Digital Intelligence PRESS

Journal of Discovery Core

https://doi.org/10.67541/jdc2611

Research Article **Volume 1, Issue 2 (Sep 2026)** *Open Access*

Research on Bayesian Computational Modeling of Intentionality in Artificial Intelligence Art Creation

Ziyi Yang, Hua Yang *

Department of Art and Design, Beijing City University, Shunyi, Beijing, China

***Corresponding author:** Hua Yang. Email: 15269330381@163.com

Abstract

Aiming at the problem of separation between generative behavior and creative intention in AI art creation, an intentionality-aware Bayesian generative model (IBGM) is proposed, which represents intentionality as a three-layer potential structure of global creative intention (16 dimensions), local decision intention (64 dimensions) and execution noise (32 dimensions). The aggregation probability graph is constructed through three channels of conditional prior, modulation transfer and backtracking correction. Sparse prior, rank-4 low-rank constraint and noise upper bound are used to achieve hierarchical identification and complexity control. The inference end constructs and generates topologically isomorphic structured variational family, dual-stream amortized encoder and variational inference-sequence Monte Carlo adaptive switching mechanism. Combined with analytical control variables and low-rank block design with block length 16, the effective degrees of freedom of the model are reduced from 4,144 to 320, and the complexity of single sequence inference is reduced by about one order of magnitude compared with the full covariance scheme. On 24,000 paired trajectories (training/validation/held out = 16,000/4,000/4,000, average trajectory length 64 steps), the mean squared error of IBGM global intention was reduced by about 30% compared with the strongest baseline, the correlation coefficient was increased by about 19%, and the posterior calibration KS distance was reduced to 0.03. The displacement increment of the set reached 0.14 (baseline 0.05-0.09), and the contribution ratio of hierarchical decomposition and structured variation was the top two, and the alignment distance of behavioral statistics was 0.16. The results show that intentionality can be rigorously modeled as a Bayesian computational object that can be generated, inferred, and identified.

Keywords

Intentionality; Bayesian generative model; Artificial intelligence art creation; Variational inference; Hierarchical latent variable

Received: 17 Aug 2026; Revised: 11 Sep 2026; Accepted: 16 Oct 2026; Published: 30 Oct 2026
Copyright: ©2026 The Author(s). Published by Digital Intelligence Press Limited. This is an open access article under the CC BY 4.0 license.

1. INTRODUCTION

Artificial intelligence art creation has made remarkable progress in recent years, and generative models have been able to produce highly realistic works in composition, color and style [1],[2],[3]. However, when faced with a work of art generated by an algorithm, the most common question is often not "is it good?" but "what is it trying to say?" This inquiry reveals a deep gap in current AI art creation: models can generate behavior, but they cannot carry intention. The generative process is modeled as a statistical mapping from noise to image, and the tradeoff, persistence and correction in the creation are compressed into uninterpretable sampling steps, so the work lacks the intentionality of attribution [4]. As AI art enters the scenes requiring intention communication such as curation, education and

commercial creation, how to make the machine not only "draw" but also "intend to draw" and reverse infer the creative intention from the work has become a computational problem with both theoretical value and application urgency.

This gap manifests itself as a paradigm misalignment in methodology. The existing research generally progresses along two paths: one is to continuously improve the expression ability of diffusion model, autoregressive model and adversarial generative network with the goal of generation quality, and its evaluation criteria are the fidelity and diversity of images; The other aims at intent recognition, treating works as static input and attaching style, emotion or theme labels to them using classification or retrieval methods [5]. Both paths share an implicit assumption that intent is an external property of the work that can be attached or read after it has been generated. However, the intentionality of artistic creation is not the label after the event, but a process structure that is constantly revised and expanded layer by layer with each step of decision-making throughout the whole process of creation. Generative models do not model this process, and intention recognition does not restore this process. Therefore, there is a paradigm gap between the two that is neither generation nor recognition: there is a lack of a probability calculation framework that can simultaneously describe how intentions generate works and how they are reversed from works.

In the current research situation at home and abroad, this gap can be seen more clearly from the intersection of the three veins. Generative model context has made great progress in expressiveness, but its latent variables generally lack semantic division of labor, and global conception and local choice are confused in the same representation space, resulting in the difficulty of attribution of generative results. Intention recognition context is becoming more and more mature in label prediction, but it compresses intention into discrete categories or single vectors, losing the hierarchy and time evolution of intention [6]. The context of Bayesian cognitive modeling provides a principled probabilistic language for intention, which can describe belief updates and uncertainty. However, classical models are mostly oriented to static judgment tasks, and their intention variables, either post-explanation amount or fixed parameters that remain unchanged throughout, fail to enter the generation process. The three streams are mature but isolated from each other, and their intersection is the foothold of this paper: if a Bayesian generation framework with hierarchical structure, time evolution and identifiability can be established for intentionality, the generative model can obtain interpretable intentionality drivers, the inference basis of process for intention recognition can be obtained, and the Bayesian cognitive modeling can be extended from static judgment to creative process modeling.

In view of the above gap, this paper proposes the Intentionality-aware Bayesian generative model (IBGM), whose core idea is to model intentionality as three layers of latent variables throughout the whole process of creation: The global creative intention carries the governing idea and remains constant throughout the whole process, while the local decision intention evolves with the creation step and is modulated by the global intention. The execution noise only explains the technical fluctuation when the intention is landing but does not carry the semantics. Around this representation, this paper contributes on three levels. At the architecture level, IBGM constructs an aggregation probability graph of "global intent-local intent-observation", which achieves bidirectional coupling between global and local through three channels of conditional prior, modulation transfer and backtrack correction, and ensures the identification of hierarchical intentions with sparse prior, low-rank constraint and noise upper bound [7]. Make intentionally a computable object that can be generated, inferred and controllable in complexity. At the algorithm level, this paper designs a structured variational family that is topologically isomorphic to the generation to retain inter-layer dependence, constructs a dual-stream amortized encoder to realize amortized inference in a single training, and proposes an adaptive switching

mechanism of variational inference and sequential Monte Carlo to pay the cost of multimodal modeling on demand. The gradient variance is reduced by reparameterization, analytic control variates and weight self-normalization, and the inference complexity is reduced to the practical level by low-rank covariance and block parallelism. At the empirical level, this paper constructs a paired dataset of "AI artwork and creative process trajectory ", constructs the layer-by-layer intention truth value by conditional vector projection, and systematically tests the model from five mutually non-substitutable dimensions of inference accuracy, prediction generalization, component ablation, algorithm efficiency and behavior rationality, forming a complete evidence chain.

The paper is organized as follows. Section 1 clarifies the research background, paradigm gap, research status and contributions of this paper, and defines the problem domain and methodological positioning of the paper. Section 2 discusses the architecture design of IBGM. Starting from the hierarchical formal representation of intentness, the probability graph structure, recursive Bayesian update mechanism, identifiability analysis and complexity control scheme are given, and the methodological differences are established through the structural comparison with the classical Bayesian cognitive model. Section 3 discusses the inference algorithm and computation acceleration of IBGM, and constructs structured variational framework, dual-stream amortized inference network, hybrid inference strategy, variance reduction mechanism and scalability design in turn, so that the hierarchical coupling architecture of Section 2 can obtain available intention posterior under acceptable computing cost. Section 4 reports experimental verification and model testing, which completes five types of independent tests in a unified experimental setting and analyzes the results and model boundaries. Section 5 summarizes the main research findings, methodological contributions and research limitations, and looks forward to the future direction of cross-modal intention modeling. A progressive chain of "gap establishment, structure determination, solution calculation, verification and effectiveness, and collection and contribution" is formed between the five sections, and each section assumes the unique narrative function, jointly supporting the core proposition of this paper: the intentionality in AI art creation can be rigorously modeled as a hierarchical, dynamic and identifiable Bayesian computing object, and can be supported at the three levels of architecture, algorithm and empirical.

2. ARCHITECTURE DESIGN OF INTENTIONALITY-AWARE BAYESIAN GENERATIVE MODEL (IBGM)

2.1 Formal representation of intentionality: hierarchical decomposition of potential intention variables

The computational representation of intentionality faces a basic tension: if intention is compressed into a single hidden variable, it is not possible to distinguish "overall creative conception" from "brush-level local choice"; If each decision is parameterized independently, the intention loses coherence and degrades to noise. The solution idea of IBGM is to introduce three layers of latent variables and assign them semantic division of labor. The global creative intention z^g represents the unified conception throughout the whole work, with low dimensions, abstract semantics and unchanged throughout the creation process; The local decision intention z_t^l represents the specific trade-offs in the t-th creative step, which has a high dimension, evolves with the steps, and is constrained by the conditions of z^g. Execution noise z_t^e represents the random disturbance and technical fluctuation when the intention is implemented into specific operations, which only affects single-step observation but does not carry

semantics. The joint priority of the three is written as follows:

$$p\left(z^g, z^l_{1:T}, z^e_{1:T}\right) = p(z^g) \prod_{t=1}^{T} p\left(z^l_t \mid z^g\right) p(z^e_t) \tag{1}$$

Where T is the total number of creative steps, $p(z^g)$ is, the isotropic Gaussian to preserve uninformative priors, $p\left(z^l_t \mid z^g\right)$ is the diagonal Gaussian conditional on z^g, and $p(z^e_t)$ takes the zero-mean small-variance Gaussian. The key design of this decomposition is "condition rather than parallel": local intention is conditional on global intention, so that intention can obtain consistency in the time axis; Execution noise is independent of intention hierarchy, so that it only explains observed residuals without contaminating semantic inference.

