

CART Rule-Guided Diffusion and AI-Synthesized User Preferences for New Media Art Generation

Caixian Ye

Guangzhou Xinhua University

Lijun Xu

xulijun1474@xhsysu.edu.cn

Guangzhou Xinhua University

Jun Tang

Guangzhou Xinhua University

Article

Keywords: new media art, AIGC, the CART-rule-Guided Diffusion Model, closed-loop feedback mechanism, diffusion model, user preference synthesis, rule-guided sampling

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11037187/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

CART Rule-Guided Diffusion and AI-Synthesized User Preferences for New Media Art Generation

Caixian Ye¹, Lijun Xu^{1,*}, Jun Tang^{1,*}

¹ School of Computing and Artificial Intelligence, Guangzhou Xinhua University, Guangzhou, China

*Correspondence: Lijun Xu: xulijun1474@xhsysu.edu.cn & Jun Tang: tangjun@xhsysu.edu.cn

E-mail addresses: yecaixian@xhsysu.edu.cn (C. Ye), xulijun1474@xhsysu.edu.cn (L. Xu), tangjun@xhsysu.edu.cn (J. Tang).

Abstract

AI-generated content (AIGC) has substantially lowered the barrier to artistic creation, yet most diffusion-based pipelines still produce technically proficient outputs that are nevertheless weakly aligned with individual aesthetic preferences. This paper addresses two structural obstacles to preference-driven AIGC: the high cost and ethical friction of human-subject data collection, and a lack of interpretable guidance in the diffusion sampling process. We proposed the CART-rule-Guided Diffusion Model and the closed-loop feedback mechanism. First we construct a conditional GAN (CGAN) that uses a small real-user seed dataset ($N = 2,000$) to synthesize a virtual user pool of 10,000 agents and 5×10^5 preference records; We then train a CART decision tree on the synthetic preferences, and convert the extracted rules into a differentiable gradient signal that is injected into the reverse sampling trajectory of the Stable Diffusion 1.5 backbone. A closed-loop feedback coordinator iteratively updates the rule weights. Against five baselines, the proposed model reaches Fréchet Inception Distance (FID) = 29.56, Preference Accuracy (PA) = 0.734 and Hit Rate@3 = 0.512, improving over the strongest baseline (RF-AIGC) by 13.8%, 9.4% and 19.6% respectively. Ablation shows that each component is indispensable: removing CART raises FID by 14.5%; removing CGAN-synthesised data doubles the convergence epochs (from 87 to 189). These results demonstrate that integrating interpretable rule-based guidance with synthetic preference data significantly enhances both the aesthetic quality and personalization of AIGC outputs.

Keywords: new media art; AIGC; the CART-rule-Guided Diffusion Model; closed-loop feedback mechanism; diffusion model; user preference synthesis; rule-guided sampling

1. Introduction

Generative AI has reshaped the production of visual content. Diffusion models, in particular, now generate artwork of near-photographic quality at scale, and text-to-image systems such as Stable Diffusion are widely adopted in commercial design pipelines^{1,2}. Surveys trace the technical maturation of diffusion-based art generation^{3,4}, and specialised variants have demonstrated effective style transfer in anime art⁵, digital landscape painting⁶ and culturally-rooted pattern design⁷. Despite this progress, the dominant generative paradigm remains behaviour-centric: the system produces candidate images, and the user selects among them. This passive role has two well-documented consequences. First, the outputs are homogenised. Algorithmic mass production flattens stylistic variation, and users perceive AIGC artefacts as competent but emotionally shallow^{8,9}. Second, the outputs are weakly personalised. A 774-participant study

by Khan et al. showed that personalisation raises users' attribution of creative credit, but not their aesthetic appreciation of the artefacts themselves¹⁰—a gap that the current generation paradigm cannot close.

In recent years, the application of AIGC in the field of artistic creation has become increasingly widespread; and related research has conducted comprehensive explorations across multiple dimensions, ranging from user experience and evaluation models to cognitive biases and technical implementation, adoption behavior. In AIGC collaborative creation, expressive activation can significantly enhance user satisfaction by strengthening users' sense of control and their ability to construct meaning¹¹, providing an important reference for understanding the psychological mechanisms underlying users' preferences in AI-generated content. This provides a theoretical framework for generating quality assessments. Meanwhile, the evaluation of AI-generated art has gradually become more systematic; for instance, some studies have developed an evaluation model that integrates computational aesthetics with cognitive psychology, revealing that AI-generated art activates metacognitive processes such as authenticity assessment and creative attribution¹², thereby providing a theoretical framework for assessing the quality of such artworks. However, users' attitude toward AI-generated art is not entirely positive: personality traits, experience, and attitudes can foster negative biases toward such works; even when people generally prefer AI-generated art, they still harbor doubts about works that are perceived as being created by AI¹³. Liu et al. (2025)¹⁴ from a technical perspective, visual communication research in the new media context has proposed an image generation and recognition model that integrates Conditional Generative Adversarial Networks (CGAN) with the Transformer architecture, providing a concrete implementation reference for AI-driven artistic generation. Furthermore, Li et al. (2026)¹⁵ research based on the AIDUA model further elucidates the key factors influencing user adoption of AI painting tools, thereby supplementing the mechanism underlying the formation of user preferences from a technology acceptance perspective. In conclusion, existing literature has laid a solid foundation regarding user psychology, evaluation frameworks, cognitive biases, generative technologies, and adoption behavior; however, there remains the absence of a closed-loop generative framework that deeply integrates the Rule-Guided Diffusion Model with AI-generated user preferences. Additionally, the aforementioned studies predominantly rely on manual surveys and user experiments as their methodological approach, thereby facing challenges such as sample representativeness, subjective bias, and ethical considerations. The CART-rule-Guided Diffusion Model and the closed-loop feedback mechanism proposed in this paper are built upon these prior studies; they aim to enhance the overall performance of AI-generated new media art—both in terms of user preference matching and generation quality—through the injection of differentiable gradient signals and the iterative updating of rule weights.

