialization rationally and balance artistic value and commercial value [1]. Encourage the expansion of diversified commercialization channels, such as virtual art exhibitions, digital art derivative development, and immersive experience projects, to promote integrated works with artistic value to achieve market value. Increase support for digital art exports, pilot export tax rebate policies for digital artworks, and improve the internationalization level of Chinese digital art enterprises [2][11][19][27]. Domestically, diversified commercialization scenarios such as cultural tourism and education should be expanded to improve the international competitiveness of local digital art [27].
- **Perfecting Regulatory Systems:** Accelerate the introduction of relevant regulatory policies, clarify review standards for AI-generated digital art works and regulatory rules for NFT art [1], guard against vulgar and 不良 content and financial risks. Establish multi-department collaborative regulatory mechanisms to strengthen industry supervision and promote healthy and orderly industry development. Meanwhile, encourage industry associations to play bridge and 纽带 roles, assist regulatory departments in carrying out work, and form a good pattern of “regulatory guidance, industry self-discipline, and enterprise self-discipline” [2][22][28].

6 Future Outlook for Integrated Development

As AI world model technology continues to iterate and the digital art industry continues to upgrade, their integration will gradually move toward deep synergy and comprehensive popularization, presenting a development trend of “technological intelligence, artistic diversification, scenario widespreadization, and industrial scaling.” Domestic integrated development will highlight localized characteristics, gradually forming a unique development path of “technological autonomy, artistic localization, and industrial standardization” [27][29].

From the technical level, future AI world models will achieve higher autonomy and controllability, capable of precisely understanding creators’ creative needs and generating works with more emotional connotations and artistic value [1]. Multimodal fusion technology will further break through, achieving seamless connection of text, images, audio, video, and virtual scenes, driving continuous innovation in digital art forms. Lightweight, low-cost integration technology will be widely popularized, allowing more creators to participate in integrated creation. Meanwhile, the development of general world models will promote the improvement of cross-scenario migration capabilities, further lowering integra-

tion thresholds [1][7][9][18][26]. Domestically, the R&D of localized AI world models will be accelerated, focusing on breakthroughs in core technologies such as Chinese creative adaptation and local art style generation [26][29].

From the artistic level, the integration of AI world models and digital art will break the boundaries of traditional art, giving birth to new art schools, new creation methods, and new artistic aesthetics [1]. The artistic value of AI-generated works will be widely recognized, and human-machine collaborative creation will become the mainstream mode, achieving the perfect balance between “technical rationality” and “artistic sensibility.” The integration of traditional cultural elements with AI technology and digital art will become an industry hotspot, driving the digital dissemination and innovative development of traditional culture [1][10][21][27]. Domestically, the focus will be on local cultural elements such as intangible cultural heritage and Guochao, creating AI digital art schools with Chinese characteristics [27][29].

From the scenario level, integration scenarios will further expand, penetrating multiple fields such as culture, education, tourism, entertainment, and advertising [2], such as AI virtual teachers, digital art cultural tourism projects, and AI-assisted advertising design. Immersive experience will become the mainstream way of digital art display. Combined with metaverse technology, more immersive and interactive virtual art worlds will be created, allowing users to fully experience the charm of digital art. Furthermore, emerging scenarios such as cultural tourism night tours, digital performances, and medical rehabilitation will become new growth points for integrated development [2][13][27]. Domestically, the deep landing of integrated scenarios in cultural tourism, education, and other fields will be promoted to create localized benchmark projects [27].

From the industrial level, their integration will promote the digital art industry to achieve scaled and high-quality development. It is expected that the industry compound growth rate will remain at 25% from 2025 to 2027, and the market scale is expected to reach 130 billion yuan in 2027 [2]. A complete industrial ecology of “technology R&D—creative creation—commercialization landing—copyright protection” will be gradually formed, cultivating a batch of enterprises and brands with core competitiveness. Meanwhile, industry standards and copyright systems will be continuously improved, and compound talent teams will gradually expand, providing solid support for integrated development [2][11][17][29]. Domestically, a batch of localized leading enterprises will be cultivated to promote the clustered development of the digital art industry [27][29].

