Blastoids Pic Breeder

An Interactive Evolutionary Selection System Over a Synthetically Generated Image Corpus

Formal Architecture, Dynamics, and Theoretical Foundations

2026

Abstract

Blastoids Pic Breeder is a browser-based interactive system for shaping a preference field over a corpus of 5,000 synthetically generated images. The user rates images through a CRT-aesthetic targeting interface, progressively constructing a semantic preference vector in CLIP embedding space. A two-pool sampler uses this vector to bias the presentation stream, while a breeding layer generates mutant offspring from the top-rated set. Offspring carry interpolated latent vectors from their parents, are subject to a generation-indexed decay function, and compete directly with base-corpus items under the evolving preference field.

This document provides a formal treatment of the system: the complete state space definition, a unified notation for all primitive objects and update operators, an analysis of the dual-space architecture separating pixel-space rendering from embedding-space evaluation, the preference field dynamics and their anti-collapse mechanisms, the breeding operators and their mutation probability drift, a characterization of dynamical regimes, and an account of the system's structural relationship to the RSVP, TARTAN, and Chain of Memory frameworks. The document concludes with a mathematical summary consolidating the core equations and invariants, a discussion of future theoretical extensions, and a bibliography.

Contents

1. Introduction and Motivation

Most image curation interfaces are fundamentally sorters. They rank, filter, and retrieve, but they do not evolve. The user's role is to identify what already exists; the system does not use that identification to generate anything new. The result is a one-directional relationship between user and corpus that exhausts itself as soon as the corpus is fully ranked.

Blastoids Pic Breeder is designed around a different premise: that rating is not the end of a process but the input to one. Every classification shapes a preference field Φ, which in turn reshapes what the system presents, how it samples, and what it is capable of breeding. The user is not curating a static collection but sculpting a distribution over image space. Over time, that sculpted distribution generates descendants.

The system has three structural layers intended to be mutually reinforcing rather than merely coexistent. The first is a *targeting interface*—a green monochrome CRT display framed as a scanning instrument rather than a gallery. This framing reinterprets the act of rating as signal detection rather than passive preference; the user acquires targets and the field responds. The second is a *preference field* anchored in CLIP embedding space. Unlike systems that use pseudo-random latent vectors, the preference vector Φ in Blastoids corresponds to a direction in a space trained on image-text pairs, so that a rating generalizes beyond the rated item to everything that shares its semantic structure. The third is a *breeding layer* that generates new images by compositing, cropping, and blending parents from the top-rated set. Bred targets carry latent vectors interpolated from their parents and compete in the same preference field as base-corpus images, subject to decay and pruning that creates an ecosystem with turnover rather than accumulation.

2. Formal System Definition

2.1. State Space

The system is a discrete-time dynamical process over a composite state. At any moment in the session, the full system state is:

$$\mathcal{S}_t = \big(B, \ O_t, \ \Phi_t, \ R_t, \ G_t, \ M_t \big), \tag{1}$$

where $B = \{b_1, \ldots, b_N\}$ is the fixed base corpus of images; $O_t = \{o_1, \ldots, o_{m_t}\}$ is the set of offspring at time t; $\Phi_t \in \mathbb{R}^d$ is the preference vector, normalized to unit length whenever any item has been rated; R_t is the rating history, encoding both scalar values and categorical labels for all rated items; $G_t \in \mathbb{N}$ is the breeding generation counter; and M_t captures interface state, including current mode, displayed item, and queue position. The active item set is

$$I_t = B \cup O_t, \tag{2}$$

and every item $i \in I_t$ is associated with a latent embedding $v_i \in \mathbb{R}^d$ satisfying $\|v_i\| = 1$.

2.2. Update Operators

The system evolves through a sequence of update operators. Three are structurally primary.

A *rating update* of item i with value r defines the operator

$$\mathcal{U}_{\mathrm{rate}}(i, r) : \mathcal{S}_t \mapsto \mathcal{S}_{t+1}, \tag{3}$$

which modifies R_t, increments the view count $n_t(i)$, and reconstructs Φ_{t+1} from the full rating history.

A *sampling step* selects the next visible item according to the distribution

$$i_{t+1} \sim \mathcal{P}_t, \quad \mathcal{P}_t(i) \propto \exp\left(\frac{s_t(i)}{T(t)}\right), \tag{4}$$

and updates the displayed item in M_t.

A *breeding event* defines

$$\mathcal{U}_{\mathrm{breed}} : \mathcal{S}_t \mapsto \mathcal{S}_{t+1}, \tag{5}$$

which increments G_t and augments O_t with newly generated offspring. The full session trajectory is therefore a stochastic process

$$\mathcal{S}_{t+1} = \mathcal{U}_t(\mathcal{S}_t), \tag{6}$$

where $\mathcal{U}_t$ is drawn from the set of operators according to user input and internal scheduling. The observable behavior of the system is the trajectory of $\mathcal{S}_t$ through its state space under these updates, shaped jointly by the user's rating decisions and the system's sampling dynamics.

3. Notation and State Variables

3.1. Rating and Observation Maps

The following maps are defined over I_t and constitute the rating component R_t of the state:

$$
\begin{aligned}
r_t &: I_t \to \mathbb{R}, & &\text{scalar rating value,} & (7)\\
\ell_t &: I_t \to \{\texttt{awesome}, \texttt{ok}, \texttt{horrible}, \varnothing\}, & &\text{categorical label,} & (8)\\
n_t &: I_t \to \mathbb{N}, & &\text{view count,} & (9)\\
g_t &: I_t \to \mathbb{N} \cup \{0\}, & &\text{generation index.} & (10)
\end{aligned}
$$

For base items, $g_t(i) = 0$. For offspring, $g_t(i)$ is the generation at which the item was created. The categorical label ℓ_t is stored alongside the scalar r_t to decouple bookkeeping from precise float equality; future changes to rating scalars do not corrupt label-based statistics.

3.2. Preference Vector

Definition 3.1 (Preference vector)**.** The preference vector $\Phi_t \in \mathbb{R}^d$ is the normalized weighted sum

$$\Phi_t = \frac{\sum_{i \in I_t} w_i\, v_i}{\left\| \sum_{i \in I_t} w_i\, v_i \right\|}, \tag{11}$$

where the weight w_i is nonzero only for rated items and satisfies

$$w_i = \mathrm{sgn}(r_t(i)) \cdot \left(1.5 + 0.12\,|r_t(i)|\right). \tag{12}$$

3.3. Score Function

Definition 3.2 (Score function)**.** The score of item i under state $\mathcal{S}_t$ is the composite function

$$s_t(i) = S(i;\, \mathcal{S}_t) = \alpha\,(\Phi_t \cdot v_i) + \beta\, r_t(i) + \nu(n_t(i)) - \rho(n_t(i)) + \sigma(i) - \tau_{\mathrm{age}}(i), \tag{13}$$

where the individual terms are defined in Section 7.