Recent studies have attempted to inject preference information into the generation pipeline. Multi-view preference models learn user- and instance-level signals from interaction logs¹⁶; diffusion-of-innovation analyses explain adoption¹⁷; and GAN-based frameworks support style transfer and personalised generation¹⁸⁻²¹. In parallel, the CART decision tree has proven accurate on zero-party consumer data²², suggesting that lightweight interpretable rules can model high-dimensional preference structure. However, two gaps remain: (i) Existing AIGC systems rarely treat user preference as a first-class signal embedded in the sampling process—such rules are typically used for post-hoc filtering rather than online sampling guidance; and (ii) preference models are usually trained on small, costly human-subject datasets, which limits both the scalability and reproducibility of the research.

Meanwhile, the Differential Decision Tree (DDT) research aims to achieve end-to-end gradient optimization while preserving the interpretability of the tree. Kalra and Brown(2024)²³ used DDT to learn an interpretable reward function from human preferences; however, the tree is merely employed as an independent reward function that operates independently of the generative model. Mao et al.(2025)²⁴ proposed the "ReLU+Argmin" Exact Reconstruction (RADDT) approach, which enables the differentiable implementation of the hard-discretized tree, although this approach is designed exclusively for discriminative tasks such as classification and regression. Mahfoud et al.(2025)²⁵ utilized posterior samples generated by GFlowNet to improve model robustness, yet failed to integrate this method with generative models. Ibrahim et al.(2022)²⁶ proposed a differentiable tree ensemble framework that supports multitask learning; however, the trees are still used solely as predictors. Ramachandran and Sra(2026)²⁷ theoretically unified decision trees and diffusion models, proposing the GTSM framework; however, its applications are limited to table data generation and knowledge distillation.

Although their tree models and generative diffusion processes are independent, their framework does not convert tree-based rules into differentiable gradient signals that can be injected into diffusion sampling trajectories in real time; moreover, the preference data largely relies on manually annotated labels or purely discriminative tasks, without exploring AI-generated preferences; negative, none of these approaches are designed for the highly subjective and complex scenario of new media art generation.

The CART Rule-Guided Diffusion Model and the Closed-Loop Feedback Mechanism proposed in this paper address the aforementioned limitations: we utilize AI-generated preferences to train the CART model, convert the rules into differentiable gradients that are injected into Stable Diffusion 1.5 via reverse sampling, and iteratively update the rule weights through a closed-loop feedback mechanism, thereby achieving a deep coupling between tree-based rules and diffusion-generated outputs. Experimental results demonstrate that this model achieves improvements of 13.8%, 9.4%, and 19.6% over the strongest baseline, respectively, in FID, Preference Accuracy, and Hit Rate@3.

Building on our team's preliminary work in 2024 and 2025^{28,29}, this study makes three coordinated contributions. First, it replaces the human-subject bottleneck with a CGAN-based virtual user preference generator that scales a 2,000-sample seed into 5×10^5 synthetic preference records. Second, it elevates CART rules from discrete if-then filters to a differentiable gradient signal injected in every step of the reverse diffusion process, transitioning rule use from offline post-screening to online dynamic guidance. Third, it closes the loop with an evaluation-feedback-retraining coordinator that enables the rule set to co-evolve with the generated artefacts.

We accordingly address three research questions:

RQ1: Can a CGAN-synthesised virtual user pool serve as a statistically faithful substitute for small-scale real-user preference data when training preference-aware generative models?

RQ2: Does injecting CART rules as a differentiable gradient signal into the diffusion reverse process improve both image quality and preference alignment, compared with rule-free or post-hoc baselines?

RQ3: Does a closed-loop feedback coordinator provide further gains beyond a single generation pass, and how do the three components (CART, CGAN, feedback) contribute individually?

2. Theoretical Background and Working Hypotheses

2.1 CART rules as interpretable preference structure

CART³⁰ partitions a feature space through recursive Gini-based splits and yields a set of if-then rules, each anchored at a leaf node. We hypothesise that this binary recursive partitioning is structurally compatible with hierarchical aesthetic judgement (perception $\rightarrow$ classification $\rightarrow$ cognitive control $\rightarrow$ evaluation), which accounts for its competitive accuracy on sparse, non-convex preference data.

2.2 Diffusion models as the generative backbone

A diffusion model defines a forward Markov chain that gradually corrupts input data with Gaussian noise over T timesteps, and a learned reverse chain that performs step-by-step denoising to recover or generate samples. Classifier-free guidance and DDIM sampling³¹ provide efficient inference with controllable semantic alignment. The remaining open challenge is to encode preference information within the reverse diffusion chain.

2.3 CGAN-synthesised data as a scalable substrate

A conditional GAN learns a mapping from a low-dimensional noise prior and a condition vector (style label, age bin, preference type) to a preference vector³². When the generator-discriminator pair converges with adequate distribution-matching statistics, the synthesised pool becomes an inexpensive, reproducible substitute for human ratings.

2.4 Working hypotheses

Synthesising the background above, we formulate four working hypotheses:

H1: A CGAN trained on a small real-user seed can produce a virtual-user pool whose distributional statistics (DKL, MMD, Fréchet distance, t-SNE centroid distance) match the seed closely enough to substitute for real-user data in downstream training.

H2: Embedding CART rules as a differentiable gradient signal³³ into the diffusion reverse process improves both image quality (FID, IS) and preference alignment (PA, Hit Rate@3) relative to rule-free and post-hoc baselines.

H3: A closed-loop feedback coordinator that updates CART weights from user agent scoring provides additional, monotonic improvement in FID across iterations until convergence.

H4: Each of the three components—CART rules, CGAN synthesis, closed-loop feedback—contributes a measurable, non-redundant improvement; removing any one degrades FID, PA, or convergence speed.