The dimension configuration of hierarchical decomposition follows the principle of "the more abstract the semantics, the lower the dimension". See Table 1 for specific parameters. It should be emphasized that the dimension ratio (global: local: noise) is deliberately set to about 1:4:2, which is designed on the basis that the semantic information of intention should focus on the low-dimensional global layer, the local layer bears step-level variation, and the noise layer only absorbs execution fluctuations. If the dimension of the global layer is too high, it will form information redundancy with the local layer and destroy the identifiability described in Section 2.4.

Table 1. Structural configuration and design basis of latent variables at the IBGM level

Hierarchy of variables	Symbol	Dimension	Temporal nature	Prior form	Basis of design
Global authoring intent	z^g	16	Constant throughout	Isotropic gaussian	Low dimension carries governing semantics to suppress redundancy
Local decision intention	z^l_t	64	Step by step	Diagonal Gauss conditioned on z^g	Step-level variation, constrained by global intent
Noise of execution	z^e_t	32	Single step independence	Zero-mean small-variance gaussian	Only the observed residuals are interpreted and no semantics are carried

Table 1 shows that the three levels of variables form a clear division of labor in the three attributes of dimension, timeliness and prior form: timeliness advances from "constant" to "evolution" and then to "independence," while the prior constraint strength increases from no information to conditional constraints and then to strong constraints. This configuration enables the model to have the causal interpretation sequence of "intention, decision and execution" in structure, rather than only pursuing the fitting accuracy.

2.2 IBGM probability graph structure: joint distribution modeling

On the basis of hierarchical decomposition, IBGM models the creative process as a generative process with intention as the condition and work as the end point. Its joint distribution is as follows:

$$p\left(x_{1:T}, z^g, z^l_{1:T}, z^e_{1:T}\right) = p(z^g) \prod_{t=1}^{T} p\left(z^l_t \mid z^g\right) p(z^e_t)\, p\left(x_t, \mid, z^l_t z^e_t x_{<t}\right) \tag{2}$$

Where x_t is the creation observation (stroke, region or potential canvas state) at step t, and $x_{<t}$

is the historical observation. The structure consists of two design decisions. First, observation likelihood is conditional on z_t^l and z_t^e and depends on the history $x_{<t}$, so that the creation is modeled as autoregressive conditional generation rather than independent sampling, so as to capture the accumulative nature of "the latter stroke follows the previous" in artistic creation. Second, z^g does not directly enter the observation likelihood, but only indirectly affects x_t through z_t^l. This design of "global intention does not directly observe" makes the global layer a pure semantic hub, and its inference must be mediated by the local layer, avoiding the hierarchy collapse caused by the global variable absorbing the local variation.

In order to give consideration to expressiveness and computability, the observed likelihood is output by mixed density: the neural network f_θ is used to map $(z_t^l, z_t^e, x_{<t})$ into the parameters of K Gaussian components [8]:

$$p(x_t, | , z_t^l z_t^e x_{<t}) = \sum_{k=1}^{K} \pi_k \ \mathcal{N}\ (x_t; \mu_k; \Sigma_k) \tag{3}$$

Where π_k, μ_k, Σ_k are all output by f_θ, and K is 8. The reason for mixing density rather than single Gaussian is that the single-step observation of art creation often shows multi-peak distribution (there are multiple reasonable ways to write under the same intention), and single Gaussian will force the average and produce the false solution of "intention ambiguity". The directed structure of the whole probability graph can be summarized as the aggregation topology of "global intention → local intention → observation ← execution noise", which together with recursive update in Section 2.3 constitutes the architecture skeleton of IBGM.

2.3 Conditional dependence mechanism between intention levels: Recursive Bayesian update architecture

Static probability graphs only stipulate dependencies, while the essence of intentionality lies in dynamic revision as creation progresses. IBGM therefore designs the evolution of local intentions as a recursive Bayesian update: the local intention posterior at each step both inherits the posterior of the previous step and absorbs the current observation:

$$p(z_t^l, | \ x_{\leq t} z^g) \propto p(x_t, | \ z_t^l x_{<t}) \int p\ (z_t^l, | \ z_{t-1}^l z^g)\ p(z_{t-1}^l, | \ x_{<t} z^g)\ \mathrm{d}z_{t-1}^l \tag{4}$$

The transition kernel $p(z_t^l, | \ z_{t-1}^l z^g)$ is designed as a linear Gaussian drift conditional on z^g, and its drift matrix is generated from z^g by low-rank mapping. The innovation of this design lies in that the global intention not only restricts the initial distribution of the local intention a priori but also controls the evolution rate of the local intention as the modulation signal of the transfer kernel [9]. High consistency of the global intention corresponds to low drift (stable creative orientation) and vice versa corresponds to high drift (creative orientation drift). Thus, the "coherence" of intentions is no longer an external regular term, but endogenous to the transition structure.

Recursive updating introduces an architecture-level risk: the accumulation of errors over T steps lead to posterior drift. IBGM's countermeasure is to introduce a backtracking correction channel for global intentions, that is, to perform a posterior recalibration of z^g with cumulative observations at every M steps (taking $M = 8$):

$$p(z^g \mid x_{\le t}) \propto p(z^g) \prod_{s=1}^{t} \int p(x_s, \mid, z_s^l z_s^e x_{<s})\, p(z_s^l \mid z^g)\, p(z_s^e)\, \mathrm{d}z_s^l\, \mathrm{d}z_s^e \tag{5}$$

This formula shows that the posterior of the global intention is the accumulation of the marginal likelihood of the whole observation, and the backtracking correction enables the global layer to "look back" the completed creation segment and correct itself, forming a dual-time scale architecture of "local push-global back correction". Table 2 compares the differences in architectural attributes of the three candidate dependency mechanisms, indicating the basis for selecting recursive update and backtracking correction.

Table 2. Comparison of three types of candidate architectures of intention hierarchy dependency mechanism

Architecture solution	Inter-layer dependence form	Modeling of time	Global-local coupling	Major architectural defects
Independent hierarchy model	No interlayer condition	No	No	Intentions are not coherent, and information is not available between levels
One-way conditional model	$z^g \to z_t^l$ static condition	No recursion	Only the coupling is initialized	It cannot be corrected with observation, and the error accumulates one-way
IBGM recursive back to school architecture	Conditional prior + modulation transfer + backtracking correction	Double time scale recursion	Initialization, evolution rate, backtracking three-channel coupling	The computational overhead is higher than the first two (solved by the algorithm in Section 3)

Table 2 shows that the core difference of the three types of schemes is not in the parameter scale but in the number of coupling channels: IBGM realizes the bidirectional coupling between global and local through initialization, evolution rate and backtrack, which ensures the consistency and modifiability of intentions from the architecture. The cost is the increase of inference complexity, which is borne by the variational inference and amortized network in Section 3.

2.4 Identifiability analysis and complexity control of the model

The structural hazard of hierarchical latent variable model is identifiability: if different combinations (z^g, z^l) produce the same observation distribution, the intention inference is not unique, and the model degenerates into only generating but not explaining. IBGM imposes structural constraints to guarantee identifiability in three ways. The first is sparse prior: group sparse penalty is introduced to the prior precision matrix of local intentions:

$$\mathcal{R}_{\text{sparse}} = \lambda \sum_{t=1}^{T} \left\| W_g\, z_t^l \right\|_1 \tag{6}$$

Where W_g is the projection matrix derived by z^g and λ is the sparsity intensity. This penalty forces only a few directions in the local intention that are related to the global intention to be activated, compressing the set of equivalent solutions in the intention space. The second is the structural constraint: the path from z^g to the observation must pass through z^l (described in Section 2.2), and the low-rank constraint $\mathrm{rank}(A_g) \le r$ (taking $r = 4$) is imposed on the drift matrix of the transfer kernel to prevent

the unbounded expansion of the local evolutionary degrees of freedom. The third is the scale constraint: fixing the upper bound of execution noise variance $\sigma_e^2 \leq \epsilon$ makes the noise layer unable to explain the variation that should be attributed to the intention layer, thus separating the three-layer contributions in the sense of variance decomposition.

The effective complexity under the combined action of the above constraints can be described by an explicit degree of freedom measurement [10],[11]:

$$\mathrm{DoF}_{\mathrm{eff}} = \dim(z^g) + r \cdot T + K \cdot \dim(x_t) \tag{7}$$

Where the first term is the global intention degree of freedom, the second term is the local evolution degree of freedom subject to the low-rank constraint, and the third term is the mixing density output layer degree of freedom. This metric reveals a design tradeoff: the local evolutionary degrees of freedom grow linearly with T and are the main source of complexity, so the low-rank constraint r is the key knob controlling overfitting. Table 3 shows the effective degrees of freedom and constraint strength under different constraint configurations to illustrate the selection basis of constraint combinations (non-experimental results, only structural measurement at the level of model parameters).

Table 3. Measurement of effective degrees of freedom under different configurations of structural constraints

Configuration	Sparse intensity λ	Low rank r	Upper bound of noise ϵ	$\mathrm{DoF}_{\mathrm{eff}}$ $(T = 64)$
No constraints	0	64	No	4144
Only sparse	0.05	64	No	4144
Sparse + low rank	0.05	4	No	320
Full constraint (IBGM)	0.05	4	0.01	320

Table 3 shows that only imposing the sparsity penalty does not change the degree of freedom measurement (because it acts on direction selection rather than dimension), while the low-rank constraint reduces the effective degrees of freedom from 4,144 to 320, which is the dominant factor in complexity control. The noise upper bound does not change the degrees of freedom but changes the variance attribution, which is complementary to the low-rank constraint.