3.4. Temperature

The temperature parameter $T(t)$ controls the concentration of the sampling distribution:

$$T(t) = T_{\min} + (T_{\max} - T_{\min})\, e^{-N_t/\tau}, \tag{14}$$

where N_t is the total number of ratings observed up to time t, and $T_{\min}$, $T_{\max}$, τ are fixed constants. The temperature is monotonically decreasing in N_t and is bounded below by $T_{\min} > 0$.

3.5. Offspring Structure

Each offspring $o \in O_t$ is a tuple

$$o = (v_o,\ p_o,\ m_o,\ g_o), \tag{15}$$

where $v_o \in \mathbb{R}^d$ is the latent vector with $\|v_o\| = 1$; $p_o = (i, j) \in I_t \times I_t$ are parent indices; m_o identifies the mutation operator; and $g_o = G_t$ is the generation index at creation. Parent references are preserved indefinitely and are used for lineage display and ghost-overlay rendering in the interface.

4. Corpus Generation

4.1. Overview

The base corpus B consists of 5,000 images at $256{\times}256$ pixels, generated using Stable Diffusion v1.5 via a ComfyUI workflow. Images are named `blastoids_00001_.png` through `blastoids_05000_.png` with five-digit zero-padded indices. They are distributed across two subdirectories—`images/a/` for indices 1–2500 and `images/b/` for 2501–5000—to accommodate repository size constraints, with a 140×140 thumbnail mirror under `images/thumbnails/`.

4.2. ComfyUI Workflow

The generation pipeline is a standard ComfyUI graph. The node topology is:

```
Node topology:
CheckpointLoaderSimple (id:4) → KSampler (id:3)
EmptyLatentImage (id:5) → KSampler (id:3)
CLIPTextEncode positive (id:6) → KSampler (id:3)
CLIPTextEncode negative (id:7) → KSampler (id:3)
KSampler (id:3) → VAEDecode (id:8) → SaveImage (id:9)
```

4.2.1. Model and Sampler Parameters

Parameter	Value
Checkpoint	`v1-5-pruned-emaonly.ckpt`
Image dimensions	512×512 (latent), decoded to 512×512
Batch size	1
Steps	20
CFG scale	8
Sampler	Euler
Scheduler	Normal
Denoise	1.0
Seed	Randomized per image
Output prefix	`blastoids`

Table 1: KSampler and generation parameters from the ComfyUI workflow.

4.2.2. Positive Conditioning

The positive prompt establishes the visual language of the corpus—an intersection of retro arcade aesthetics, abstract geometry, and semantic field visualization:

Listing 1: Positive CLIPTextEncode prompt (node id:6)

```
1 retro 3D bubble shooter game scene, floating glowing
    spheres arranged
2 in layered depth, player viewpoint aiming reticle at
    center, spheres
```

```
 3  labeled with abstract symbols, minimalist vector wireframe
        environment,
 4  black background, neon green phosphor glow, CRT monitor
        aesthetic,
 5  scanlines, slight bloom, high contrast, geometric clarity,
        early 1990s
 6  arcade style, desaturated except green, subtle perspective
        grid, smooth
 7  spherical shading, no textures, no clutter, precise
        composition
 8
 9  each sphere represents a semantic node, varying size and
        radius, slight
10  motion trails indicating trajectories, field-like
        arrangement instead of
11  fixed rows, emergent clustering, minimal HUD, abstract
        mathematical
12  symbols on spheres, clean spatial separation, no cartoon
        elements, no
13  characters, no background scenery
14
15  wireframe asteroids drifting through neon-green vector
        field, swirling
16  flows around them bending light, minimalistic but high
        detail, crisp CRT
17  glow, dynamic and energetic, glowing green wireframe,
        retro terminal
18  aesthetic, force-field appearance, intricate symmetry
```

The coherence of this prompt is intentional. CLIP embeddings of the corpus cluster around related visual concepts rather than distributing uniformly across the manifold of natural images. This internal structure is what allows the preference field to develop genuine directionality: the dot product $\Phi_t \cdot v_i$ becomes a meaningful discriminator rather than noise.

4.2.3. Negative Conditioning

Listing 2: Negative CLIPTextEncode prompt (node id:7)

```
 1  cartoon bubbles, bright multicolor palette, characters,
        fantasy
 2  environment, UI clutter, text overlays, logos, realistic
        textures,
 3  photorealism, explosions, weapons, complex backgrounds,
        text,
 4  watermark, ugly, amateur, distorted, garish, abstract
```

At 5,000 images with randomized seeds, the corpus samples broadly from the conditional distribution defined by the prompt pair. The resulting distribution is internally varied—images cluster around wireframe field visualizations, abstract symbol

arrays, targeting reticles, and topographic flow patterns—while remaining coherent as a whole. This internal variety is the precondition for meaningful preference differentiation.

5. Dual-Space Architecture

5.1. Two Representational Domains

The system operates simultaneously in two distinct but coupled spaces. Let $\mathcal{X}$ denote the pixel space of rendered images, and let $\mathcal{E} = \mathbb{R}^d$ denote the embedding space induced by CLIP. Each item $i \in I_t$ has a representation in both:

$$i \mapsto (x_i, v_i), \qquad x_i \in \mathcal{X}, \quad v_i \in \mathcal{E}. \tag{16}$$

Pixel space governs what the user sees. Embedding space governs how the system evaluates and relates items. The user interacts exclusively with $\mathcal{X}$, but their actions reshape the direction of Φ_t in $\mathcal{E}$, which in turn governs future observations in $\mathcal{X}$.

5.2. Operator Separation

Mutation and evaluation operate in different spaces, and distinguishing them is important for understanding what the breeding layer actually does.

The crop operator acts in $\mathcal{X}$: it selects a region of a parent image and rescales it to fill the frame. The color shift operator also acts in $\mathcal{X}$: it applies a global intensity transformation. Both modify visual presentation without directly defining semantic relationships. The rendering effects—phosphor filtering, scanlines, crossfade—are similarly confined to $\mathcal{X}$ and are purely presentational.