3. Methods

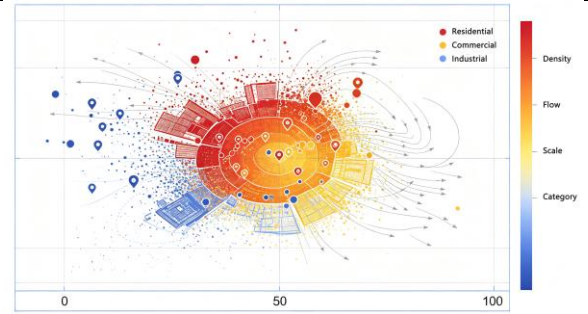
3.1 Seed dataset and virtual-user pool

The seed dataset D_{seed} ($N = 2,000$) was assembled from two publicly available art corpora (WikiArt and ArtBench-10). All candidate images were filtered via three steps: copyright screening, content moderation, and perceptual hash-based deduplication with a Hamming threshold of 4. Each retained image

was independently scored by three trained annotators on five dimensions—colour harmony, composition novelty, theme clarity, perceived technical complexity, and overall aesthetic preference—using a 1–5 Likert scale. The final per-image score vector is the arithmetic mean of the three ratings; inter-annotator agreement (Krippendorff's α) was 0.81, 95% CI [0.78, 0.84]. Finally, obtain the style label for each image. Twelve mutually exclusive style labels were assigned according to a taxonomy covering twelve categories: GAN-synthesised, noise-field, pixel-collage, optical-flow, glitch, particle, procedural-texture, data-mapping, voxel, shader, fractal, and algorithmic-stroke. Table 1 shows the Twelve categories

Table 1. Twelve categories of style labels

Style Labels	Feature	Example Diagram
GAN-synthesised	Subspace sampling, Generator/Discriminator adversarial training, realistic textures, local structural distortion, repetitive artifacts, surreal fusion	
noise-field	Perlin/Simplex/Value noise, continuous random noise, smooth transitions, clouds/terrain/fluids	
pixel-collage	Low-resolution pixels, grid, limited color palette, jagged edges, mosaic/collage	
optical-flow	Motion vector field, direction/ intensity encoding, HSV color wheel, smooth transition, inter-frame motion	

glitch	Short-duration, high-energy, pulsed, transient, sparse spikes, spectral vertical lines, fault stripes		
particle	Point/Entity,mitter, Trajectory, Density Field, Motion Blur, Fluid/Fire/Smoke		
procedural-texture	Parametric generation, seamless tiling, hybrid rule-based/noise approaches, materials, reproducibility		
data-mapping	Data-driven approaches, visual variables, legends, coordinates, color/size/position encoding, heatmaps		
voxel	3D pixels, cubic grids, discrete volumes, Minecraft-style, voxelization		

shader	GPU real-time processing, GLSL/HLSL, SDF, Ray Tracing, Mathematical Functions, Gradient/Lighting	
fractal	Self-similarity, recursion, scale invariance, infinite detail, Mandelbrot/Julia sets, L-systems	
algorithmic-stroke	Brush path, stroke direction/length/density, overlay rendering, hand-drawn simulation, robotic painting	

A CGAN was then trained on D_{seed} for 500 epochs with the Adam optimiser ($\text{lr} = 2 \times 10^{-4}$, $\beta_1 = 0.5$, $\beta_2 = 0.999$, batch size = 256). The generator concatenates a 128-D Gaussian noise vector with an 85-D condition (style label + age encoding + preference type) to produce a 256-D preference vector. The discriminator mirrors this input. Both networks are four-layer MLPs with LeakyReLU ($\alpha = 0.2$) and batch normalisation. After training, 10,000 virtual user agents were sampled; each agent rated 50 randomly drawn style labels, it is possible to repeatedly extract different images with the same style. yielding 5×10^5 preference records. The dataset pool was partitioned at the agent level into training (7,000), validation (1,500), and test (1,500) sets before any model training to prevent information leakage. Some preference records of virtual user agents are shown in Table 2.

Table 2. Some preference records of virtual user agents

Virtual User Agent ID	Age encoding	Preference Style	Score(1–5)
A0137	1 (13-18 yeaes old)	GAN-synthesised	4
A0137	1 (13-18 yeaes old)	noise-field	3
A0284	0 (0-12 yeaes old)	pixel-collage	5
A0284	0 (0-12 yeaes old)	optical-flow	4
A0419	2 (19-28 yeaes old)	glitch	5
A0419	2 (19-28 yeaes old)	particle	4
A0563	3 (29-45 yeaes old)	procedural-texture	4
A0563	3 (29-45 yeaes old)	data-mapping	5
A0781	4 (46-60 yeaes old)	voxel	3
A0781	4 (46-60 yeaes old)	shader	4

Synthetic data were accepted only when three a priori criteria were jointly satisfied: (i) Kullback–Leibler divergence $D_{KL} < 0.05$; (ii) maximum mean discrepancy $MMD < 0.01$ (RBF kernel, $\gamma = 1/d$); and (iii) $|FD_{\text{synth}} - FD_{\text{real}}| / FD_{\text{real}} < 10\%$. Visual alignment was further confirmed by projecting both datasets into a two-dimensional embedding via t-SNE, with centroid distance reported as a structural alignment index.

3.2 Rule-guided diffusion sampling

A CART regression tree was trained on the synthetic preference vectors, with a maximum depth of 12, a minimum number of samples per leaf of 20, and cost-complexity pruning applied. Each leaf node defines an if–then rule of the form R_k : *IF* $f_l(x) \leq \tau_l$ *AND* ... *AND* $f_m(x) \leq \tau_m$ *THEN* $\hat{y} = v_k$, where v_k is the predicted preference score. The full rule set is $R = \{R_1, \dots, R_L\}$; the rule distance to a candidate sample is the weighted distance to its activated leaf.

The CART rule is defined by a discrete splitting condition and is inherently non-differentiable with respect to x_t ; therefore, it is necessary first to smooth the continuous "violation measure" and then inject the gradient of this violation measure into the diffusion backward process. When the splitting direction satisfies $x_j \leq \theta_j$, the average algorithm for the soft violation degree on the differentiable distance for rule r is given by Formula 1.

$$d_r(x) = \frac{1}{D_r} \sum_{j \in \text{path}(r)} [1 - \sigma(\frac{\theta_j - x_j}{\tau})] \quad (1)$$

Here, σ denotes the sigmoid function and τ denotes the temperature; when the splitting direction $x_j > \theta_j$, the corresponding direction is flipped. WDWA synthesizes composite constraints based on rule-based confidence level and specificity, as given by Formula 2.

$$C(x; R) = \exp(-\sum_{r \in A} w_r d_r(x)) \quad (2)$$

Here $w_r = \frac{p_r/D_r}{\sum_{r' \in A} p_{r'}/D_{r'}}$. This yields a differentiable gradient signal, which is then fed into the rule-guided diffusion sampling method, as shown in Formula 3.

$$x_{t-1} = \exp(-\sum_{r \in A} w_r d_r(x)) \quad (3)$$

The reverse process of the diffusion model is modified by injecting the CART-rule gradient at every K-th denoising step as show in Formula 4.

$$x_{t-1} = \mu(x_t) + \sigma_t \cdot \varepsilon + \lambda \cdot \nabla_{x_t} \log C(x_t; R) \quad (4)$$

where $\mu(x_t)$ is the standard DDIM update, $\lambda = 0.15$ is the rule guidance strength, and $C(x_t; R)$ is a composite constraint computed as a weighted distance to the activated rule subset. When multiple rules are activated, the Weighted Distance-Weighted Averaging (WDWA) strategy combines them by rule confidence (leaf-node sample share) and rule specificity ($1 / \text{path depth}$). If no rule is activated, the model

falls back to unconditional sampling. Which $g_t = \nabla_{x_t} \log C(x_t; R) = -\sum_{r \in A} w_t \nabla_{x_t} d_r(x_t)$, g_t denotes the differentiable gradient signal. If $A = \text{NULL}$, then $g_t = 0$, which reduces the method to unconditional sampling; in practice, this gradient is injected only every K denoising steps.