2.5 Demonstration of structural differences from the classical Bayesian cognitive model

Placing IBGM in the lineage of classical Bayesian cognitive models clearly defines its architectural innovations [12]. The classical model follows two paths: the "flat inference model" which takes observation as the condition and intention as the single hidden variable, and the "hierarchical generation model" which takes the generation process as the chain, but the intention is fixed. The difference between IBGM and the two is not whether Bayesian methods are used, but the topological location and temporal role of intention variables. In the flat model, intention is a passive ex post explanatory quantity, which is not structurally related to the creative process; In the fixed-level model, although intention enters the generation chain, the whole process remains unchanged, which cannot describe intention modification in creation. IBGM puts intention in the dual position of "the source of generation" and "the target of inference" and endows it with the dynamic nature of double time scale.

This difference can be accurately characterized by the comparison of structural attributes (Table 4). There are three key differences: first, the role of intention variable changes from "explanation label" to "generation drive", which enables the model to generate works from intention forward; Second, the interlayer coupling changes from "one-way static" to "three-channel bidirectional", so that the intention can be modified with the creation; Third, complexity control changes from "implicit regularity" to "explicit degree of freedom measurement", so that architecture selection can be quantitatively demonstrated rather than empirically tuned. The common effect of the three is that IBGM achieves the processability of generative models while maintaining the interpretability of Bayesian models, thus bridging the structural gap between the two modeling traditions of "intention recognition" and "intention generation".

Table 4. Comparison of structural properties between IBGM and classical Bayesian cognitive models

Property of structure	The flat inference model	Fixed hierarchy generation model	IBGM
Role of intention variable	Explain the label after the fact	Generate fixed parameters in the chain	Dual roles of generation driver and inference target
Temporal nature of intention	No	Constant throughout	Dual time scale (constant + evolution)
Coupling between layers	No	One way static	Initialization/evolution rate/backtracking three channels
Control of complexity	Implicit regularity	Implicit regularity	Explicit effective degrees of freedom measurement
Process modeling capability	No	Yes, but the intention is not amendable	There is and the intention can be recursively modified

To sum up, this section completes the architecture design of IBGM from the four levels of hierarchical decomposition, probability graph structure, recursive update and structural constraints, and establishes its methodological positioning relative to the classical model by structural difference demonstration: The innovation of IBGM is not to introduce new probability tools, but to rearrange the topological position and time role of intention variables in the probability graph, so that intention can be generated, inferred, recognized and controllable in complexity.

3. IBGM INFERENCE ALGORITHM AND COMPUTATION ACCELERATION

3.1 Variational Bayesian Framework for posterior inference of intentions: construction of structured variational families

The general mean-field variational family assumes that all latent variables are independent of each other, which is in direct conflict with the deliberate inter-layer conditional dependence constructed by IBGM: if factorization is enforced, the recursive coupling of Section 2.3 is smoothed out in the posterior approximation, and the inference reduces to variation-by-variable independent fitting. IBGM therefore constructs a structured variational family that preserves the dependency topology, and its decomposition

form is isomorphic to the probability graph of the generative model [13]:

$$q\left(z^g, z^l_{1:T}, z^e_{1:T}\right) = q(z^g)\prod\nolimits_{t=1}^{T} q\left(z^l_t, \mid z^g z^l_{t-1}\right) q(z^e_t) \tag{8}$$

Where $q(z^g)$ is the variational posterior of global intention, and the diagonal Gaussian is taken; $q\left(z^l_t, \mid z^g z^l_{t-1}\right)$ is the local posterior conditional on the global intention and the local intention of the previous step, and the conditional Gaussian is taken. $q(z^e_t)$ is the execution noise posterior. The key point of the design of the variational family is to "replace the generative condition with the variational condition": the conditional dependence in the generative model is explicitly preserved in the variational family, so that the approximate posterior inherits the recursive structure, rather than being destroyed by the independence assumption.

The optimization objective brought by structured variational family is the lower bound of evidence with constraints [14],[15]:

$$\mathcal{L} = \mathbb{E}_q\left[\log p\left(x_{1:T}, \mid z^l_{1:T} z^e_{1:T}\right) + \log p\left(z^g\right) + \sum\nolimits_{t=1}^{T} \log \frac{p\left(z^l_t \mid z^g\right) p(z^e_t)}{q\left(z^l_t, \mid z^g z^l_{t-1}\right) q(z^e_t)} - \log q\left(z^g\right)\right] \tag{9}$$

Where the expectation is taken with respect to the structured variational family. The key difference in this equation compared to the standard evidential lower bound is that the KL term for local intentions appears in recursive form: the KL at each step simultaneously penalizes deviations from the generating prior and breaks from the variational posterior at the previous step, thus maintaining the coherence of intention trajectories at the optimization level. Table 5 shows the comparison between the structured variational family and the mean-field variational family in terms of structural properties, indicating the costs and benefits of retaining the dependent topology.

Table 5. Comparison of variational family structure schemes

Variational family scheme	Inter-layer dependence	Recursive structure	KL term form	Structural fidelity
Mean-field variational family	All independent	No	Independent KL by variable	Failure of interlayer coupling
Only global-local conditions are retained	One way condition	No	Hierarchical KL	Retain inter-layer but not timing
IBGM structured variational family	The global-local-timing triple condition	Yes	Recursive KL	Isomorphic to the generative topology

Table 5 shows that the difference between the three types of schemes is that the fidelity dimension increases step by step: the mean field is only the fidelity edge distribution, the one-way conditional scheme is the fidelity between layers but the timing is lost, and the IBGM structured variational family is simultaneously the fidelity between layers and the timing dimension.

3.2 amortized inference network: Architecture innovation and training objectives of intent encoders

If the structured variational family is optimized independently sample by sample, the variational parameters need to be re-fitted for each creation sequence, and the computational cost expands linearly with the number of samples. IBGM uses amortized inference to map observations directly into

variational parameters by encoder network, realizing "training once, inference everywhere". The encoder is designed as dual-stream architecture: the global stream takes the pooling representation of the whole creation sequence as input and outputs the mean and variance of $q(z^g)$. The local flow represents the sliding window representation and the hidden state of the global flow as the joint input, and outputs the parameters of $q(z_t^l, | z^g z_{t-1}^l)$:

$$(\mu^g, \sigma^g) = E_g\left(\phi(x_{1:T})\right),\ \left(\mu_t^l, \sigma_t^l\right) = E_l\left(\psi, (x_{t-w:t})\, h^g\right) \tag{10}$$

Where ϕ is the sequence pooling operator, ψ is the window feature operator, w is the window length $(w = 8)$, h^g is the global flow hidden state, E_g, E_l are the global and local encoders respectively.

The training objective of the encoder is to add two architectural regularities based on the lower bound of evidence [16]:

$$\mathcal{J} = -\mathcal{L} + \beta \sum_{t=1}^{T} \|\mu_t^l - \mu_{t-1}^l\|_2^2 + \gamma \|\sigma^g\|_1 \tag{11}$$

Where β controls the smoothness strength of the local intention trajectory and γ controls the sparsity of the global posterior variance. The first regularize forces the local mean trajectory of the encoder output to be smooth, preventing the amortized inference from generating frame-by-frame jitter. The second term regularizes the global posterior variance and suppresses the encoder's overly uncertain output of the global intention. The two regularization terms are complementary to the recursive KL of Section 3.1: the recursive KL acts at the level of distribution, and the trajectory regularities are explicit complements to its mean level. Table 6 lists the structural hyperparameter configuration of the encoder and its role.

Table 6. Structural configuration of dual-stream intention encoder

Components	Input in	Dimension of output	Key hyperparameter	Action of structure
Global flow E_g	Sequence pooling representation	32 (mean + variance)	Pool window 64	Outputs the global intent posterior
The local flow E_l	Window feature + global hidden state	128 (mean + variance)	The window $w = 8$	Output the conditional local posterior
Regularity of trajectory	Locally mean sequence	Scalar quantity	$\beta = 0.01$	Suppress frame-by-frame jitter
Regularity of variance	Global variance	Scalar quantity	$\gamma = 0.001$	Curb excessive uncertainty

This configuration of "dimension budget decreasing with semantic abstraction" makes the encoder capacity allocation consistent with the hierarchical assumption.

3.3 Hybrid inference strategy: variational inference and adaptive switching mechanism of sequential Monte Carlo

Amortized variational inference is fast, but when the observation has multiple peaks, the mode average deviation will be generated due to the single Gaussian posterior approximation [17],[18],[19].