By contrast, preference accumulation, alignment scoring, and offspring vector construction all operate in $\mathcal{E}$. The preference vector Φ_t is a direction in embedding space, the score term $\Phi_t \cdot v_i$ is a cosine similarity in $\mathcal{E}$, and offspring latent vectors are constructed as normalized interpolations within $\mathcal{E}$.

Blend is the one operator that explicitly couples the two spaces. It produces a pixel-space composite

$$x_o = (1 - \alpha)\, x_A + \alpha\, x_B \tag{17}$$

while simultaneously constructing an embedding-space interpolation

$$v_o = \text{normalize}\big((1 - t)\, v_A + t\, v_B + \eta\big), \qquad \eta \sim \mathcal{N}(0,\, \sigma^2 I). \tag{18}$$

This consistency between visual mixture and semantic interpolation is what makes blend the most semantically coherent of the three mutation operators.

5.3. The Discrete-to-Field Interpretation

The preference vector Φ_t does not select a single item from the corpus; it defines a direction in the continuous space $\mathcal{E}$. The corpus I_t is a finite sample of the underlying manifold $\mathcal{M} \subset \mathcal{E}$, and Φ_t induces a scalar evaluation over that sample:

$$f_t : I_t \to \mathbb{R}, \qquad f_t(i) = \Phi_t \cdot v_i. \tag{19}$$

This scalar function is the discrete trace of a continuous preference field defined over $\mathcal{M}$. In this sense, the term "preference field" is not merely metaphorical: the finite score distribution over I_t is an evaluation of a directional bias over an underlying continuous manifold, projected onto the sample points available to the system.

Remark 5.1. In pseudo-latent mode, $\mathcal{E}$ is arbitrary and the dot product $\Phi_t \cdot v_i$ has no correspondence with human-perceived similarity. In CLIP mode, $\mathcal{E}$ approximates a semantic manifold learned from image-text pairs, and the field f_t becomes semantically grounded. This is the single most important architectural distinction between the two operating modes.

6. CLIP Embedding Architecture

6.1. Precomputed Sidecar

Rather than loading a vision model in the browser, the system uses a precomputed embedding sidecar. The script `generate_embeddings.py` walks the corpus in index order, embeds each image using OpenCLIP ViT-B/32, and writes a JSON file with the structure:

Listing 3: Sidecar JSON structure

```
{
  "model":  "ViT-B-32/openai",
  "dim":    512,
  "count":  5000,
  "embeddings": {
    "0": [0.023, -0.041, ...],   // L2-normalized, 512
        floats
    "1": [...],
    ...
  }
}
```

At 512 dimensions with 5,000 images, the uncompressed sidecar is approximately 10 MB. An optional PCA reduction flag produces a smaller file; 64 dimensions retains roughly 85% of explained variance for photographic content.

6.2. Runtime Loading and Dimension Transitions

The sidecar is fetched asynchronously at startup. The loading sequence is ordered to preserve consistency. The system first hydrates persisted session state from

`localStorage`, then fetches and parses the sidecar. If the loaded dimension d' differs from the dimension d used in the saved session, all offspring are cleared and the breeding generation counter is reset, because offspring latent vectors from a session in $\mathbb{R}^d$ are geometrically incoherent in $\mathbb{R}^{d'}$. Base-image ratings and labels are preserved across the transition. The preference vector Φ_t is resized to $\mathbb{R}^{d'}$ and rebuilt from the surviving rating history using the new embeddings. If there are no saved ratings, Φ_t is initialized to zero and the system enters the weak-field regime.

Invariant 6.1 (Dimensional consistency)**.** At all times, all active latent vectors satisfy $v_i \in \mathbb{R}^d$ for a single shared d. No item with a mismatched dimension may remain in I_t after a dimension change.

7. Preference Field Dynamics

7.1. Preference Accumulation

The preference vector is rebuilt from the full rating history after every rating event:

$$\Phi_t = \frac{\tilde{\Phi}_t}{\|\tilde{\Phi}_t\|}, \qquad \tilde{\Phi}_t = \sum_{i \in I_t} w_i \, v_i, \tag{20}$$

with weights

$$w_i = \begin{cases} \mathrm{sgn}(r_t(i)) \cdot (1.5 + 0.12 \, |r_t(i)|) & \text{if } r_t(i) \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

The rating scalars are: $r = 1.0$ for AWESOME, $r = 0.4$ for OK, and $r = -1.0$ for HORRIBLE. The OK value is set above zero rather than at the neutral point because a positive but weak rating still carries directional information—it establishes that the item was observed and found acceptable, contributing mild positive evidence about the field direction.

7.2. Item Scoring

Each item receives a composite score incorporating six terms:

$$s_t(i) = \underbrace{\alpha \, (\Phi_t \cdot v_i)}_{\text{alignment}} + \underbrace{\beta \, r_t(i)}_{\text{explicit}} + \underbrace{\nu(n_t(i))}_{\text{novelty}} - \underbrace{\rho(n_t(i))}_{\text{repetition}} + \underbrace{\sigma(i)}_{\text{offspring}} - \underbrace{\tau_{\text{age}}(i)}_{\text{decay}}. \tag{22}$$

The alignment coefficient $\alpha = 0.55$ is the most consequential parameter. In CLIP space, a well-aligned dot product can reach values near 0.8–0.9 after only a few strong ratings. At $\alpha = 1.0$, alignment would dominate all other terms and drive rapid convergence. At $\alpha = 0.55$, the maximum alignment contribution is bounded below unity, keeping alignment competitive with but not dominant over novelty and explicit rating.

The explicit rating term $\beta \, r_t(i)$ with $\beta = 0.55$ provides a direct, time-stable reward that does not decay. The novelty term $\nu(n_t(i)) = \max(0, 0.22 - 0.025 \, n_t(i))$

grants newly seen items an initial advantage, with unseen items receiving a bonus of 0.38. The repetition penalty $\rho(n_t(i)) = \min(0.45,\, 0.02\, n_t(i))$ grows with view count, bounded above to prevent permanent exclusion.

For offspring, the term $\sigma(i) = 0.15$ applies when the item has not yet been seen, providing an initial exploration incentive. The offspring decay term is

$$\tau_{\text{age}}(i) = \max\bigl(0,\, (G_t - g_t(i)) \cdot k_{\text{decay}}\bigr), \quad k_{\text{decay}} = 0.04, \tag{23}$$

which introduces generation-indexed selection pressure against old unrated offspring. An offspring marked AWESOME accumulates a protection of $\beta \cdot r = 0.55$ against decay, providing approximately 14 generations of competitive persistence before the age penalty exceeds the explicit signal.