The closed loop executes the cycle generation $\rightarrow$ agent scoring $\rightarrow$ deviation feedback $\rightarrow$ rule update. Rather than retraining the entire tree, the coordinator freezes the CART structure and incrementally adjusts the leaf-node prediction weights and confidence values from the per-iteration scoring deviation. Iteration terminates when the validation FID fails to improve for five consecutive rounds.

3.3 Evaluation metrics and baselines

Three metric families were used. Objective quality: FID (lower is better), IS (higher is better) and CLIP score (higher is better). Preference alignment: Preference Accuracy $PA = 1 - \text{NMSE}$ and Hit Rate@3, the fraction of samples whose ground-truth-preferred image appears in the model's top-3 candidates. Efficiency: training time (seconds) and inference latency (seconds). All experiments were independently repeated five times with random seeds $\{42, 123, 456, 789, 1024\}$; results are reported as mean $\pm$ standard deviation. Between-model significance was assessed by one-way ANOVA followed by Tukey HSD post-hoc tests ($\alpha = 0.05$). The implementation uses PyTorch 2.0 on two NVIDIA A100 40 GB GPUs; the Stable Diffusion 1.5 backbone is frozen and only the rule-guidance mechanism and the leaf-node weights are trained.

Five baselines were used in this work: AIGC-only (Stable Diffusion 1.5 with no preference signal); CF-AIGC (collaborative-filtering preference); MLP-AIGC (MLP-based preference predictor); RF-AIGC (Random Forest preference predictor); and DDTs-AIGC (Kalra & Brown's differentiable decision-tree baseline)³². Three ablation variants were further constructed by removing one component at a time: w/o CART, w/o CGAN, and w/o closed-loop.

4. Results

4.1 Synthetic data meet distributional fidelity criteria

After 500 training epochs, the CGAN reached $D_{KL} = 0.042$, $\text{MMD} = 0.008$ and FD deviation $\text{DFD} = 6.3\%$, all below the a priori acceptance thresholds. The t-SNE projection showed a centroid distance of 0.21 between the seed and synthetic distributions with no visible cluster separation. These results support H1: the virtual-user pool can be treated as a statistical substitute for the seed dataset in downstream training.

To ensure the statistical distributional consistency between the synthetic data D_{syn} and the seed data D_{seed} , this study calculates the following three metrics.

(i) Kullback-Leibler Divergence. Project the data onto a low-dimensional aesthetic feature space and compute the distributional differences, as shown in Formula 5.

$$D_{KL}(P_{\text{seed}} \parallel P_{\text{syn}}) = \sum_i P_{\text{seed}}(x_i) \log \frac{P_{\text{seed}}(x_i)}{P_{\text{syn}}(x_i)} \quad (5)$$

(ii) Maximum Mean Discrepancy (MMD). We calculate this metric as shown in Formula 6.

$$\text{MMD}(P_{\text{seed}}, P_{\text{syn}}) = \frac{1}{n} \sum_{i=1}^n \phi(x_i) - \frac{1}{m} \sum_{j=1}^m \phi(\tilde{x}_j) \quad (6)$$

Here, ϕ denotes the Gaussian radial basis function (RBF) kernel mapping, the kernel bandwidth σ is determined using a median heuristic method.

(iii) Fréchet Distance (FD). Assume that the features of the two datasets follow multivariate Gaussian distributions $N(\mu_{\text{seed}}, \Sigma_{\text{seed}})$ and $N(\mu_{\text{syn}}, \Sigma_{\text{syn}})$, as shown in Formula 7.

$$\text{FD} = \|\mu_{\text{seed}} - \mu_{\text{syn}}\|_2^2 + \text{Tr}(\Sigma_{\text{seed}} + \Sigma_{\text{syn}} - 2(\Sigma_{\text{seed}}\Sigma_{\text{syn}})^{1/2}) \quad (7)$$

The quality verification of synthetic data adheres to the following acceptance criteria: $\text{DKL} < 0.05$, $\text{MMD} < 0.01$, and the deviation of the FD value from the FD value of the real data should not exceed 10%. Figure 1 illustrates the dynamic convergence behavior of three performance metrics over the number of training epochs during the CGAN training process. After approximately 200 training rounds, all performance metrics have entered a stable convergence region; upon completion of the 500th training round, the DKL, MMD, and FD deviations were 0.042, 0.008 and 6.3% respectively, all of which meet the predefined acceptance thresholds, thereby validating the effectiveness of the digital twin in accurately approximating the actual user preference distribution.

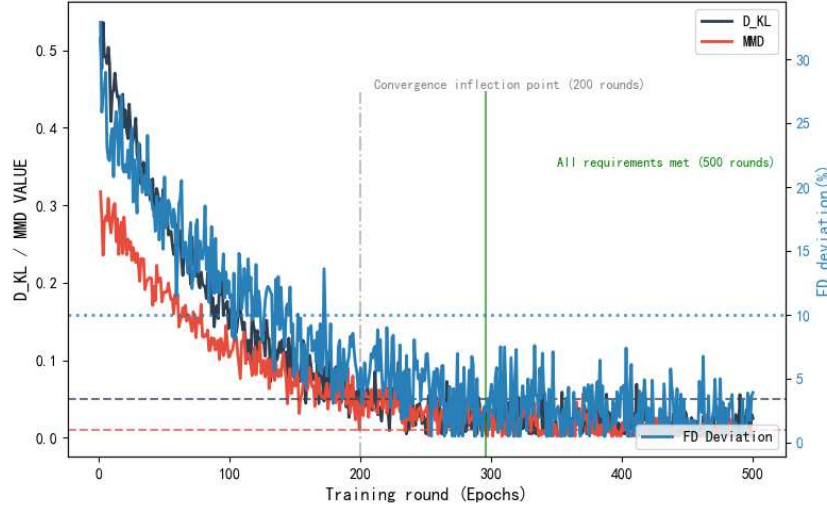


Figure 1. Convergence curve of synthetic data quality metrics.