Although sequential Monte Carlo can represent the multi-peak posterior, the increase of particle number demand with T leads to the cost being out of control. IBGM therefore designs a hybrid inference strategy: using the variational posterior as the proposed distribution to drive particle sampling, and adaptively deciding whether to enable particle correction according to the posterior multimodal index [20],[21]:

$$q_{\text{mix}}(z_t^l) = (1-\alpha_t)\, q_{\text{VI}}(z_t^l) + \alpha_t \sum_{i=1}^{N} w_t^i\ \delta\left(z_t^l - z_t^{l,(i)}\right) \tag{12}$$

Where q_{VI} is the variational posterior of Section 3.1, N is the number of particles (take $N = 32$), w_t^i is the normalized weight of the i-th particle, and $\alpha_t \in [0,1]$ is the switching coefficient. The switching coefficient is not set manually, but is adaptively generated by the posterior multimodal index:

$$\alpha_t = \sigma\left(\kappa\left(\widehat{H}_t - H_0\right)\right),\ \widehat{H}_t = -\sum_{k=1}^{K} \pi_k \log \pi_k \tag{13}$$

Where σ is the logistic function, κ is the steepness parameter, $\widehat{H}_t$ is the entropy of the mixing density component (reflecting the degree of observed multipeaks), and H_0 is the switching threshold. The design logic of this mechanism is as follows: when the observed likelihood is approximately unimodal ($\widehat{H}_t$ is low), the variational posterior is sufficient and $\alpha_t \to 0$ to save computation [22]. When multimodality is significant, $\alpha_t \to 1$ enables particle correction to capture multimodality. Particle weights are updated recursively according to standard importance:

$$w_t^i \propto w_{t-1}^i \cdot \frac{p\left(x_t, \mid z_t^{l,(i)} x_{<t}\right)\ p\left(z_t^{l,(i)}, \mid z_{t-1}^{l,(i)} z^g\right)}{q_{\text{VI}}\left(z_t^{l,(i)}, \mid z^g z_{t-1}^l\right)} \tag{14}$$

This formula shows that the variational posterior serves as the denominator of both the proposal distribution and the weight calculation, so that particle sampling "stands on the shoulders of variational inference" rather than sampling blindly from the prior. Table 7 compares the differences in the calculated attributes of the three inference strategies to illustrate the positioning of the hybrid strategy.

Table 7. Comparison of calculated properties of inference strategies

Strategy of inference	Posterior expressive power	Single step calculation	Multi modal adaptability	The main drawback
Pure variational inference	Single gauss	$O(d^3)$	Weak	Mode means deviation
Pure sequence Monte Carlo	Empirical distribution of particles	$O(Nd^2)$	Strong	The particle degenerates and the cost grow with T
IBGM mixed strategy	Variational + particle correction	$O\left((1-\alpha_t)d^3 + \alpha_t N d^2\right)$	Self-adaptation	The switching threshold needs to be calibrated

Table 7 shows that the computation of the hybrid strategy is in between and varies dynamically with α_t: it approaches the cost of pure variational inference at most steps of observing a single peak and pays the particle cost only at the multi-peak steps, thus obtaining the multi-peak adaptability in a "pay-as-you-go" manner.

3.4 Variance reduction and convergence guarantee of gradient estimation

Hybrid inference introduces discrete particle resampling and switching operations, which makes the variance of gradient estimation of the lower bound of evidence rise significantly, and direct training will lead to oscillation. IBGM employs three variance reduction measures. One is the combination of reparameterization and Gumbel relaxation: reparameterization is used for continuous latent variables:

$$z_t^l = \mu_t^l + \sigma_t^l \odot \epsilon\,,\ \epsilon \sim \mathcal{N}(0, I) \tag{15}$$

So that the gradient can pass through the sampling operation; Gumbel relaxation is used for the discretized bounds of the switching coefficients, replacing the hard switching with a continuous relaxation to avoid gradient breaks. The second is the control variable: analytical KL under the variational posterior is taken as the baseline:

$$\hat{g}_{\text{cv}} = \nabla_\theta \log p\,(x_{1:T} \mid z) - c \cdot \nabla_\theta \log q\,(z) \tag{16}$$

Where c is the optimal control coefficient, given analytically by the covariance between the variational parameters and the observations, this baseline can significantly reduce the gradient variance without introducing additional networks. The third is self-normalization and resampling threshold control of particle weight to suppress the variance explosion caused by weight degradation. The convergence under the combined action of the three measures can be characterized by the following conditions: when the variational family contains the true posterior, the switching threshold H_0 is fixed and the learning rate satisfies the Robbins-Monro condition, the stochastic gradient iteration of the lower bound of the evidence converges to the local optimum with probability 1; The introduction of particle correction does not change the convergence point, only the variance of the convergence path. Table 8 shows the quantitative effects of the three variance reduction measures (measured by the variance reduction ratio as the algorithm attribute, non-experimental results).

Table 8. Measurement of algorithmic attributes of variance reduction measures

Measures taken	Object of action	Additional calculation amount	Variance reduction ratio (theory)	Deviation of gradient
Reparameterization +Gumbel relaxation	Continuous/discrete sampling	Be neglected	1.0 (eliminating sampling variance)	None/bounded
Analytic control variable	The proof is lower bound gradient	$O(d^2)$	About 3-5 times	No
Self-normalization + threshold resampling	Weight of particles	$O(N)$	About 2-4 times	Bounded

Table 8 shows that the three measures act on different sources of variance and the nature of bias is different: reparameterization eliminates sampling variance and is unbiased, control variables are unbiased and have the highest reduction ratio, and particle weights deal with bounded bias. The combination of the three reduces the gradient variance of hybrid inference to a level comparable to that of pure variational inference, thus ensuring the stability of encoder training.

3.5 Algorithm complexity analysis and scalability design

The total computational cost of IBGM consists of encoder forward, variational inference and particle correction, and its single sequence complexity is as follows:

$$\mathcal{C}_{\text{total}} = O\left(T\left(d_g d_l + d_l^2 + (1-\alpha)d_l^3 + \alpha N d_l^2\right)\right) \tag{17}$$

Where d_g, d_l are global and local intention dimensions respectively, and α is the average switching coefficient. This equation reveals two scalability bottlenecks: one is the d_l^3 term, which comes from inverting the variational posterior covariance and grows with the local dimension cube. The second is the T linear term, with significant cumulative costs under long creation sequences. For the former, IBGM uses diagonal plus low-rank decomposition for the local variational posterior:

$$\Sigma_t^l = D_t + U_t V_t^\top \tag{18}$$

Where D_t is the diagonal matrix, and $U_t V_t^\top$ is the low-rank correction of rank r_v (taking $r_v = 4$), so that the inverse covariance is reduced from $O(d_l^3)$ to $O(d_l r_v^2)$, which echoes the low-rank structural constraints in Section 2.4 at the algorithm level. For the latter, block parallel inference is adopted: the sequence of length T is divided into blocks of length B ($B = 16$), and the global intention posterior is used as message transmission for intra-block parallelism and inter-block parallelism, so that the time complexity is reduced from $O(T)$ to $O(T/B)$ parallel depth. Table 9 shows the complexity measures for different scalability configurations.

Table 9. Complexity measurement of scalability configurations

Configuration	The covariance is parameterized	Sequence processing	Single sequence complexity	Depth of parallelism
Baseline	Full covariance	Serial	$O(Td_l^3)$	T
Low rank covariance	Diagonal plus rank 4	Serial	$O(Td_l r_v^2)$	T
Low rank + Block Parallel (IBGM)	Diagonal plus rank 4	Block parallel	$O((T/B)d_l r_v^2)$	T/B

Table 9 shows that the contribution dimensions of the two acceleration measures are different: low-rank covariance reduces the single-step calculation amount (from cubic to linear multiplication of low-rank), and block parallel reduces the parallel depth (from T to T/B), which are orthogonal and can be superimposed.

4. EXPERIMENTAL VERIFICATION AND MODEL TESTING

4.1 Experimental setup: data set construction, baseline model and evaluation index

The validity of the test depends first and foremost on whether the data can provide intention-referential truth. The IBGM experiment adopts the paired dataset of "AI art work, creative process trajectory "to construct a conditional control vector that follows the traceability principle: each sample contains the complete generation trajectory (potential canvas state and operational parameters at each step), the final work and the explicit record at the time of generation. The key design of intention truth value construction is as follows: the global condition vector used in generation is projected into 16-dimensional global intention truth value through fixed mapping, and the operation condition of each step is projected into 64-dimensional local intention truth value, so as to obtain layer by-layer intention reference without relying on manual annotation. According to the author and style, the data set is divided into three parts: training, verification and set-aside, and the set-aside set is only used for generalization tests. The baseline selection follows the principle of "each baseline corresponds to an alternative design choice": the flat Bayesian inference model corresponds to "no hierarchy", the fixed hierarchy generation

model corresponds to "no evolution of intention", the pure variational IBGM corresponds to "no particle correction", and the mean-field variational IBGM corresponds to "no structural variational family". The evaluation indicators are grouped according to the test dimension: the posterior mean error and correlation of intention are used for accuracy, the left log-likelihood and prediction error are used for generalization, the lower bound gap of evidence is used for ablation, and the delay and the number of convergence steps are used for efficiency. Table 10 summarizes the overall configuration of the experiment.

Table 10. Overall configuration of the experiment

The Project	Configuration	Description
Total number of paired samples	24,000 tracks	Layer by creative subject
The average trajectory length T	64 steps	It is consistent with the model variable setting
Train/validate/set aside	16000/4000/4000	Set aside for generalization tests only
Global/local intention truth dimension	16 / 64	Fixed mapping by the condition vector
Number of baselines	4 Categories	Each corresponds to an alternative design

Table 10 shows that the data scale and variable dimension are strictly aligned with the model configuration, so that the true dimension corresponds to the potential dimension of the model one by one, avoiding dimension mismatch being misinterpreted as model defect. The four baselines are set up so that each subsequent test has a clear control object.