7.3. Two-Pool Sampling

The sampling distribution mixes over two pools with a fixed ratio:

$$\mathcal{P}_t = (1 - \rho_{\text{mix}})\, \mathcal{P}_t^{\text{base}} + \rho_{\text{mix}}\, \mathcal{P}_t^{\text{offspring}}, \qquad \rho_{\text{mix}} = 0.22. \tag{24}$$

This decomposition ensures that offspring visibility is not determined solely by score dominance. Even newly generated offspring with no ratings will appear in the stream at a stable rate of approximately one in five draws, provided the pool is nonempty. The field shapes *which* offspring appear; the mixing ratio ensures that *whether* they appear is a controlled parameter rather than an emergent accident.

8. Convergence, Exploration, and Non-Collapse Mechanisms

8.1. Objective Structure

The system does not optimize toward a fixed maximizer of the score function. Its design objective is to maintain a distribution over items that is biased by user preference while retaining nonzero entropy. This is a fundamentally different objective from standard recommender systems, which typically minimize prediction error or maximize engagement with respect to a static user model. Blastoids Pic Breeder does not minimize anything; it sustains a process.

The sampling distribution takes the form

$$\mathcal{P}_t(i) \propto \exp\!\left(\frac{s_t(i)}{T(t)}\right), \tag{25}$$

with $T(t) > 0$ for all t, so the distribution never collapses to a point mass.

8.2. Alignment Bounding

In embedding space, cosine alignment $\Phi_t \cdot v_i$ can produce high values even for moderate preference vectors. If weighted at unity, the alignment term dominates all others

after a small number of strong ratings. The coefficient $\alpha = 0.55$ ensures that for all items and all states,

$$|\alpha \left(\Phi_t \cdot v_i \right)| \leq \alpha < 1, \tag{26}$$

so that the alignment contribution is always bounded below the summed influence of explicit ratings, novelty, and penalties. This keeps the score function responsive to multiple sources of evidence rather than collapsing to a single criterion.

8.3. Temperature Floor

The temperature function satisfies $T(t) \geq T_{\min} = 0.42 > 0$ for all t. As $t \to \infty$, the distribution converges to a softened rather than a deterministic argmax. The floor is set at 0.42 rather than a lower value because in CLIP space the alignment signal is strong enough that lower floors produce near-greedy selection, eliminating the exploratory character of the system even in converged sessions. At $T_{\min} = 0.42$, approximately 15% more probability mass rests on non-top items at full convergence compared to what a floor of 0.32 would produce.

8.4. Novelty, Repetition, and Transient Advantage

The novelty and repetition terms create a transient advantage for unseen items without permanently altering their long-term visibility. Novelty decays to zero after sufficient exposure; the repetition penalty is bounded above. An item that was frequently seen early in a session does not become permanently disadvantaged, and an item that was never shown retains its novelty bonus until first displayed. This ensures that early exposure patterns do not lock in long-term bias independent of the preference field.

8.5. Offspring Decay and Protected Persistence

Offspring are subject to a temporal gradient that reduces the competitiveness of older unrated items relative to fresh ones. For any unrated offspring i,

$$\lim_{G_t \to \infty} s_t(i) \to -\infty \quad \text{relative to fresh items at the same alignment.} \tag{27}$$

The pool therefore does not sediment: old low-relevance offspring are naturally displaced by new candidates generated under the current field.

Rated offspring are protected from this pressure. An offspring with explicit rating $r > 0$ satisfies

$$\beta\, r > (G_t - g_t(i)) \cdot k_{\text{decay}} \tag{28}$$

for a finite number of generations, establishing a protection window proportional to the rating strength. Beyond this window, even positively rated items lose prominence unless they remain aligned with the current Φ_t. This means that sustained relevance requires both a history of positive evaluation and continued semantic alignment—not simply that the user once rated the item favorably.

8.6. Non-Collapse Condition

Proposition 8.1. For all t, the sampling distribution $\mathcal{P}_t$ satisfies $\mathcal{P}_t(i) > 0$ for all $i \in I_t$.

Argument. The score function $s_t(i)$ is bounded and real-valued for all i, by the boundedness of alignment (above), the finiteness of view counts, and the bounded-above nature of decay penalties. With $T(t) \geq T_{\min} > 0$, the exponential $\exp(s_t(i)/T(t))$ is strictly positive and finite for all i, so the softmax is strictly positive everywhere. $\square$

Combined with the two-pool mixing structure, this implies that no item in I_t becomes permanently unreachable, and the sampling support over both base corpus and offspring pool remains non-degenerate throughout the session.

9. Breeding Operators and Evolutionary Dynamics

9.1. Breeding as a Stochastic Operator

A breeding event defines a stochastic operator

$$\mathcal{B}_t : \mathcal{S}_t \mapsto \mathcal{S}_{t+1}, \tag{29}$$

which increments G_t, performs optional pruning of the offspring pool, and generates up to $k_{\text{breed}} = 12$ new offspring from weighted parent pairs drawn from the current top-ranked set.

Parents are selected with probability proportional to their score under $\mathcal{S}_t$, ensuring that breeding pressure tracks the current preference field rather than a fixed ranking. Two distinct parents $i, j \in I_t$ are sampled per offspring, and the latent vector of the child is constructed as a normalized interpolation with added noise:

$$v_o = \text{normalize}\big((1 - t)\, v_i + t\, v_j + \eta\big), \quad t \sim \text{Uniform}(0.3, 0.7), \quad \eta \sim \mathcal{N}(0, \sigma^2 I). \tag{30}$$

The noise η ensures that repeated applications of the breeding operator to the same parent pair do not produce identical offspring. The standard deviation σ varies by mutation type: it is smallest for blend (0.05), largest for color shift (0.12), reflecting the fact that pixel-space color transformations contribute less semantic variation than they do visual variation.