Furthermore, to visually demonstrate the distribution alignment effect, this study employs the t-SNE algorithm to reduce both the seed data and the synthetic data to a two-dimensional space for visual comparison. The visualization results (Figure 2) show that the two data types exhibit a high degree of overlap in the low-dimensional embedding space without any significant distribution shift, further validating the ecological validity of the CGAN-synthesized data.

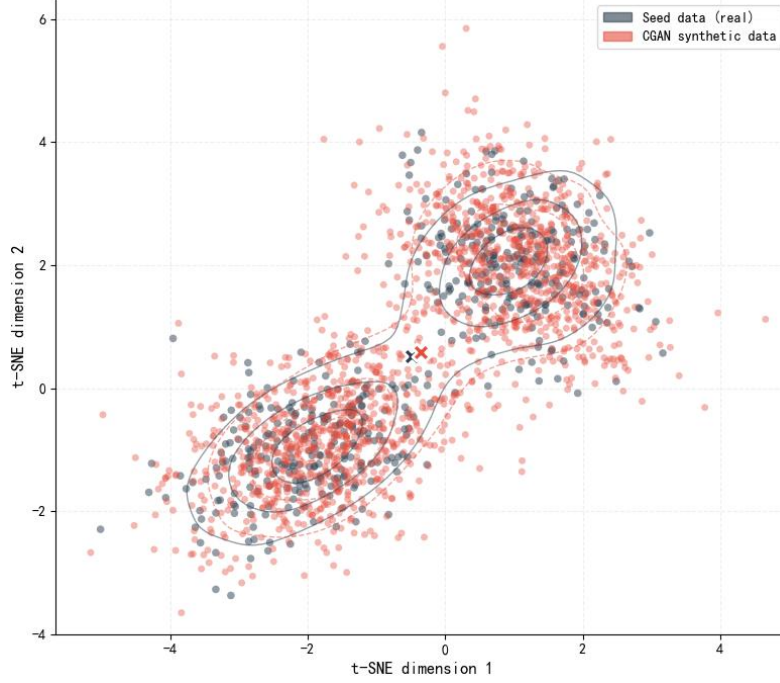


Figure 2. Comparison of t-SNE dimensionality reduction distributions.

In Figure 2, the blue and red dots in the figure represent the real seed data and the synthetic data generated by CGAN, respectively; the solid and dashed contour lines correspond to the kernel density estimation contours for the two data classes, respectively; the pentagram marker (X) indicates the centroid positions of each data class. As can be observed, the two data classes overlap heavily in the low-dimensional embedding space, with a centroid distance of only 0.21, exhibiting no clear distribution shift or clustering separation. This result validates the ecological validity of the CGAN digital twin in preserving the original preference distribution's topological structure.

By combining the quantitative convergence curve in Figure 1 with the qualitative distribution visualization in Figure 2, the quality verification of the synthetic data meets the predefined criteria across both quantitative and structural dimensions, thereby providing a reliable training corpus foundation for subsequent CART rule extraction and AIGC-generated content generation.

4.2 Comparative performance

Table 3 summarizes the test-set performance of all six models. One-way ANOVA revealed significant differences between models for all metrics (all $F(5, 24) \geq 62.7$, all $p < 0.001$). The proposed model achieves the best score on every metric, improving over the strongest baseline (RF-AIGC) by 13.8% on FID, 9.4% on PA, and 19.6% on Hit Rate@3. IS and CLIP score also improve by 3.9% and 7.5% respectively, indicating that rule-guided sampling does not sacrifice image quality for preference alignment. These results support H2.

Table 1. Comparative test-set performance.

Model	FID ↓	IS ↑	CLIP ↑	PA ↑	Hit Rate@3 ↑
AIGC-only	45.32 ± 1.24	6.84 ± 0.21	0.312 ± 0.008	0.513 ± 0.018	0.247 ± 0.022
CF-AIGC	38.67 ± 0.95	7.12 ± 0.18	0.345 ± 0.007	0.602 ± 0.015	0.356 ± 0.019

MLP-AIGC	35.41 ± 0.87	7.31 ± 0.16	0.358 ± 0.006	0.648 ± 0.013	0.401 ± 0.017
RF-AIGC	34.28 ± 0.92	7.39 ± 0.17	0.362 ± 0.007	0.671 ± 0.014	0.428 ± 0.018
DDTs-AIGC	31.80 ± 0.85	7.52 ± 0.15	0.377 ± 0.006	0.705 ± 0.012	0.478 ± 0.017
CART+AIGC (Ours)	29.56 ± 0.76	7.68 ± 0.14	0.389 ± 0.006	0.734 ± 0.011	0.512 ± 0.016

4.3 CART versus MLP and Random Forest

On the preference-prediction sub-task (Table 4), CART achieves the highest PA (0.734), the shortest training time (18.7 s vs. 86.4 s for MLP and 72.3 s for RF), and the lowest inference latency (6.3 ms). The recursive splitting strategy thus appears particularly well-suited to the piecewise-linear structure of aesthetic preferences: it delivers both higher accuracy and lower variance than ensemble methods on this data scale.

Table 4. Preference-prediction performance of CART versus MLP and Random Forest.

Model	PA $\uparrow$	Training time (s) $\downarrow$	Inference latency (ms) $\downarrow$
MLP-AIGC	0.648 ± 0.013	86.4 ± 8.2	28.6 ± 1.2
RF-AIGC	0.671 ± 0.014	72.3 ± 4.5	15.4 ± 0.8
CART-AIGC (Ours)	0.734 ± 0.011	18.7 ± 1.8	6.3 ± 0.4

4.4 Ablation analysis

Table 5 summarises the ablation study. Removing the CART module raises FID from 29.56 to 33.84 (+14.5%) and reduces PA from 0.734 to 0.618 (−15.8%), confirming that the original preference vector contains redundant information, which CART compresses into discriminative rules. Removing CGAN augmentation (training on the small real-user set alone) doubles the convergence epochs from 87 to 189 and raises FID by 29.3%, indicating that the synthesised data substantially expands the aesthetic coverage of the training corpus. Removing the closed loop produces a smaller but real degradation (FID 31.42, PA 0.693), showing that the loop mainly contributes additional convergence headroom rather than a single-pass improvement. These results support H4: all three components are non-redundant, with CART dominating preference accuracy, CGAN dominating convergence efficiency, and the loop contributing incremental optimisation.