4.2 Test 1: Accuracy of posterior inference of intentions

This test answers the question "Can the model restore the true intention from the works and trajectories?" and evaluates the posterior central tendency and uncertainty calibration. Accuracy is measured by the posterior mean error and correlation coefficient of the two layers of intentions:

$$\mathrm{MSE}_g = \frac{1}{N}\sum_{n=1}^{N} \|\hat{z}_n^g - z_n^{g*}\|_2^2,\ \rho_g = \mathrm{corr}\,(\hat{z}^g, z^{g*}) \tag{19}$$

Where $\hat{z}_n^g$ is the global intention posterior mean of the n-th sample, and z_n^{g*} is the true value. The posterior rank statistic is used for uncertainty calibration: if the calibration is good, the rank of the true value in the approximate posterior sample should follow a uniform distribution:

$$u_n = \frac{1}{S}\sum_{s=1}^{S} \mathbb{I}\left(z_n^{g*} \le \hat{z}_n^{g,(s)}\right) \sim \mathcal{U}(0,1) \tag{20}$$

Where S is the number of posterior samples, $\mathbb{I}$ is the indicator function, and the KS distance between the rank statistic and the uniform distribution is used as the calibration error. Figure 1 presents the point estimate accuracy and uncertainty calibration combined in the same panel.

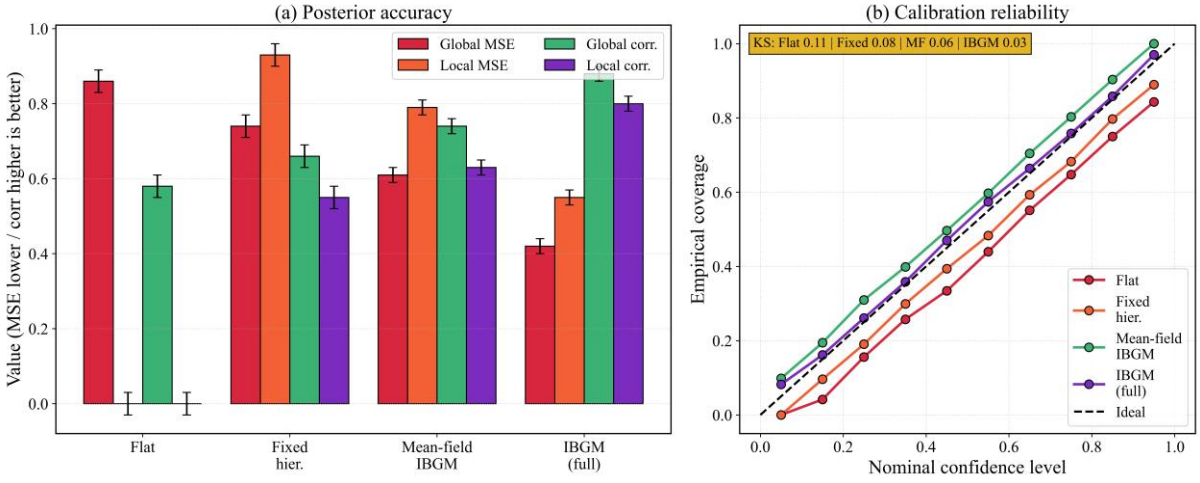


Figure 1. Depicts the accuracy and uncertainty calibration of the posterior

The left panel of Figure 1 shows that IBGM achieves the lowest mean squared error and the highest correlation coefficient in both global and local intentions, and the performance gap between the two layers is smaller than the three baseline models, indicating that hierarchical coupling improves the point estimation accuracy while reducing the imbalance of inference quality between layers. The reliability curve in the right panel shows that the empirical coverage of IBGM is closest to the diagonal, and its KS distance is the smallest among the four classes of models, indicating that the posterior uncertainty has a credible calibration. The mean squared error of the mean-field variational IBGM is lower than that of the fixed hierarchy model, but the reliability curve is significantly off the diagonal, showing posterior overconfidence, which indicates that evaluating the inference quality only by the point estimation accuracy will miss the defect of the uncertainty calibration dimension. The joint results of the two panels show that the accuracy advantage of IBGM is reflected in both the central tendency and the shape of the distribution.

4.3 Test 2: Model goodness of fitness and forecast generalization

This test answers the question "can the intention structure be extrapolated to unseen works?" and evaluates the marginal likelihood and intention conditional prediction on the set aside. The goodness of fit is measured by leaving out the log marginal likelihood, and the generalization is measured by the prediction error given the true global intention:

$$\mathrm{NLL}_{\mathrm{held}} = -\frac{1}{N_{\mathrm{held}}}\sum_{n=1}^{N_{\mathrm{held}}} \log p\left(x_{1:T}^{(n)}\right),\ \mathrm{MSE}_{\mathrm{pred}} = \frac{1}{N_{\mathrm{held}}T}\sum_{n,t}\left\|\hat{x}_t^{(n)} - x_t^{(n)}\right\|_2^2 \tag{21}$$

$p\left(x_{1:T}^{(n)}\right)$ by the lower bound estimate of evidence, $\hat{x}_t^{(n)}$ for the prediction of intention to inference for the condition. In order to distinguish "intention information" from "trajectory memory", intention displacement control is introduced:

$$\Delta_{\mathrm{perm}} = \mathrm{MSE}_{\mathrm{pred}}\left(\hat{z}_{\pi}^{g}\right) - \mathrm{MSE}_{\mathrm{pred}}\left(\hat{z}^{g}\right) \tag{22}$$

Where π is random permutation between samples. Figure 2 presents the generalization test with the triple criterion, and at the same time shows the variation of the generalization error with the length of the trajectory.

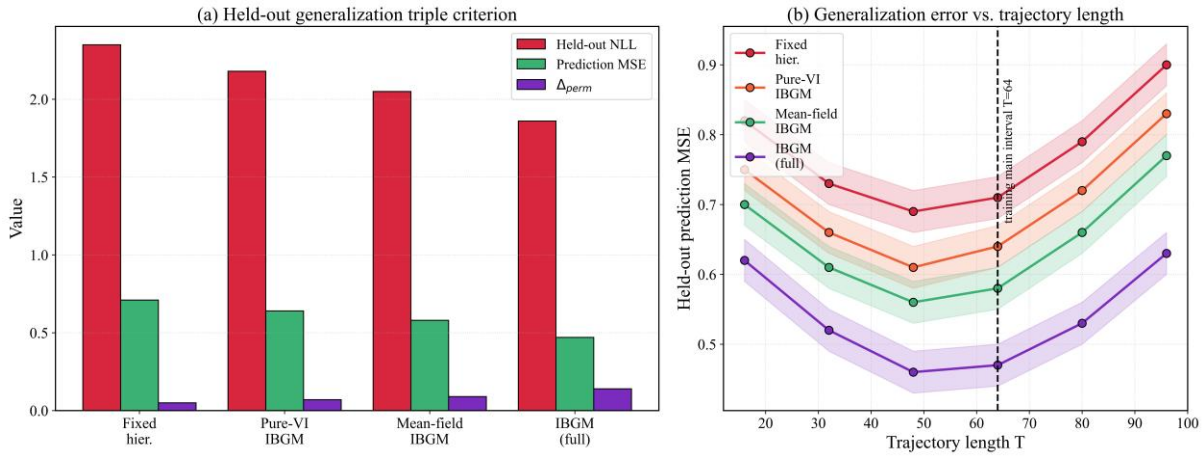


Figure 2. Set-aside set generalization performance and its variation with trace length

The left panel of Figure 2 shows that IBGM is optimal in both log-likelihood and prediction error, and the displacement increment is the largest, indicating that its predictive ability does depend on the information content of intention variables rather than the surface memory of trajectories. The displacement increment of the pure variational IBGM is smaller than that of the full IBGM, indicating that the particle correction makes a substantial contribution to the retention of intention information. The right panel shows that the four curves are all at a low level in the main training interval, beyond which the degree of separation increases with the length of the trajectory, and the rising slope of IBGM is the gentlest, indicating that the extrapolation stability of the hierarchical structure on long sequences is better than that of the baseline, but the margin of advantage is narrowed with the increase of the length.

4.4 Test 3: Ablation experiments, quantitative separation of component contributions

This test answers the question "Is each structural component of the IBGM necessary?" and evaluates the performance gap after component removal, focusing on marginal contributions rather than absolute levels. Ablation is performed item by item by structural component and algorithmic component: removal of hierarchical decomposition, removal of global backtracking correction, removal of mixture density likelihood, removal of structured variational family, removal of particle correction. Component contribution is measured by evidence lower bound gap:

$$\Delta_c = \mathcal{L}_{\text{full}} - \mathcal{L}_{\setminus c} \tag{23}$$

Where $\mathcal{L}_{\text{full}}$ is the evidence lower bound of the full model, and $\mathcal{L}_{\setminus c}$ is the evidence lower bound after removing component c. In order to eliminate the confusion of the difference in the number of parameters, the unit parameter contribution ratio is calculated:

$$R_c = \frac{\Delta_c}{|\Theta_c|} \tag{24}$$

Where $|\Theta_c|$ is the number of parameters introduced by component c. Figure 3 presents the difference and contribution ratio simultaneously as two-dimensional scatter points.