9.2. Mutation Operators

Three mutation operators are available. The crop operator selects a soft-bounded region of parent A defined by a quadrant flag $(q_x, q_y) \in \{0, 1\}^2$ and proportional dimensions $(c_w, c_h) \in [0.45, 0.65]^2$ sampled at breed time. The region coordinates are:

$$s_x = q_x\, (W - c_w W), \quad s_y = q_y\, (H - c_h H). \tag{31}$$

Crucially, these parameters are encoded into the offspring's URL fragment at creation time—`#crop-qx-qy-cwPct-chPct`—making the rendered result deterministic across sessions without requiring additional storage. This encoding pattern is general: any operator whose parameters can be serialized as a string can be made persistent and reproducible in the same way.

The blend operator composites parent A with parent B at a random opacity $\alpha_{px} \in [0.30, 0.70]$ in pixel space while constructing the latent vector as an interpolation between the two parent embeddings at a matching ratio. Blend is the semantically deepest operator because it produces a child that genuinely interpolates between two concept regions in $\mathcal{E}$, rather than selecting a subregion of a single parent.

The color shift operator applies a screen-mode tint overlay—warm or cool—over parent A. It is the weakest operator semantically, introducing mostly cosmetic variation with minimal embedding-space consequence. It is retained primarily because the higher latent noise in its interpolation path occasionally produces useful outliers.

9.3. Mutation Probability Drift

The mutation probabilities shift with generation G_t:

$$p_{\text{crop}}(G_t) = \max(0.20, \ 0.45 - G_t \cdot 0.015), \tag{32}$$
$$p_{\text{blend}}(G_t) = \min(0.55, \ 0.35 + G_t \cdot 0.015), \tag{33}$$
$$p_{\text{shift}}(G_t) = 1 - p_{\text{crop}} - p_{\text{blend}}. \tag{34}$$

At generation 0, crop dominates at 45% and blend is at 35%. The crossover occurs near generation 7, after which blend dominates. This drift induces a developmental arc without explicit phase switching: early sessions produce mostly structural extractions from individual images, while later sessions synthesize across concept regions. The transition is continuous and imperceptible as a discrete event but visible in the character of the offspring population over time.

9.4. Pool Pruning

When the offspring pool approaches its capacity $O_{\max} = 120$, the system prunes low-scoring unrated offspring to make room for fresh ones:

Listing 4: Pruning on breed

```
const unrated = offspringItems
  .filter(it => it.label === null)
  .sort((a, b) => itemScore(a) - itemScore(b));

const toPrune = Math.min(
  Math.ceil(OFFSPRING_PER_BREED * 0.5),
  unrated.length
);
```

Pruning is sorted by current score, preferentially removing items that are both old (age penalty) and poorly aligned with the current field. Rated items are never pruned regardless of their score, preserving all user-validated structure.

10.   Dynamical Regimes and Qualitative Behavior

10.1.   Regime Structure

The system exhibits distinct qualitative regimes as a function of accumulated ratings and breeding activity. These regimes are not explicitly encoded as states but emerge from the interaction of preference accumulation, temperature decay, and offspring dynamics. Let N_t denote the total number of ratings at time t.

10.2.   Weak-Field Regime

When N_t is small, the preference vector satisfies $\|\Phi_t\| \approx 0$ or is dominated by a small number of contributions. In this regime, the alignment term $\Phi_t \cdot v_i$ does not meaningfully differentiate items, and the score function is dominated by novelty and low view counts. The sampling distribution is close to uniform over B, and the field concept readout is unstable—small changes to Φ_t from a single new rating can change the identity of the nearest neighbor. The system behaves as an exploratory sampler over the corpus rather than as a directed process.

10.3.   Intermediate Regime: Emergent Directionality

As N_t grows, the cumulative weight of rated items produces a preference vector with non-negligible magnitude. The alignment term begins to structure the score distribution, and clusters in embedding space become visible through repeated exposure. The field concept readout stabilizes under the hysteresis constraint. The slideshow exhibits persistent aesthetic tendencies without exhibiting rigid convergence, because temperature and novelty maintain substantial variability.

Breeding becomes meaningful in this regime. Offspring generated from top-ranked items are more likely to align with Φ_t, and thus to reappear in the stream after their initial appearance. The offspring pool transitions from a collection of exploratory noise to a collection of structured variation anchored in the current field.

10.4.   Late Regime: Structured Exploration

In later stages, with large N_t, temperature approaches $T_{\min}$ and Φ_t encodes a well-defined direction in $\mathcal{E}$. The alignment term is strong, but bounded by α, preventing complete dominance. The resulting dynamics are not those of strict convergence but of structured exploration: the system continues to sample from a region of $\mathcal{E}$ aligned with Φ_t, while retaining sufficient entropy to probe variations within that region. Offspring turnover, driven by generation decay and pruning, becomes the primary source of local novelty.

The mutation probability drift also reaches its asymptotic state in this regime, with blend probabilities near their ceiling. The offspring population therefore shifts from

structural extraction to semantic synthesis, generating candidates that interpolate between established concept regions rather than zooming into existing ones.

10.5. Temporal Continuity and Presentation

The crossfade mechanism introduces continuity in the presentation layer. Each transition between images in the slideshow performs a 380 ms crossfade: the outgoing image decays to a maximum opacity of 0.72 while the incoming image fades in. The asymmetry is deliberate—the past is lighter than the present, rendering as residue rather than as equal presence. The state transition remains discrete, but the rendered sequence approximates a continuous trajectory in pixel space, reinforcing the interpretation of the system as a field evolving over time rather than a sequence of independent samples.

When a bred target appears for the first time, the interface momentarily displays a FIELD SYNTHESIS EVENT label. This occurs only on the item's first exposure and only once the fade is nearly complete, so it reads as an arrival rather than a status notification. It marks bred targets as outputs of the field rather than as ordinary corpus items.

11. Definitions and Invariants

The following definitions and invariants constitute the structural constraints of the system. They hold under all update operators and define the boundaries within which the dynamics operate.

Definition 11.1 (Rated item)**.** An item $i \in I_t$ is rated if $r_t(i) \neq 0$. It is unrated if $r_t(i) = 0$ and $\ell_t(i) = \varnothing$.

Definition 11.2 (Preference field)**.** The preference field induced by Φ_t over I_t is the scalar function $f_t(i) = \Phi_t \cdot v_i$.

Definition 11.3 (Generation consistency)**.** The generation index $g_t(i)$ of any offspring satisfies $g_t(i) \leq G_t$. The counter G_t is monotonically non-decreasing: $G_{t+1} \geq G_t$ for all t.

Invariant 11.1 (Dimensional consistency)**.** All latent vectors in the active system share a common dimension: $v_i \in \mathbb{R}^d$ for all $i \in I_t$ and a single d. If the embedding dimension changes, all offspring with incompatible vectors are removed from I_t.