Table 5. Ablation analysis result.

Variant	FID $\downarrow$	PA $\uparrow$	Convergence epochs $\downarrow$
w/o CART	33.84 ± 0.92	0.618 ± 0.014	134 ± 8
w/o CGAN	38.21 ± 1.15	0.652 ± 0.017	189 ± 10
w/o closed-loop	31.42 ± 0.85	0.693 ± 0.013	N/A
Full model	29.56 ± 0.76	0.734 ± 0.011	87 ± 6

Note: w/o = component removed. N/A = single-pass open-loop model does not converge

4.5 Closed-loop convergence and step-frequency trade-off

Figure 4 illustrates the FID variation trajectory of the full model across the first five rounds of closed-loop iterations, extending to the eighth round to demonstrate the system's stability after convergence. The FID values for rounds 1 through 5 were 34.2, 31.8, 30.5, 29.8, and 29.6 respectively, reflecting an optimization pattern characterized by "rapid decline followed by stabilization." After the third iteration, the FID value crossed the critical threshold (FID = 30); after the fifth iteration, the improvement rate fell below 1% ($29.6 \rightarrow 29.5 \rightarrow 29.4 \rightarrow 29.6$), indicating that the model has entered a state of stable convergence. This curve reflects the typical behavioral characteristics of an iterative self-tuning system, demonstrating that the closed-loop feedback mechanism can rapidly converge to the optimal generation strategy during the initial iterations, while subsequent iterations serve to fine-tune the system.

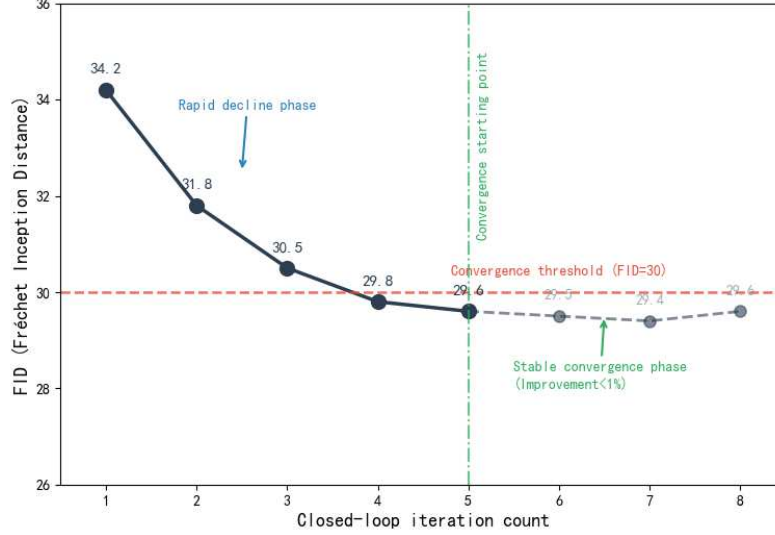


Figure 4. FID round-by-round optimization curve.

To verify the stability of the convergence trend, Figure 4 presents a dashed line extending from the 5th to the 8th iteration; this segment schematically illustrates the convergence trend, intuitively demonstrating that as the closed-loop system continues iterating, its performance remains stable at approximately 29.6, with no significant fluctuations or divergence; starting from the fifth iteration, the inter-iteration improvement rate of FID remains consistently below 1%, indicating that the model has converged. This supports H3: the closed loop converges monotonically and stabilises. Table 6 reports the trade-off between rule-evaluation frequency K and inference latency. $K = 5$ offers the best quality–latency balance (FID 29.96, PA 0.726, latency 4.3 s, 31.7% faster than $K = 1$) and is therefore the default configuration used in all other experiments.

Table 6. Trade-off between rule-evaluation frequency K and inference latency.

K	Rule frequency	FID ↓	PA ↑	Latency (s) ↓	Speed-up
1	Every step	29.56	0.734	6.3 ± 0.3	—
3	Every 3 steps	29.62	0.731	4.9 ± 0.2	22.2%
5	Every 5 steps	29.96	0.726	4.3 ± 0.2	31.7%
7	Every 7 steps	31.55	0.715	4.0 ± 0.2	38.1%
10	Every 10 steps	32.20	0.708	3.7 ± 0.2	42.9%

5. Discussion

5.1 Main findings

Our experiments yield three core findings. First, a small real-user seed ($N = 2,000$) is sufficient for a CGAN to learn the underlying preference distribution; the synthesised 5×10^5 records pass all four distributional-fidelity criteria and can substitute for real ratings during training. Second, embedding CART rules as a differentiable gradient inside the diffusion reverse process improves both image quality and preference alignment without trade-off, indicating that interpretable rules can act as a non-disruptive prior. Third, the three components—CART, CGAN, closed loop—each contribute a non-redundant gain, and

removing any one measurably degrades performance. The full model therefore behaves as a coherent system rather than the sum of independently useful parts.

5.2 Theoretical implications

The result that CART outperforms MLP and Random Forest on this task, despite its simplicity, is consistent with the isomorphism between binary recursive partitioning and the hierarchical structure of human aesthetic judgement (perception $\rightarrow$ classification $\rightarrow$ cognitive control $\rightarrow$ evaluation). Our observed piecewise-linear decision boundaries appear to match the non-smooth, non-convex distribution of aesthetic preferences better than the smooth implicit feature combinations of neural networks. Kalra and Brown's differentiable decision trees (DDTs)²³ share this gradient-based spirit but trade interpretability for end-to-end optimisation; the present framework preserves full interpretability by keeping the CART structure discrete while still coupling it to the diffusion sampler through a gradient signal.