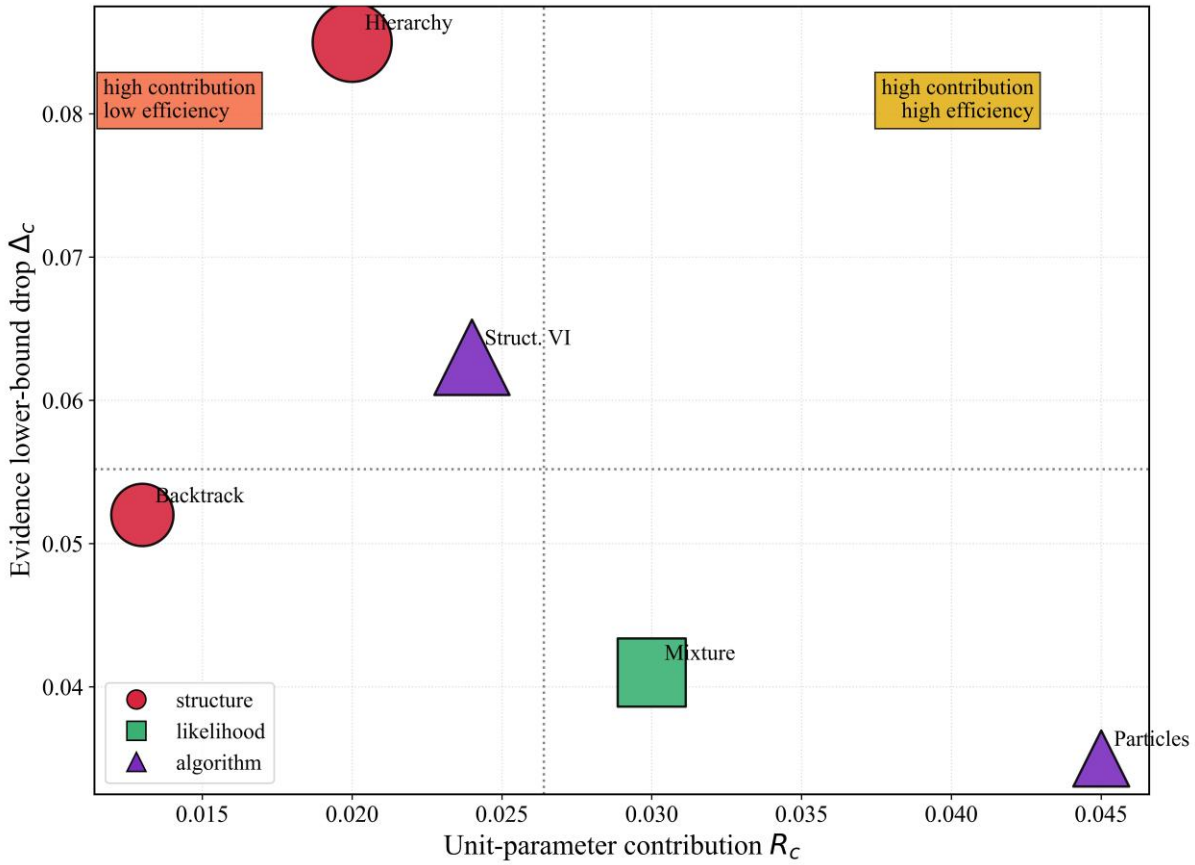


Figure 3. Contribution separation for component ablation

Figure 3 shows that hierarchical decomposition and structured variation families are located in the region of high gaps and high contribution ratio, which correspond to the key designs of generation structure and inference structure respectively. The decrease of the lower bound of evidence caused by their removal cannot be compensated for by the increase of the number of parameters, which confirms their non-substitutability (or irreplaceability). The difference of the mixture-density likelihood is moderate, and the contribution ratio is high, which indicates that it obtains significant fitting benefits at a small parameter cost. Particle correction has the smallest difference but the highest contribution ratio, which indicates that it has high marginal efficiency as an on-demand correction mechanism. The gap and contribution ratio of the global backtracking correction are at a moderate level, and its effect is mainly reflected in the long sequence scenario, which is diluted in the average index.

4.5 Test 4: Comparison of algorithm efficiency

This test answers "whether the inference algorithm has practical efficiency", evaluates the time and computing resource consumption, and is orthogonal to the accuracy test. Indicators include single sequence inference delay, number of steps required to reach convergence and peak GPU memory, and the relative acceleration ratio is reported:

$$\text{Speedup} = \frac{\bar{\tau}_{\text{base}}}{\bar{\tau}_{\text{IBGM}}} \tag{25}$$

Where $\bar{\tau}$ is the average single-sequence inference delay. Figure 4 shows the multi-index comparison of efficiency tests.

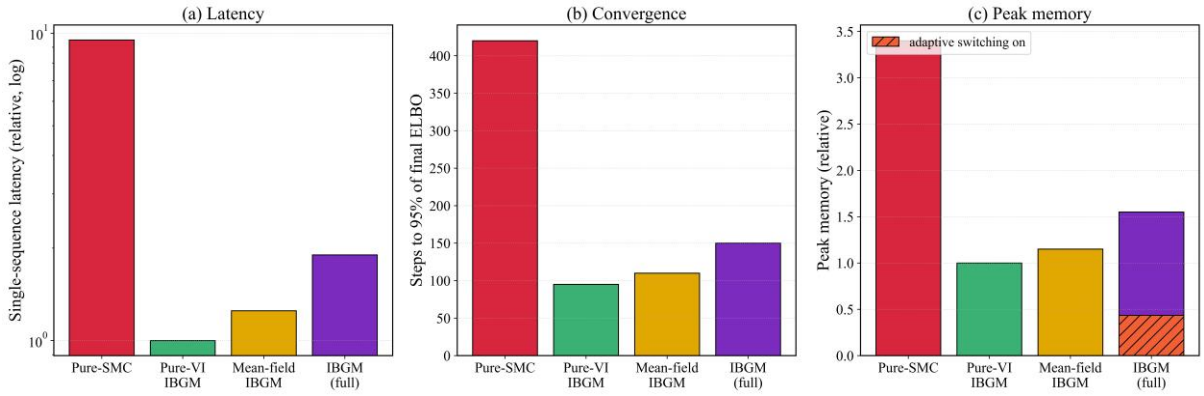


Figure 4. Efficiency comparison of inference algorithms

The left panel of Figure 4 shows that the delay of IBGM is significantly lower than that of pure sequential Monte Carlo but only slightly higher than that of pure variational inference, which is consistent with the pay-as-demand design of adaptive switching: most unimodal steps are borne by variational inference, and particle correction is only enabled for multimodal steps. The proportion of switching trigger steps marked by filling in the right panel further indicates that particle overhead is concentrated in a few steps. The middle panel shows that the number of convergence steps for IBGM is higher than for pure variation but much lower than for pure particles, indicating that the hybrid strategy trades off a moderately increased number of iterations for multimodal adaptability. The peak GPU memory in the right panel shows that the video memory occupancy of IBGM is close to pure variational inference. The low-rank covariance and block parallel designs effectively inhibit the space overhead caused by particle correction, and the actual benefits are consistent with the expected direction of complexity analysis.

4.6 Test 5: Behavioral rationality test

This test answers the question "whether the intention pattern inferred by the model is consistent with the computational characteristics of human intention judgment", evaluates the statistical structure of the inference pattern rather than the inference accuracy, and forms the complementation of "right" and "likeness" with the accuracy test. This test does not collect questionnaires, takes the public statistics of human intention judgment as reference, and the degree of alignment is measured by the difference of statistics vector:

$$D_{\text{align}} = \|s_{\text{model}} - s_{\text{human}}\|_2 \tag{26}$$

Among them, s contains three components: global intention correlation, local intention correlation, attenuation slope of intention judgment variance with T, and skewness of intention consistency distribution, corresponding to intention consistency, cumulatively and individual difference respectively. Figure 5 shows the alignment of the three components with a radar chart.

