The Art of Art

A Strategic Treatise on Human Creativity and Generative AI

Justin A. Fritz
The Canonical Art LLC, Fort Collins, CO, USA
justin@thecanonicalart.com

June 2026

"The modern industrial revolution is similarly bound to devalue the human brain, at least in its simpler and more routine decisions."

— Norbert Wiener, *The Human Use of Human Beings*, 1950

Abstract

Sun Tzu's *Art of War* is not about war. It is a strategic framework for any domain where resources are limited, adversaries are adaptive, and knowing the terrain determines survival. *The Art of Art* is its contemporary equivalent: a strategic framework for human creativity in a terrain reshaped by generative AI. The conflation of *AI-generated* with *AI-assisted* is the first terrain error – it misidentifies the adversary and misallocates the defense.

This paper develops a dynamical systems framework to make the distinction precise, grounded in four hand-drawn conceptual diagrams created in 2022–2023 – prior to the public availability of large-scale generative AI systems – and in the author's experience as a hardware engineer whose AI-assisted technical papers have been challenged on provenance grounds.

We propose four interlocking claims: (I) human creative and intellectual cognition is a knowledge state machine $K_{t+1} = f(K_t, \theta_t, \epsilon_t)$ with a trajectory that constitutes authorship; (II) the economic value of intellectual work derives from the *rate of change* of the knowledge field it induces ($V \propto d\Phi/dt$), not from static output quality – the same mechanism that allows AI to commoditize generic art does not reach expert-directed work at the distribution frontier; (III) generative AI is a derived state-space transformer that exploits the existing distribution without possessing an independent knowledge trajectory; and (IV) current AI provenance evaluation ("AI-generated" judgments) measures distributional proximity to training data, not knowledge-trajectory contribution – the wrong metric for the right question.

The synthesis: human expert and AI are complementary operators. The expert provides direction, domain K_{past}, and the capacity to shift the field; AI provides throughput. Labeling expert-directed AI-assisted work as "AI-generated" mistakes the tool for the author.

1 Introduction

This paper has a concrete occasion. The author is a hardware engineer with twenty years of experience in ASIC and PHY design: polar forward error correction (FEC) codec implementation at Xilinx prior to 3GPP 5G NR standardization; math engine silicon for AI and wireless workloads; die ownership of the display and sensor subsystem in the Microsoft HoloLens v2 silicon; FPGA design for Azure cloud infrastructure. In 2026, the author submitted a technical paper to arXiv in the computer architecture (cs.AR) category describing a unified bandwidth efficiency model for HBM4 memory. Potential endorsers raised the concern that the work appeared "AI-generated."

It was not. It was *AI-assisted*: written with language model support under the continuous direction of a domain expert whose K_{past} includes the design of circuits that do not exist in any publicly available training corpus. The polar FEC implementation, the HoloLens silicon partition, the specific HBM4 bandwidth efficiency model – none of these were in the training data. The engineering judgment that produced the claims in that paper resided in the human author, not in the tool.

The conflation of AI-assisted with AI-generated is the engineering instance of a broader problem that the author had already drawn – literally, on poster paper – in 2022 and 2023, and argued publicly in a June 2025 Nerd Nite Fort Collins talk titled *Robo Norbert Wiener* [1]. That talk used an AI model of Norbert Wiener – the founder of cybernetics – to diagram modern AI technology and make predictions about its future. The talk itself was an instance of the two-operator model this paper formalizes: the author's K_{past} (knowledge of cybernetics, AI systems, and their social consequences) directed an AI to instantiate Wiener's perspective, producing analysis neither could generate alone. The medium changed; the logic did not.

The Central Claim. Authorship is a property of the *knowledge trajectory* $\{K_t\}$ that produced a work, not of the tools used to produce it. AI assistance is a tool – structurally equivalent to LaTeX (which formats equations the author cannot hand-compose in print-quality typography), to Python simulation (which executes numerical models the author specified), or to a calculator (which multiplies numbers the author told it to multiply). Labeling a paper "AI-generated" because it was written with language model assistance is equivalent to labeling it "LaTeX-generated" because it was typeset in LaTeX. The tool is not the author. The K_t is the author.

1.1 The Drawings as Primary Evidence

In 2022 and 2023, before the public release of ChatGPT and DALL-E 2 had saturated public discourse on AI-generated art, the author produced a series of hand-drawn diagrams on poster paper – conceptual sketches made with markers and tape, pinned to a wall. These were working notes, not illustrations of published theory.

The drawings turned out to anticipate, in rough but precise form, four claims about the relationship between human creativity and generative AI that the subsequent three years of public deployment have made urgent. This paper argues for those claims using the drawings as primary evidence: artifacts that crystallized the argument before the argument was fashionable, produced by a human K_{past} that no model in 2022 contained.

Why drawings? The medium is self-demonstrating. These diagrams were produced under the same cognitive constraints the paper analyzes: incomplete information, iterative refinement, a mapping from an internal knowledge state to an external artifact. An AI cannot have produced them in 2022 – not because the generation capability did not exist, but because the specific K_{past} that generated the Faraday analogy, the Robocat punchline, and the USAF resolution chart argument was not in any training set. The same is true of the HBM4 bandwidth efficiency model: it exists in the world because a specific human K_{past} produced it, not because a model sampled a high-probability region of its training distribution.