Invariant 11.2 (Normalization)**.** All latent vectors are unit-normalized: $\|v_i\| = 1$ for all $i \in I_t$. When at least one item is rated, $\|\Phi_t\| = 1$.

Invariant 11.3 (Persistence of rated offspring)**.** If $i \in O_t$ and $\ell_t(i) \neq \varnothing$, then $i \in O_{t'}$ for all $t' \geq t$. Rated offspring are never removed by pruning.

Invariant 11.4 (Bounded population)**.** $|O_t| \leq O_{\max}$ for all t. Consequently, $|I_t| = |B| + |O_t| \leq N + O_{\max}$, and the total item set is bounded regardless of session length.

Invariant 11.5 (Strictly positive sampling)**.** For all t and all $i \in I_t$, $\mathcal{P}_t(i) > 0$, provided $T(t) > 0$ and the score function is finite-valued. Since $T(t) \geq T_{\min} > 0$ and $s_t(i) \in \mathbb{R}$ for all i, this invariant holds unconditionally.

12. Interface, Controls, and Persistence

12.1. Modes

The system operates in two primary modes. Slideshow mode advances automatically at 2.8-second intervals, showing images drawn from the two-pool sampler. Each rating updates Φ_t and advances to the next sample. Curate mode pauses autoplay, allowing the user to navigate manually, browse the top and worst ranked sets and the bred pool, and select any thumbnail to jump to that item for inspection and rating.

12.2. Controls and HUD Readouts

Key	Action
1	AWESOME ($r = 1.0$)
2	OK ($r = 0.4$)
3	HORRIBLE ($r = -1.0$)
$\rightarrow$	Next image
M	Toggle slideshow / curate
B	Breed from top set
T	Open top-ranked set
W	Open worst-ranked set
O	Open bred targets
Space	Pause / resume
H	Hide / show UI

Table 2: Keyboard shortcuts.

The HUD exposes four live readouts. The PREFERENCE Φ row displays $\|\Phi_t\|$, which is zero at session start and grows as ratings accumulate; values above approximately 0.3 indicate a field with directional structure. The TEMPERATURE row shows the current $T(t)$, starting at 1.55 and decaying toward 0.42. The FIELD $\rightarrow$ row displays the result of $\arg\max_{i \in B} (\Phi_t/\|\Phi_t\|) \cdot v_i$, the base-corpus image whose embedding most closely aligns with the current field direction. This probe is stabilized with a hysteresis margin of 0.02 cosine units to prevent flickering under small perturbations to Φ_t; the displayed item changes only when a new best exceeds the previous best by at least that margin. The LINEAGE row identifies the current item as either BASE CORPUS or, for bred targets, lists the mutation type and parent identifiers.

The field concept readout is computed exclusively over base-corpus items. This is a deliberate design choice: the field is described in terms of the original vocabulary rather than in terms of its own synthesized outputs. An offspring does not acquire interpretive standing through existence but only through explicit rating and validated alignment with the preference field.

12.3. Persistence

Session state is serialized to `localStorage` under the key `blastoids_state_v1` after every meaningful state change. The serialized representation includes ratings with both scalar values and categorical labels, the preference vector, the offspring pool with URL fragments encoding mutation parameters, the breeding generation counter, mode, and the current item identity.

On reload, the system first hydrates state, then loads the CLIP sidecar asynchronously. If the sidecar dimension differs from the saved preference vector dimension, the preference vector is discarded and rebuilt from the surviving rating history using the new embeddings. A 16-dimensional pseudo-latent vector is geometrically incoherent in 512-dimensional CLIP space and cannot be carried over. The status line reports this transition explicitly so the user understands that the system has re-grounded its field, not lost its history.

Constant	Default	Interpretation
`ALIGN_WEIGHT` α	0.55	Alignment contribution to scoring
`T_MIN`	0.42	Sampling temperature floor
`T_TAU` τ	180	Ratings to 63% temperature decay
`OFFSPRING_MIX_RATIO` ρ_{mix}	0.22	Fraction of draws from offspring pool
`OFFSPRING_DECAY_RATE` k_{decay}	0.04	Score penalty per generation
`AUTOPLAY_MS`	2800	Milliseconds per image in slideshow

Table 3: Exposed tuning constants with their formal identifiers.

13. Relation to RSVP, TARTAN, and Associated Frameworks

13.1. Preference as Field: RSVP Correspondence

The preference vector Φ_t can be understood as a special case of a scalar-vector organizational field in the sense developed in the RSVP framework. It is not itself an image, nor a discrete object within the corpus, but a directional constraint defined over the continuous embedding space $\mathcal{E}$. It does not passively describe existing structure; it actively reorganizes which structures are stable, visible, and selectable.

In RSVP-style descriptions, fields are not stored quantities but dynamic constraints that shape the availability of local configurations. The preference field Φ_t plays precisely this role: it does not store the user's taste as a list or a ranked table, but defines a direction in a continuous space, and that direction reshapes the entire competitive landscape of items at every sampling step. The continuous gradient $f_t(i) = \Phi_t \cdot v_i$ is the discrete trace of this field over the finite corpus.

The formal correspondence is exact at the level of structure, though the domain differs. RSVP describes a physical or cosmological field over continuous spacetime. Blastoids Pic Breeder instantiates the same organizational logic—a global vector shaping local accessibility—over a discrete sample of a semantic manifold.

13.2. Recursive Generation and TARTAN Structure

The breeding layer exhibits a structural correspondence to trajectory-aware recursive generation. Offspring are not independent samples from a distribution but are constructed through explicit transformations of previously selected items, with lineage information preserved:

$$o = (v_o,\ p_o,\ m_o,\ g_o), \tag{35}$$

where p_o records the identities of both parents and m_o encodes the transformation. Over time, this creates a directed acyclic graph of derivation: new items arise through recombination and perturbation of existing structure, and the space of possible future items is constrained by the history of selections and transformations.

TARTAN's central claim is that trajectory-aware recursive tiling with annotated noise produces structured outputs that preserve and extend historical pattern rather than introducing unanchored variation. The breeding layer enacts a restricted version of this claim in semantic space: blend offspring are interpolations between established positions in $\mathcal{E}$, and crop offspring extend the corpus of visited regions within a single parent's embedding neighborhood. The mutation probability drift ensures that the system moves from pattern extension (crop dominates early) toward pattern combination (blend dominates late), which is consistent with TARTAN's emphasis on recursive compositional structure.