Overall, the integration of AI world models and digital art is an inevitable trend of technological development and artistic innovation. Although currently facing many challenges, with the synergistic efforts of all parties, continuous technological breakthroughs, and continuous system improvements, their integration will certainly push the digital art industry into a brand new development stage, bringing more possibilities for human artistic creation and cultural dissemination. Domestic integrated development will rely on local advantages to

achieve coordinated advancement of technological autonomy, artistic characteristics, and industrial standards, assisting cultural 强国 construction [27][29].

7 Research Conclusions

Through systematic analysis of the background, current status, challenges, paths, and future outlook of AI world model and digital art integrated development, this research draws the following core conclusions:

- The integration of AI world models and digital art has clear internal logic and core value, representing an important manifestation of technology empowering art and art enriching technology. It has currently entered a rapid development period, forming diversified integration scenarios and modes, with commercialization landing pace continuously accelerating [1][2][11][27]. Domestic integrated development has formed a localized ecology, showing distinct characteristics in scenario landing and technical adaptation [27][29].
- Their integrated development has achieved certain results, forming a batch of typical cases in digital art creation, virtual experience, NFT art, and other scenarios, driving efficiency improvement and form innovation in the digital art industry. However, they also face core challenges including insufficient technical adaptation, artistic value controversies, imperfect copyright systems, and shortages of compound talents [1][2][10][17][28][29]. Domestically, they also face special challenges such as insufficient localized technology and lack of industry standards [26][29].
- Promoting their deep integration and high-quality development requires synergistic efforts from five levels: technology, art, copyright, talent, and industry. By strengthening technological innovation, adhering to artistic original intent, improving copyright systems, cultivating compound talents, and standardizing industry development, current dilemmas can be resolved [1][2][6][17][22][28][29]. Domestically, localized technology R&D, local copyright system improvement, and local compound talent cultivation need to be strengthened [26][28][29].
- In the future, with continuous technological iteration and continuous system improvements, the integration of AI world models and digital art will present broader development prospects, giving birth to new art forms and new industrial ecosystems, promoting high-quality development of the digital art industry, and injecting new momentum into the upgrading of cultural and creative industries [1][2][7][11][27][29]. Domestic integrated development will highlight local characteristics, assisting traditional culture digital dissemination and cultural 强国 construction [27][29].

References

- [1] Ding J, Zhang Y, Shang Y, et al. Understanding world or predicting future?

- a comprehensive survey of world models [J]. *ACM Computing Surveys*, 2025, 58 (3): 1-35.
- [2] Hurst W, Spyrou O, Tekinerdogan B, et al. Digital art and the metaverse: benefits and challenges [J]. *Future Internet*, 2023, 15 (6): 188.
- [3] Zhang J, et al. Text2NeRF: Text-Driven 3D Scene Generation With Neural Radiance Fields [J]. *IEEE Transactions on Visualization and Computer Graphics*, 2024, 30 (12): 7749-7762.
- [4] Mo W, et al. Dynamic Prompt Optimizing for Text-to-Image Generation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024: 26627-26636.
- [5] Li J, Liu M. Neural Radiance Fields (NeRF) for Virtual Art Scene Construction [J]. *ACM Transactions on Graphics (TOG)*, 2024, 43 (12): 356-364.
- [6] Zhang M, Chen M. Copyright Attribution and Infringement Risk Prevention of AI-Generated Digital Art [J]. *Intellectual Property*, 2024 (7): 45-56.
- [7] Xu C, Zou H, Li J, et al. Multimodal image generation and fusion through content-style hybrid disentanglement [J]. *Knowledge-Based Systems*, 2025, 330 (Part B): 114597.
- [8] Sun H, Wu L, Li X, et al. Style-woven Attention Network for Zero-shot Ink Wash Painting Style Transfer [C]//*Proceedings of the 2022 International Conference on Multimedia Retrieval (ICMR ' 22)*, New York: Association for Computing Machinery, 2022: 277-285.
- [9] Chen M, Zhang L. Development and Promotion of Lightweight AI Integration Tools—Assisting the Development of Small and Medium Digital Art Studios [J]. *Science & Technology Progress and Policy*, 2024, 41 (16): 123-130.
- [10] Wang Y, Li J. Artistic Value Controversies and Evaluation System Construction of AI-Generated Digital Art [J]. *Literature and Art Studies*, 2024 (5): 156-168.
- [11] Li L, Zhang Q. China Digital Art Industry Development Report (2024)—AI Integration Perspective [J]. *Cultural Industry Research*, 2024 (2): 89-106.
- [12] Liu M, Chen J. Communication Characteristics of Digital Art and Communication Innovation under AI Empowerment [J]. *Journalism & Communication Review*, 2024, 77 (3): 102-114.
- [13] Wen B, Xie H, Chen Z, et al. 3D Scene Generation: A Survey [EB/OL]. [2025]. <https://arxiv.org/abs/2505.05474>.
- [14] Wang L, Chen M. Copyright Protection Dilemmas and Improvement Paths of AI-Generated Works—Based on International Judicial Practice [J]. *Journal of Intellectual Property Law & Practice*, 2024, 19 (4): 178-190.
- [15] Li M, Wang H. Research on Talent Training Models for AI World Model and Digital Art Integration [J]. *Higher Education Research*, 2024, 45 (8): 89-98.