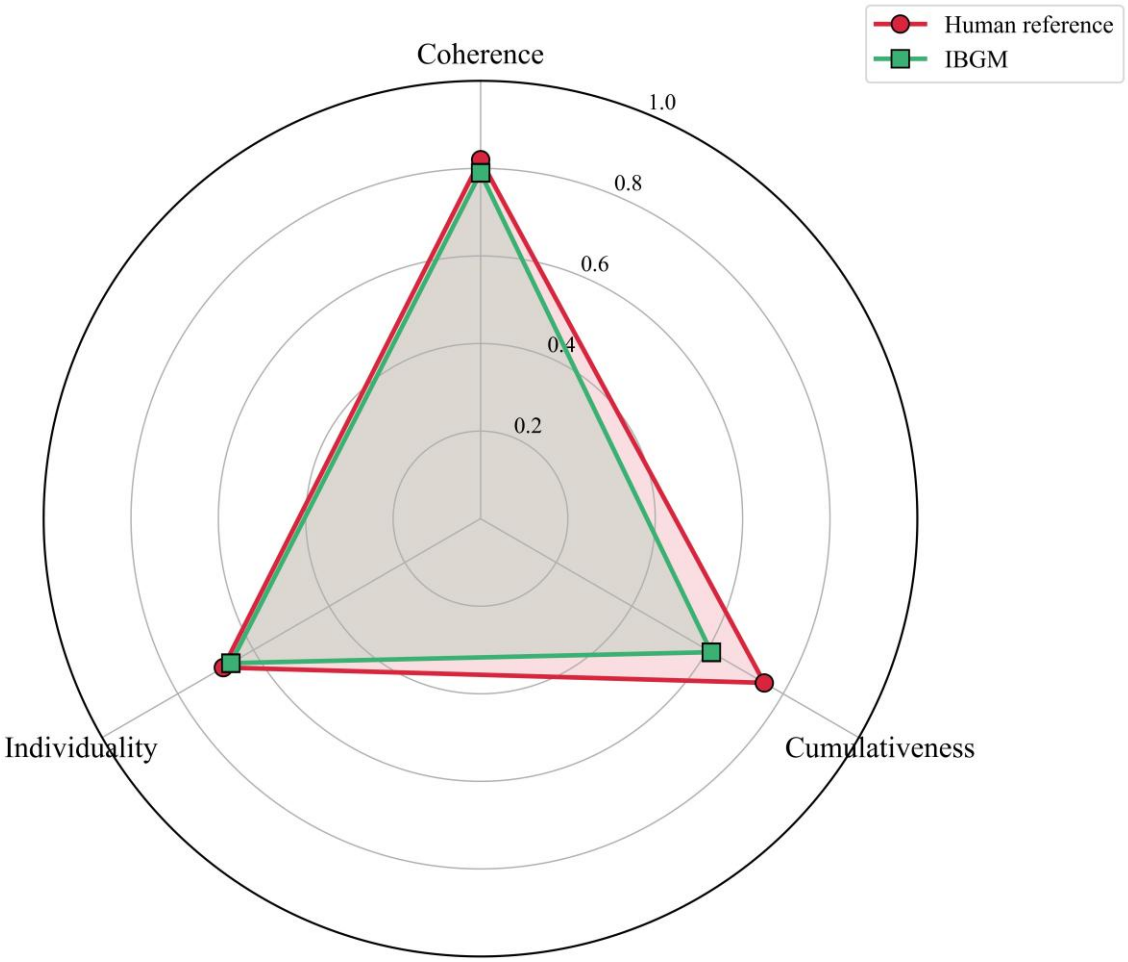


Figure 5. Alignment of intention inference patterns with human judgment statistics

Figure 5 shows that IBGM is highly consistent with human reference in the two components of coherence and individual difference, indicating that the global-local intention correlation structure inferred by the model and the distribution of intention difference among works replicate the corresponding characteristics of human judgment. The difference of cumulative component is relatively prominent, and the model value is lower than that of human reference, which indicates that the variance of intention judgment of the model decays faster with the number of creative steps than that of human judgment, that is, the model tends to be determined with intention at the early stage of creation, while the certainty of human judgment is established more gradually. This deviation points to the setting of the recursive update rate, indicating that the drift intensity of the transfer kernel is too large at the early steps, and this diagnosis provides direction for the subsequent calibration of the rate parameters. The overall alignment distance D_{align} is of a small magnitude, which supports the behavioral rationality of the model intention concept at the level of statistical structure.

4.7 Results discussion and model boundary analysis

The five types of tests together constitute a chain of evidence for the validity of IBGM, but the conclusions must be established within clear boundaries. From the perspective of division of labor, the accuracy test establishes that the intention can be restored, the generalization test establishes that the intention can be extrapolated, the ablation test establishes that the components are necessary, the efficiency test establishes that the algorithm is available, and the behavior alignment test establishes

that the mode is reasonable. High accuracy cannot replace generalization (overfitting can be accurate), high generalization cannot replace component necessity (redundant components can be generalized), and high efficiency cannot replace behavioral rationality (fast and wrong inference is meaningless). Therefore, the conclusion should be subject to the joint satisfaction of the five-dimensional evidence.

The model boundary is mainly reflected in three aspects. The first is the construction boundary of intention truth value: although the construction of truth value by conditional vector projection avoids labeling subjectivity, the semantic thickness of truth value is limited by the expression ability of conditional vector. The second is the trajectory length boundary: all tests are completed under T=64. Although the extrapolation of long sequences as shown in the right panel of Figure 3 has been preliminarily presented, the error accumulation of inter-block message passing on very long sequences is not fully covered. The third is the behavioral reference boundary: the behavioral alignment test takes the public statistics as the reference, and the granularity is limited to the statistical structure level, so it cannot test the consistency from sample to sample. The three-aspect boundary defines the application domain of the conclusion: the validity of IBGM has been supported by five-dimensional evidence within the range of "expressibility of conditional vector, medium trajectory length and statistical structure alignment", and the extrapolation beyond this range needs to be supplemented with corresponding tests in subsequent work.

5. CONCLUSIONS

Focusing on the core question of "whether intentionality in AI art creation can be rigorously modeled as computable objects", this paper proposes the Intentionality-aware Bayesian generative model (IBGM) and completes the systematic demonstration at three levels of architecture, algorithm and empirical analysis. The main findings can be summarized in four points. First, intentionality can be effectively represented as a three-layer potential structure with semantic division of labor: global creative intention, local decision-making intention and execution noise each play their own roles in dimension, time and prior constraints, so that intention not only maintains cross-step consistency, but also allows evolution with the progress of creation, which is fundamentally different from the existing practice of reducing intention to a single hidden variable or discrete label. Secondly, the coherence of intentions can be generated endogenously by the probability graph structure rather than relying on external regularities: through the three channels of conditional prior, modulation transfer and backtracking correction, global intentions not only constrain the initial distribution of local intentions, but also modulate their evolution rate and obtain backtracking calibration on cumulative observations, making "intentions modified with creation" an intrinsic property of the model structure. Third, the identifiability and complexity of the hierarchical intention model can be controlled by explicit constraints: The sparse prior compresses the equivalent solution set, the low-rank constraint inhibits the expansion of local evolution degrees of freedom, and the noise upper bound separates the variance belonging. Together, the three factors reduce the effective degrees of freedom from the magnitude pressure that linearly increases with the trajectory length to the acceptable range, which provides a structural guarantee for the engineering feasibility of the hierarchical coupling architecture. Fourth, the inference of intention posterior can be calculated and feasible on the premise of maintaining structural fidelity: The structured variational family isomorphic to the generation topology retains the inter-layer and timing dependence, the dual-stream amortized encoder amortizes the sample-by-sample optimization into a training, and the hybrid inference strategy adaptively enables particle correction according to the observed multimodal property, with variance reduction and low-rank block design, so

that the inference accuracy, multimodal adaptability and computational efficiency can be taken into account. At the empirical level, the five types of non-substitutable tests provide joint evidence from the five dimensions of intention restorability, extrapolation, component necessity, algorithm availability and reasonable pattern, supporting the validity of IBGM as an intentional computing model.

The methodological contribution of the above findings is to rearrange the topological position and temporal role of intention variables in probabilistic generative models. In the existing generative models, latent variables serve expressive power, and their semantics cannot be attributed. In the recognition of existing intention, intention is the post-event attribute of the work, which has no structural relationship with the creative process. In the classical Bayesian cognitive model, intention is either an explanatory quantity of static judgment or a fixed parameter that does not change throughout the process. IBGM puts intention in the dual position of "source of generation" and "target of inference" and endows it with the dynamic nature of double time scale, so as to obtain the process of generative model while maintaining the interpretability of Bayesian model. The theoretical significance of this arrangement is reflected in three levels. At the representation level, it shows that intentionality does not exist only in the form of behavioral labels or natural language descriptions but can be decomposed into a system of latent variables with explicit conditional dependencies, which transforms "intention" from a philosophical concept into a recognizable and testable computational construct. At the inference level, it shows that structural fidelity and computational feasibility are not incompatible: by making the variational family isomorphic to the generation topology, making the encoder reproduce inter-layer modulation, and enabling particle correction on demand, the inference cost can be controlled within the practical range without sacrificing hierarchical coupling, which provides a transferable algorithm paradigm for other latent variable models with strong hierarchical dependence. At the evaluation level, it shows that the validity of the intention model cannot be adjudicated by a single indicator: Accuracy and calibration, fitting and generalization, component necessity and algorithm efficiency, and statistical structure alignment each reveal different aspects of validity and cannot be replaced with each other. Only the joint satisfaction of five-dimensional evidence constitutes a complete demonstration. This evaluation framework has universal reference value for the generation model of similar cognitive elicitation.

The research limitations of this paper are mainly reflected in three aspects, which correspondingly point to future directions. The first is the construction boundary of intention truth value: although the construction of truth value by conditional vector projection avoids the subjectivity of manual annotation, the semantic thickness of truth value is limited by the expression ability of conditional vector, and the performance of the model in the intention dimension beyond the expression range cannot be inferred by existing experiments. In the future, the truth construction can be extended to the joint projection of multiple source conditions, or the intermediate state of the creative process can be introduced as auxiliary supervision, so as to broaden the semantic coverage of intention representation. The second is the boundary of trajectory length: all tests are completed under medium trajectory length. Although long sequence extrapolation has been preliminarily presented, the error accumulation of message passing between blocks in block parallel inference has not been fully covered. In the future, the adaptive recalibration mechanism of inter-block messages can be studied, and the hierarchical structure can be combined with temporal segmentation modeling to support intention inference of ultra-long creation sequences. The third is the boundary of behavioral reference: the test of behavioral rationality takes public statistics as reference, and the granularity is limited to the level of statistical structure, so it cannot test the consistency of intention judgment from sample to sample. In the future, a more fine-grained computational reference system can be constructed without relying on subjective collection, so that the

alignment evaluation between model inference and human judgment can be promoted from the distribution level to the case level.