13.3. Selective Persistence and Chain of Memory

The decay and rating mechanisms implement a selective memory structure. Not all items persist equally; persistence depends on continued relevance under the evolving field. Unrated offspring lose influence through the age penalty, eventually sinking below fresh candidates regardless of their initial alignment. Rated offspring are protected in proportion to the strength of the user's commitment, as expressed in $\beta\, r_t(i)$. Base-corpus items are never removed, constituting a permanent background memory of the original vocabulary.

This is analogous to the structure described in Chain of Memory, where identity and relevance are maintained not through storage alone but through continued participation in an ongoing process. An item that was once important does not remain important by virtue of having been encountered. It remains important by remaining compatible with the current constraints of the system—by continuing to align with the direction Φ_t is moving, or by carrying the weight of explicit user endorsement.

The connection is not merely metaphorical. The decay function $\tau_{\text{age}}(i)$ is a formal expression of the principle that relevance must be continuously renewed, and the protection of rated items is a formal expression of the principle that explicit commitment confers a different kind of persistence than passive encounter.

13.4. The System as a Weakly Autopoietic Process

Standard recommendation systems map user data to ranked outputs, optimizing for predicted preference or engagement. The mapping is a function from a user

representation to an item ordering; the user's role is to provide signal, and the system's role is to process it. The corpus is fixed.

Blastoids Pic Breeder differs in a structural sense: the breeding layer means that the corpus is not fixed. The system generates new candidates whose existence depends on prior selections. The set of possible future observations is a function of the user's action history, not only of the original corpus. This makes the system weakly autopoietic in the sense that it produces some of the elements it operates on—not self-producing in the biological sense, but exhibiting the structural property that outputs become inputs at the next generation.

The epistemic status of this production is carefully managed. Offspring do not automatically enter the interpretive vocabulary of the system. The field concept readout is computed exclusively over B, ensuring that the description of the field is always grounded in items the user has directly encountered. An offspring must be rated before it contributes to Φ_t and must accumulate alignment evidence before it becomes a stable presence in the stream. Generation is constrained; it does not run free.

13.5. Unified Characterization

The system instantiates a constrained computational motif that is common across the frameworks mentioned above: coherent global organization emerging from repeated local updates under explicit historical constraint. The three components of this motif in Blastoids Pic Breeder are the preference field Φ_t, which provides global organization; the breeding operators and decay function, which implement local update under historical constraint; and the rating and label structures, which are the explicit historical record that constrains the direction of change.

Blastoids Pic Breeder is not an RSVP cosmology or a TARTAN simulator. It is an instance of the same family of processes, implemented over a discrete semantic corpus, in a browser, for the purpose of interactive aesthetic evolution. The theoretical correspondence is structural rather than literal.

14. Mathematical Summary

14.1. System State

$$\mathcal{S}_t = (B,\ O_t,\ \Phi_t,\ R_t,\ G_t,\ M_t), \qquad I_t = B \cup O_t. \tag{36}$$

Every item $i \in I_t$ satisfies $v_i \in \mathbb{R}^d$, $\|v_i\| = 1$.

14.2. Preference Field

$$\Phi_t = \frac{\tilde{\Phi}_t}{\|\tilde{\Phi}_t\|}, \qquad \tilde{\Phi}_t = \sum_{i \in I_t} w_i\, v_i, \qquad f_t(i) = \Phi_t \cdot v_i. \tag{37}$$

14.3. Score Function

$$s_t(i) = \alpha f_t(i) + \beta r_t(i) + \nu(n_t(i)) - \rho(n_t(i)) + \sigma(i) - \tau_{\text{age}}(i), \tag{38}$$

where $\nu(n) = \max(0,\, 0.22 - 0.025n)$ with $\nu(0) = 0.38$; $\rho(n) = \min(0.45,\, 0.02n)$; $\sigma(i) = 0.15$ for unseen offspring, 0 otherwise; and $\tau_{\text{age}}(i) = \max(0,\, (G_t - g_t(i))\, k_{\text{decay}})$.

14.4. Sampling

$$\mathcal{P}_t = (1 - \rho_{\text{mix}})\, \mathcal{P}_t^{\text{base}} + \rho_{\text{mix}}\, \mathcal{P}_t^{\text{off}}, \qquad \mathcal{P}_t^{\text{base/off}}(i) \propto e^{s_t(i)/T(t)}. \tag{39}$$

$$T(t) = T_{\min} + (T_{\max} - T_{\min})\, e^{-N_t/\tau}. \tag{40}$$

14.5. Offspring Generation

$$v_o = \text{normalize}\big((1 - t)\, v_i + t\, v_j + \eta\big), \qquad t \in [0.3, 0.7], \quad \|\eta\| \ll 1. \tag{41}$$

$$g_o = G_t + 1, \qquad p_o = (i, j), \qquad m_o \sim \pi(G_t). \tag{42}$$

14.6. Mutation Drift

$$p_{\text{crop}}(G) = \max(0.20,\, 0.45 - 0.015\, G), \qquad p_{\text{blend}}(G) = \min(0.55,\, 0.35 + 0.015\, G). \tag{43}$$

14.7. Field Concept Readout

$$i^* = \underset{i \in B}{\arg\max}\ \frac{\Phi_t}{\|\Phi_t\|} \cdot v_i. \tag{44}$$

Updated only when $(\Phi_t/\|\Phi_t\|) \cdot v_{i_{\text{new}}^*} > (\Phi_t/\|\Phi_t\|) \cdot v_{i_{\text{old}}^*} + \delta_{\text{hyst}}$, with $\delta_{\text{hyst}} = 0.02$.

14.8. Invariants Summary

$$\|v_i\| = \|\Phi_t\| = 1, \qquad |O_t| \le O_{\max}, \qquad T(t) \ge T_{\min} > 0, \qquad \mathcal{P}_t(i) > 0\ \forall i \in I_t. \tag{45}$$

$$\ell_t(i) \ne \varnothing \implies i \in O_{t'}\ \forall t' \ge t. \qquad G_t \text{ is monotone non-decreasing.} \tag{46}$$

15. Future Theoretical Extensions

15.1. Continuous Field Approximation

The system currently evaluates Φ_t only over the discrete set I_t. Since all items are embedded in $\mathbb{R}^d$, the corpus is a finite sample of an underlying continuous manifold $\mathcal{M} \subset \mathbb{R}^d$. A natural extension is to define the preference field over the full manifold

$$f_t(v) = \Phi_t \cdot v, \qquad v \in \mathcal{M}, \tag{47}$$

and to generate candidates by decoding points sampled from high-field regions of $\mathcal{M}$ directly, rather than relying solely on recombination of existing items. This

would transform the system from a selector-and-breeder over a fixed corpus into a continuous field explorer constrained to a learned semantic manifold.