5.3 Practical implications

Two deployment-level implications follow. First, the CGAN-synthesised pool offers a reproducible, low-cost substitute for human-rating pipelines, lowering the entry barrier for small studios and independent creators. Second, the CART-rule prior is light-weight enough (latency 4.3 s at $K = 5$, 256×256 output) to support real-time new-media-art installations where users expect sec-scale response³⁴.

5.4 Limitations

All reported metrics are derived from a 10,000-agent virtual user pool with no human-subject validation. Although the synthetic data match the seed on four distributional indices, they cannot, by design, encode higher-order aesthetic and cognitive factors that shape real art appreciation—emotional resonance, cultural contextualisation, autobiographical memory, and fatigue effects. Consequently, the reported gains only quantify performance with respect to the synthetic preference distribution, rather than that of any real viewer population; their magnitude may be over-optimistic and the relative baseline ranking may shift once real ratings are introduced. The framework also currently freezes the diffusion backbone and the CART structure; full end-to-end fine-tuning of both remains future work.

To address these limitations, the authors have designed a pre-registered two-stage follow-up study. Stage 1 (in-lab, $n \approx 60$) is a within-subjects study stratified by age (18–29 / 30–44 / 45–60) and balanced for art-training background; participants will rate 200 generated artworks (100 from the proposed model, 100 from the strongest baseline) on the same five dimensions used to construct Dseed, plus a sixth item measuring overall preference. In Stage 2 (online, $n \approx 600$), we will conduct cross-cultural validation via a crowdsourcing platform (Prolific or equivalent) with samples collected from China, the US and the EU. The pre-registered success criterion for transferring the simulation-trained model to real deployment without retraining is a Spearman correlation $\rho > 0.6$ between synthetic and real ratings together with a calibration error < 0.15 . Ethical approval will be obtained before Stage 1 commences.

6. Conclusions

This study proposes a preference-driven AIGC framework that integrates three coordinated innovations: (i) a CGAN-based virtual-user pool that replaces costly real-user ratings with reproducible synthetic records that pass four distributional-fidelity criteria; (ii) a differentiable CART-rule signal injected into the diffusion reverse process, enabling transition of rule application from offline post-

screening to online dynamic guidance; and (iii) a closed-loop coordinator that iteratively refines rule weights from agent scoring. On the test set, the proposed model achieves FID = 29.56, PA = 0.734 and Hit Rate@3 = 0.512, improving over the strongest baseline by 13.8%, 9.4% and 19.6% respectively, while cutting training time by 78% relative to MLP and by 74% relative to Random Forest. Ablation studies confirm that each component is indispensable. The reported gains, however, are obtained under virtual-user simulation; a pre-registered two-stage Real-user validation ($n \approx 60$ in-lab, $n \approx 600$ cross-cultural online) will quantify ecological validity before deployment.

Data availability

The seed dataset D_{seed} ($N = 2,000$) is assembled from the publicly available WikiArt and ArtBench-10 corpora. The full source code, the trained CGAN / CART model weights, the synthesised virtual-user pool (10,000 agents), and all scripts required to reproduce the reported results will be released at GitHub under the MIT licence upon acceptance <https://github.com/yecaixian/year-2026/tree/main/Scientific%20Reports>. Raw annotator scores and intermediate feature files are available from the corresponding author upon reasonable request.

Funding

This research was funded by the Key Scientific Research Project of Guangzhou Xinhua University (Grant No.2025KYZDZK02) and the 2023 Guangdong Undergraduate University Teaching Quality and Teaching Reform Project (Grant No.2023CYXY003, 2023CYXY003-2).

Author contributions

Conceptualisation, data curation, formal analysis: C.Y. and L.X.; funding acquisition, investigation, methodology: J.T.; project administration, resources, software: C.Y. and J.T.; supervision, validation, visualisation, writing—original draft, Writing – review and editing: L. X. All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

Written informed consent was obtained from the three expert annotators who rated the seed images. The annotators were informed that their ratings would be used to generate synthetic virtual-user preference data and for publication in anonymized form.

References

1. Chen, H., Xiang, Q., Hu, J., et al. Comprehensive exploration of diffusion models in image generation: A survey. *Artif. Intell. Rev.* 58, 99. <https://doi.org/10.1007/s10462-025-11110-3> (2025).
2. Cao, P., Zhou, F., Song, Q., & Yang, L. Controllable generation with text-to-image diffusion models: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 48, 4771–4791. <https://doi.org/10.1109/TPAMI.2025.3646548> (2026).
3. Po, R., Yifan, W., Golyanik, V., et al. State of the art on diffusion models for visual computing. *Comput. Graph. Forum* 43, e15063. <https://doi.org/10.1111/cgf.15063> (2024).
4. Wang, B., Chen, Q., & Wang, Z. Diffusion-based visual art creation: A survey and new perspectives. *ACM Comput. Surv.* 57, 268:1–268:37. <https://doi.org/10.1145/3728459> (2025).
5. Shen, C.-C., Luo, S.-N., Fan, L., & Dai, C. Leveraging AI and diffusion models for anime art creation: A study on style transfer and image quality evaluation. *Comput. Sci. Inf. Syst.* 22, 1331–1346. <https://doi.org/CSIS241212045S> (2025).
6. Li, Z., Wang, Y., Li, C., Zhang, J., Xu, S., & Gao, Y. LFMDiff: Generation of Chinese traditional landscape paintings based on diffusion model. *npj Herit. Sci.* 13, 230. <https://doi.org/10.1038/s40494-025-02136-5> (2025).
7. Zhou, L., Jin, Y., Li, F., & Zhang, L. Construction of an AIGC paradigm for pattern generation in the style of Wuxi Jiahua. *Prog. Text. Sci. Technol.* 47, 1–5, 32. <https://doi.org/10.19507/j.cnki.1673-0356.20250905.002> (2025) (in Chinese).
8. Cao, Y., Li, S., et al. A survey of AI-generated content (AIGC). *ACM Comput. Surv.* 57 (5). <https://doi.org/10.1145/3704262> (2025).
9. Voinea, D. V. The algorithmic auteur: AI, cultural production, and the reconfiguration of audio visual media. *Soc. Sci. Educ. Res.* 256–268. <https://doi.org/10.5281/zenodo.15804554> (2025).
10. Khan, M. A., Mikalonytė, E. S., Porsdam Mann, S., Liu, P., Chu, Y., Attie-Picker, M., Buyukbabani, M. B., Savulescu, J., Hannikainen, I. R., & Earp, B. D. Personalizing AI art boosts credit, not beauty. *Technol. Soc.* 84, 102553. <https://doi.org/10.1016/j.techsoc.2025.103055> (2026).
11. Guan, S. & Feng, Y. Expressive activation enhances user satisfaction in AIGC co-creation through perceived control and meaning-making. *Sci. Rep.* 16, 20991. <https://doi.org/10.1038/s41598-026-53744-5> (2026).