- [16] Chen J, Liu F. International AI Digital Art Integration Case Analysis and Implications for China [J]. *International News World*, 2024, 46 (7): 134-150.
- [17] Wang Q, Zhang M. Research on Creative Subjectivity in Digital Art Creation under Human-Machine Collaboration Mode [J]. *Art Observation*, 2024 (8): 28-35.
- [18] Kong L, Yang W, Mei J, Liu Y, Liang A, Zhu D, ...Liu Z. 3D and 4D world modeling: A survey[J]. *arXiv preprint arXiv:2509.07996*, 2025.
- [19] Li J, Wang Y. Research on Internationalization Development Paths of Chinese Digital Art Enterprises—AI Integration Perspective [J]. *International Trade Issues*, 2024 (10): 145-158.
- [20] Xu M, Cui J, Ma X, et al. Image enhancement with art design: a visual feature approach with a CNN-transformer fusion model [J]. *PeerJ Computer Science*, 2024, 10 (11): e1896.
- [21] Wang X. AI-Assisted Practice of Visual Design for Traditional Cultural Symbols [J]. *Academic Journal of Management and Social Sciences*, 2025 (1): 45-52.
- [22] Chen M, Liu F. Regulatory Dilemmas and Collaborative Governance Paths of AI-Generated Digital Art [J]. *Chinese Public Administration*, 2024 (9): 102-110.
- [23] Li M, Wang H. Homogenization Problems and Innovation Paths of AI Digital Art Works [J]. *Art Review*, 2024 (6): 98-108.
- [24] Zhang G, Bhagat S, Lee S H, et al. From game to beyond game: Exploring the role of 3D rendering technology in user immersion and virtual consumption in the Metaverse [J]. *Decision Support Systems*, 2025, 189: 114383.
- [25] Zhou H, Chen Y, Zhou J, et al. Blockchain Implementation for Visual Art Copyright Protection [C]//2024 IEEE 24th International Conference on Software Quality, Reliability, and Security Companion (QRS-C), Cambridge: IEEE, 2024: 232-239.
- [26] Guo T, Lu Z, Wang F. Multi-modal social media analysis via shap-based explanation: a framework for public art perception [J]. *IEEE Access*, 2025, 13: 68753-68765.
- [27] Wang Y, Chen M. Construction and Development of China' s AI Digital Art Integration Ecology—Taking Baidu and Alibaba as Examples [J]. *Industrial Economics Research*, 2024 (5): 134-146.
- [28] Li J, Zhang L. Analysis of AI-Generated Digital Art Infringement Cases and Improvement of Rights Protection Mechanisms in China [J]. *Studies in Law and Business*, 2024, 41 (5): 123-135.
- [29] Wang L, Liu F. Construction of Industry Standards and Quality Evaluation Systems for AI and Digital Art Integration in China [J]. *Academic Research*,

2024 (7): 102-110.

Source: ChinaXiv – Machine translation. Verify with original.