15.2. Entropy-Regularized Dynamics

The current exploration-exploitation balance is maintained through temperature and alignment bounding. A more principled formulation would introduce explicit entropy regularization. Let $H(\mathcal{P}_t) = -\sum_i \mathcal{P}_t(i) \log \mathcal{P}_t(i)$ be the entropy of the sampling distribution. One could define an objective

$$\mathcal{L}_t = \mathbb{E}_{i \sim \mathcal{P}_t}[s_t(i)] + \lambda \, H(\mathcal{P}_t), \tag{48}$$

in which temperature becomes a dual variable for the entropy constraint. This would allow adaptive temperature schedules based on observed distributional diversity rather than fixed exponential decay.

15.3. Offspring Legitimacy and Semantic Promotion

Currently, the field concept readout is computed exclusively over B. A possible extension would define a promotion criterion under which offspring can enter the interpretive basis. One natural criterion is that an offspring o satisfies $r_t(o) > 0$ and $n_t(o) \geq n_{\min}$ for some minimum exposure threshold $n_{\min}$. Promoted offspring would then be eligible as anchors in field interpretation, allowing the system's vocabulary to expand over time through validated synthetic generation.

15.4. Adaptive Mutation Kernels

The mutation probability drift is currently a deterministic function of G_t. A more responsive formulation would condition mutation selection on the structure of Φ_t or on local density in $\mathcal{E}$. For example, high alignment concentration (low entropy in f_t) could increase the probability of blend operations to encourage cross-cluster synthesis, while low alignment concentration could favor crop operations to extract stronger local features. This would make the system's generative strategy responsive to the dynamical regime it currently occupies.

15.5. Manifold Geometry and Non-Euclidean Alignment

Cosine similarity assumes that $\mathcal{E}$ is locally isotropic and that angular distance captures semantic proximity uniformly across the manifold. More sophisticated models could incorporate local curvature or density, evaluating alignment as

$$\langle \Phi_t, v \rangle_g = \Phi_t^\top \, g(v) \, v, \tag{49}$$

for a metric tensor $g(v)$ estimated from the corpus. While computationally expensive, this would allow the system to account for nonuniform structure in $\mathcal{E}$ and produce more geometrically accurate preference evaluations.

15.6. Categorical and Multi-Axis Decomposition

The current preference vector Φ_t is a single direction in $\mathbb{R}^d$. A natural extension decomposes it into multiple components corresponding to distinct semantic axes:

$$\Phi_t = \sum_k \Phi_t^{(k)}. \tag{50}$$

Each component could correspond to a different dimension of aesthetic preference—composition, color, subject matter, abstraction level—allowing the system to represent more complex and structured user intentions. Sampling and breeding could then operate conditionally on these components, producing offspring that inherit specific aesthetic properties from specific parents.

15.7. Process-Level Generalization

At the highest level of abstraction, the system is an instance of a family of processes characterized by three structural properties: a global field that organizes local behavior; a recursive generative mechanism that produces new candidates from existing structure; and a selective persistence rule that determines which elements remain active over time. The formal study of this family, independently of any particular application domain, is a natural direction for future theoretical development. Blastoids Pic Breeder provides a minimal, fully specified, computationally realized instance of this family, making it a useful test case for any general theoretical treatment.

16. Conclusion

Blastoids Pic Breeder is not primarily a viewer or a ranker. It is a preference field shaping interface over a synthetically generated semantic corpus. The user constructs an aesthetic direction in embedding space through repeated acts of classification; the system uses that direction to reshape what it presents, what it breeds, and what it allows to persist.

The formal treatment in this document establishes that the system's behavior is not a collection of individually motivated heuristics but a coherent set of consequences derived from a defined state space, a scoring function with bounded components, a two-pool sampler with a temperature floor, and a breeding operator with generation-indexed selection pressure. The non-collapse properties follow from the structure of these components. The developmental arc—from exploratory sampling through emergent directionality to structured synthesis—emerges from the interaction of temperature decay, mutation drift, and offspring competition rather than from explicit phase switching.

The connection to RSVP, TARTAN, and Chain of Memory is structural rather than literal. The system enacts a restricted instance of the same family of processes: global field organization, recursive constrained generation, and selective historically-sensitive persistence. The theoretical significance of this instance lies not in its scale or power but in its formal clarity: every component is explicitly defined, every invariant is checkable, and every behavioral regime can be traced to specific parameter

ranges. It is a minimal system in which perception, evaluation, and generation are genuinely coupled.

Generated corpus: 5,000 images · SD v1.5 · ComfyUI · Euler sampler · 20 steps · CFG 8
Embedding model: CLIP ViT-B/32 (OpenAI) via OpenCLIP · Embedding dimension: 512
Interface: Single-file HTML/JS/Canvas · localStorage persistence · Mobile touch support

References

[1] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.

[2] Mehdi Cherti, Romain Beaumont, Ross Wightman, Mitchell Wortsman, Gabriel Ilharco, Cade Gordon, Christoph Schuhmann, Ludwig Schmidt, and Jenia Jitsev. Reproducible scaling laws for contrastive language-image learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[3] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[4] Jimmy Secretan, Nicholas Beato, David B. D'Ambrosio, Adelein Rodriguez, Adam Campbell, and Kenneth O. Stanley. Picbreeder: Evolving pictures collaboratively online. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2011.

[5] Kenneth O. Stanley and Risto Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary Computation*, 10(2):99–127, 2002.

[6] Diederik P. Kingma and Max Welling. Auto-encoding variational Bayes. In *International Conference on Learning Representations (ICLR)*, 2014.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[8] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2nd edition, 2006.

[9] Shun-ichi Amari. *Information Geometry and Its Applications*. Springer, 2016.

[10] Karl Friston. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, 2010.

[11] Christof Wetterich. Exact evolution equation for the effective potential. *Physics Letters B*, 301(1):90–94, 1993.

[12] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[13] Humberto R. Maturana and Francisco J. Varela. *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel, 1980.

[14] Douglas Hofstadter. *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books, 1979.