12. Jin, C. An AI-generated art evaluation model that integrates computational aesthetics and cognitive psychology. *Sci. Rep.* 16, 15309. <https://doi.org/10.1038/s41598-026-42766-8> (2026).
13. Grassini, S. & Koivisto, M. Understanding how personality traits, experiences, and attitudes shape negative bias toward AI-generated artworks. *Sci. Rep.* 14, 4113. <https://doi.org/10.1038/s41598-024-54294-4> (2024).
14. Liu, W. & Kim, H.-G. The visual communication using generative artificial intelligence in the context of new media. *Sci. Rep.* 15, 11577. <https://doi.org/10.1038/s41598-025-96869-9> (2025).
15. Li, Y., Siek, H. L. & Guo, J. Factors influencing user adoption of AI painting tools based on the AIDUA model. *Sci. Rep.* <https://doi.org/10.1038/s41598-026-48130-0> (2026).
16. Zhang, H., Wu, T., & Wei, Y. Multi-view user preference modeling for personalized text-to-image generation. *IEEE Trans. Multimedia* 27, 3082–3091. <https://doi.org/10.1109/TMM.2025.3557683> (2025).
17. Liu, S., Xia, L., & Lan, Y. An empirical study on the innovation diffusion of generative artificial intelligence in audiovisual creation. *New Media Res.* (4), 17–22, 50. <https://doi.org/10.16604/j.cnki.issn2096-0360.2025.04.013> (2025) (in Chinese).
18. Luo, K., & Tao, R. Creative design of digital media art based on computer visual aids. *Paladyn* 16, 1–9. <https://doi.org/10.1515/pjbr-2025-0009> (2025).
19. Wei, Y., Zheng, Y., Zhang, Y., Liu, M., Ji, Z., Zhang, L., & Zuo, W. Personalized image generation with deep generative models: A decade survey. *Comput. Vis. Media* 11, 1141–1194. <https://doi.org/10.26599/CVM.2025.9450495> (2025).
20. Rui, T., Xiao, L., et al. Using Pix2Pix conditional generative adversarial networks to generate personalized poster content: Style transfer and detail enhancement. *J. Vis. Cult.* <https://doi.org/10.1177/14727978241307472> (2024).
21. Hu, D., Wang, E., & Arshad, M. Design of personalized creation model for cultural and creative products based on evolutionary adaptive network. *PeerJ Comput. Sci.* 11, e3288. <https://doi.org/10.7717/peerj-cs.3288> (2025).
22. Choi, W., Kim, H. Y., Choi, B., & Moon, J. Development of a fashion recommendation system with consumers' zero-party data applying the CART decision-tree model. *J. Fash. Mark. Manag.* 30, 233–257. <https://doi.org/10.1108/JFMM-07-2024-0284> (2026).
23. Kalra, A. & Brown, D. S. Can differentiable decision trees enable interpretable reward learning from human feedback? *Reinforcement Learning Journal* 4, 1887–1910. <https://doi.org/10.48550/arXiv.2306.13004> (2024).

24. Mao, Q., Ren, J., Wang, Y., Zou, C., Zheng, J., & Cao, Y. Differentiable decision tree via “ReLU+Argmin” reformulation. In *Advances in Neural Information Processing Systems (NeurIPS 2025, Spotlight)*. <https://neurips.cc/virtual/2025/poster/119074> (2025).
25. Mahfoud, M., Boukachab, G., Koziarski, M., Hernández-García, A., Bauer, S., Bengio, Y. & Malkin, N. Learning decision trees as amortized structure inference. In *ICLR 2025 Workshops: FPI*. <https://mlanthology.org/iclrw/2025/mahfoud2025iclrw-learning/> (2025).
26. Ibrahim, S., et al. Flexible modeling and multitask learning using differentiable tree ensembles. *arXiv* <https://arxiv.org/abs/2205.09717> (2022).
27. Ramachandran, S. N. & Sra, S. Trees to flows and back: Unifying decision trees and diffusion models. In *Proceedings of the 43rd International Conference on Machine Learning (ICML 2026)*. <https://icml.cc/virtual/2026/poster/62447> (2026).
28. Ye, C., Xu, L., & Tang, J. Construction of new media art model by using the CART algorithm and AIGC. In *Proc. 2024 Int. Conf. Informatics Education and Computer Technology Applications (IECA)*, 1–6. <https://doi.org/10.1109/IECA62822.2024.00007> (2024).
29. Ye, C., Xu, L., & Tang, J. Intelligent generation of new media art based on user experience design. *Commun. Appl. Nonlinear Anal.* 32 (3s), 112–132. <https://doi.org/10.52783/cana.v32.2566> (2025).
30. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. *Classification and Regression Trees* (Wadsworth International Group, Belmont, CA, 1984), pp. 168–258.
31. Song, J., Meng, C., & Ermon, S. Denoising diffusion implicit models. In *9th Int. Conf. Learning Representations (ICLR 2021)*, Virtual Event, Austria, May 3–7, 2021. <https://doi.org/10.48550/arXiv.2010.02502> (2021).
32. Wen, J., Zhu, X. R., Wang, C. D., & Tian, Z. (2022). A framework for personalized recommendation with conditional generative adversarial networks. *Knowledge and Information Systems*, 64(10), 2637–2660. <https://doi.org/10.1007/s10115-022-01719-z>[reference:0]
33. Neftci, E. O., Mostafa, H., & Zenke, F. (2019). Surrogate Gradient Learning in Spiking Neural Networks. *IEEE Signal Processing Magazine*, 36(6), 51–63. <https://doi.org/10.1109/MSP.2019.2931595>
34. Zhang, Y. et al. Research on multimedia data transmission optimization of interactive art installation integrating 6G technology: a simulation-based framework. *Discov. Comput.* 29, 235. <https://doi.org/10.1007/s10791-026-09919-9> (2026).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [informedconsent.pdf](#)