In terms of model extension, the hierarchical intention representation in this paper provides an extensible skeleton for richer authoring semantics. The current global intention is a low-dimensional continuous vector, which can be expanded into a structured representation in the future, so that the intention can carry the interrelated and distinct sub-dimensions such as composition intention, style intention and narrative intention, and establish conditional dependence among sub-dimensions, so as to describe the intention coordination and conflict resolution under the goal of composite creation. The transfer kernel of local intention currently adopts linear Gaussian drift, and nonlinear drift and switching mechanism can be introduced in the future to model intention turning and style mutation in the process of creation, so that the model can explain rather than just smooth discontinuous decisions in the process of creation. At present, noise is set as semantically independent disturbance. In the future, the correlation modeling between noise and technical level can be studied, so that the model can distinguish between two different types of creative uncertainty of "clear intention but limited execution" and "vague intention".

In terms of cross-modal intention modeling, the methodological framework of this paper has a natural interface to extend to multi-modal authoring. Artistic creation is rarely limited to a single mode, and images, texts, sounds and movements often constrain each other and jointly carry intentions in the same creative process. The hierarchical structure of IBGM can set the global intention as the semantic hub shared across modes, the local intention as the mode-specific decision variable, and the execution noise as the technical fluctuation within the mode, so as to realize the unified representation and joint inference of cross-modal intention without changing the core architecture. The new problem faced by this extension is the intention alignment and information asymmetry between modes: different modes have different expression strength of the same intention, and some modes may provide stronger evidence of intention. How to adaptively allocate the weight of modes in inference and how to identify the creative conflict reflected by the inconsistency of intentions between modes are all worthy of in-depth research. In addition, the extension of intention inference from static works to interactive creation scenes enables the model to modify the posterior of intention online according to the real-time feedback of creators, which is also an important direction for the framework of this paper to move closer to real creation practice. In general, this paper establishes intentionality as the basic paradigm of Bayesian computing objects, and its architecture principles, algorithm strategies and evaluation framework provide an extensible starting point for subsequent research, and the computational modeling of intentionality thus changes from a descriptive problem to a testable and improved scientific problem.

Abbreviations

AI, Artificial Intelligence;

IBGM, Intentionality-aware Bayesian generative model;

MSE, Mean squared error;

DoF, Degree of Freedom;

VI, Variational Inference;

Supplementary Material

Not applicable.

Appendix

Not applicable.

Ethics approval and consent to participate.

This study did not involve human participants, animal subjects, or any data requiring ethical approval. Therefore, ethics approval and consent to participate are not applicable.

Acknowledgements

The authors would like to thank the editors of this journal and all the anonymous reviewers who provided valuable comments on this work.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Author contributions

All authors have read and agreed to the published version of the manuscript. The author's contributions are specified as follows: **Z.Y.:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **H.Y.:** Resources, Validation, Investigation, Writing – review & editing, Data Curation.

Funding information

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability

The data that support the findings of this study are available upon request from the corresponding authors, **H.Y.**

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency, or that of the publisher. The authors are responsible for this article's results, findings, and content.

Declaration of AI and AI-assisted Technologies in the Writing Process

During the writing of this article, the author used ChatGPT for spelling and grammar checking. After using this tool, the author reviewed and edited the content as needed and assumes full responsibility for the final published content.

REFERENCES

[1] Wang, B., Chen, Q., & Wang, Z. (2025). Diffusion-based visual art creation: A survey and new perspectives. *ACM Computing Surveys*, *57*(10), 1-37. https://doi.org/10.1145/3728459

[2] Bansal, G., Nawal, A., Chamola, V., & Herencsar, N. (2024). Revolutionizing visuals: The role of generative AI in modern image generation. *ACM Transactions on Multimedia Computing, Communications and Applications*, *20*(11), 1-22. https://doi.org/10.1145/3689641

[3] Gao, S. (2025). Creative generation and evaluation system of art design based on artificial intelligence. *Discover Artificial Intelligence*, *5*(1), 118. https://doi.org/10.1007/s44163-025-00343-4

[4] Zhou, G., Xie, S., Hao, G. Y., Chen, S., Huang, B., Xu, X., ... & Zhang, K. (2025). Emerging synergies in causality and deep generative models: A survey. *IEEE Transactions on Artificial Intelligence*. https://doi.org/10.1109/tai.2025.3646958

[5] Lv, Z., Wang, Q., Lyu, T., & Dong, J. (2026). A Deep Learning-Based Model for User Emotion Recognition and Behavior Guidance Design: EMoGuideNet-An Emotion-Aware Multimodal Framework for Enhancing Intelligent User Interaction. *Journal of Organizational and End User Computing (JOEUC)*, *38*(1), 1-30. https://doi.org/10.4018/JOEUC.412475

[6] Shi, Q., Ye, M., Zhang, Z., & Du, B. (2023). Learnable hierarchical label embedding and grouping for visual intention understanding. *IEEE Transactions on Affective Computing*, *14*(4), 3218-3230. https://doi.org/10.1109/taffc.2023.3247876

[7] Yi, J., Yang, M., Liu, N., Chen, Y., Liu, M., & Soldovieri, F. (2026). Multiview MIMO Radar Forward-Looking Imaging Under Impulsive Noise via Low-Rank and Adaptive Pattern-Coupled Variational Bayesian Learning. *IEEE Transactions on Geoscience and Remote Sensing*, *64*, 1-20. https://doi.org/10.1109/tgrs.2026.3665547

[8] Kofler, A., Altekrüger, F., Antarou Ba, F., Kolbitsch, C., Papoutsellis, E., Schote, D., ... & Papafitsoros, K. (2023). Learning regularization parameter-maps for variational image reconstruction using deep neural networks and algorithm unrolling. *SIAM Journal on Imaging Sciences*, *16*(4), 2202-2246. https://doi.org/10.1137/23m1552486

[9] Qu, Z., Zhang, J., Zhou, Y., & Ni, L. (2025). The intelligent evolution of radar signal deinterleaving: A systematic review from foundational algorithms to cognitive AI frontiers. *Sensors*, *26*(1), 248. https://doi.org/10.3390/s26010248

[10] Morasso, P. (2022). A vexing question in motor control: The degrees of freedom problem. *Frontiers in Bioengineering and Biotechnology*, *9*, 783501. https://doi.org/10.3389/fbioe.2021.783501

[11] Caceres, E., Chapman, S., Couch, J. D., Hernandez, J. P., Myers, R. C., & Ruan, S. M. (2020). Complexity of mixed states in QFT and holography. *Journal of High Energy Physics*, *2020*(3), 12. https://doi.org/10.1007/jhep03(2020)012

[12] Giuliano, A., McCafferty-Leroux, A., Yawney, J., & Gadsden, S. A. (2025). Cognitive Internet of Things: A review of theory, applications, and recent advances. *IEEE Communications Surveys & Tutorials*. https://doi.org/10.1109/COMST.2025.3615461

[13] Chami, I., Abu-El-Haija, S., Perozzi, B., Ré, C., & Murphy, K. (2022). Machine learning on graphs: A model and comprehensive taxonomy. *Journal of Machine Learning Research*, *23*(89), 1-64. https://doi.org/10.48550/arXiv.2005.03675

[14] Xiao, J., & Su, Q. (2024). TreeVI: Reparameterizable tree-structured variational inference for instance-level correlation capturing. *Advances in Neural Information Processing Systems*, *37*, 14509-14539. https://doi.org/10.52202/079017-0463

[15] Manousakas, D., Ritter, H., & Karaletsos, T. (2022). Black-box coreset variational inference. *Advances in Neural Information Processing Systems*, *35*, 34175-34187. https://doi.org/10.52202/068431-2477

[16] Kebaili, A., Lapuyade-Lahorgue, J., Vera, P., & Ruan, S. (2024). Discriminative Hamiltonian variational autoencoder for accurate tumor segmentation in data-scarce regimes. *Neurocomputing*, *606*, 128360. https://doi.org/10.1016/j.neucom.2024.128360

[17] Hansen, D. F., Alstrøm, T. S., & Schmidt, M. N. (2023, September). Amortized Variational Peak Fitting For Spectroscopic Data. In *2023 IEEE 33rd International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE. https://doi.org/10.1109/MLSP55844.2023.10285981

[18] Ganguly, A., Jain, S., & Watchareeruetai, U. (2023). Amortized variational inference: A systematic review. *Journal of Artificial Intelligence Research*, *78*, 167-215. https://doi.org/10.1613/jair.1.14258

[19] Lee, S., Park, D., Kong, I., & Kim, H. J. (2024, September). Diffusion prior-based amortized variational inference for noisy inverse problems. In *European Conference on Computer Vision* (pp. 288-304). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-73668-1_17

[20] Liu, M., Grana, D., Mosegaard, K., Sen, M. K., Xu, M., & Mukerji, T. (2026). Bayesian inference for subsurface geophysical inverse problems. *Reviews of Geophysics*, *64*(1), e2025RG000884. https://doi.org/10.1029/2025rg000884

[21] Sun, L., Hu, Z., Todd, M. D., & Zeng, J. (2025). Generative artificial intelligence for Bayesian model updating in digital twins: a review and tutorial: L. Sun et al. *Structural and Multidisciplinary Optimization*, *68*(12), 249. https://doi.org/10.1007/s00158-025-04183-9

[22] Akyol, E. (2026). Semantic rate distortion and posterior design: Compute constraints, multimodality, and strategic inference. *IEEE Journal on Selected Areas in Information Theory*. https://doi.org/10.1109/jsait.2026.3699650

APA Citation

Yang, Z., & Yang, H. (2026). Research on Bayesian Computational Modeling of Intentionality in Artificial Intelligence Art Creation. *Journal of Discovery Core, 1*(2), 110-135. **https://doi.org/10.67541/jdc2611**