1.2 Scope and Contributions

This is a position paper grounded in a dynamical systems reading of creative cognition. The primary contributions are:

1. A formal state-machine model of human creative cognition (the *Atomic Process* framework, §3).
2. A derivation of the economic value of creative and intellectual work from a flux-rate law – explaining why AI assistance drives down the price of generic output without eliminating the value of expert-directed work (§4).
3. A characterization of generative AI as a *derived* system – one that approximates learned transition functions without possessing an independent knowledge trajectory (§5).
4. An argument that current AI provenance judgments ("AI-generated" vs. "AI-assisted") are metrically equivalent to evaluating spatial frequency response and inferring scene content (§6).
5. A two-operator synthesis model specifying how human K-trajectory and AI throughput combine productively (§7).

2 Background

2.1 Cybernetics and the Feedback Roots of Creative Systems

The framework developed in this paper is cybernetic in origin. Norbert Wiener's foundational definition – cybernetics as "the scientific study of control and communication in the animal and the machine" [2] – established that biological and mechanical systems share a common formal structure: feedback loops that update internal state in response to the difference between desired and observed output. W. Ross Ashby extended this to adaptive systems, showing that any sufficiently complex feedback system tends toward regulatory stability [3].

The K-trajectory model developed in Section 3 is a cybernetic system in Wiener's sense: K_{t+1} is updated by comparing the current creative output against the creator's internal goal state, with the error driving the next action. The KC-130 navigator's dead reckoning is dead-reckoning cybernetics – a biological feedback controller predating its formalization by Kalman [4]. The LSTM cell update (Equation 2) is a differentiable implementation of the same feedback principle, trained rather than designed.

Wiener did not stop at the engineering formalism. In *The Human Use of Human Beings* [5], he extended cybernetics directly to the economics of labor:

> The modern industrial revolution is similarly bound to devalue the human brain, at least in its simpler and more routine decisions. — Wiener, 1950

This is the Iron Law of Creative Flux (Section 4) stated seventy-five years early. Wiener identified that automation devalues the cognitive outputs that lie *within* a machine's feedback envelope – the "simpler and more routine decisions." What he did not formalize, but gestured toward, is that outputs requiring the *redefinition* of the feedback envelope retain value precisely because no machine can automate that step. The $d\Phi/dt$ term in Equation (3) is the mathematical statement of Wiener's intuition.

In his final book, *God and Golem, Inc.* [6], Wiener posed three questions that define the boundary of machine capability: Can a machine learn? Can a machine reproduce? Can a machine *create*? He answered the first two affirmatively and left the third unresolved. The present paper is, in part, a formal answer to Wiener's third question: machines can generate; they cannot create, in the sense that creation requires a K-trajectory with an independent Fogbank. The Fogbank is where Wiener's third question lives.

What distinguishes the present paper from classical cybernetics is the explicit treatment of that Fogbank – the region of state space beyond the current feedback loop's reach. Classical cybernetics optimized within a defined error signal; the transformational creativity this paper analyzes requires redefining the error signal itself. This is the move from control to co-creation: from optimizing a fixed objective to discovering that the objective needs to change.

2.2 Dynamical Systems and Creativity

A dynamical system is characterized by a state space $\mathcal{X}$, a transition operator $f : \mathcal{X} \to \mathcal{X}$, and a trajectory $\{x(t)\}_{t=0}^{\infty}$ induced by iterating f from an initial condition $x(0)$. Creative systems – both biological and artificial – can be cast in this language. Boden's foundational taxonomy [7] distinguishes *combinatorial* creativity (recombining existing elements), *exploratory* creativity (traversing the boundaries of a known style space), and *transformational* creativity (restructuring the space itself). In dynamical terms: combinatorial is sampling within a basin of attraction; exploratory is traversing the attractor's boundary; transformational is finding or creating a new attractor.

Generative AI systems as of 2025 are powerful combinatorial and early exploratory systems [8, 9]. The question this paper addresses is whether they have the capacity for the transformational – and whether that question even matters economically.

2.3 Generative AI: A Brief Taxonomy

For the purposes of this paper we treat generative AI as any learned system of the form $x(t+1) = f_\theta(x(t))$ or, in the autoregressive case, $p(x_{t+1} \mid x_{0:t})$, where the parameters θ are optimized against a corpus of human-produced artifacts. Three families matter here:

- **Autoregressive language models** (GPT-3 [10], GPT-4 [11]): parameterize $p(x_{t+1}|x_{0:t})$ over token sequences. Creative output is conditional generation.
- **Diffusion models** (DALL-E 2 [9], Stable Diffusion [12]): learn to reverse a noise process over image latent codes. Creative output is conditional denoising.
- **RLHF-aligned models**: post-train either family against human preference labels [13], shaping the output distribution toward human-judged quality. The alignment step is, in the Robocat framing of Section 5, the "thank you" layer.

2.4 Economic Context: Labor Markets and Automation

Ricardo's Iron Law of Wages holds that in a competitive labor market, wages converge to subsistence level [14]. The mechanism is supply elasticity: when labor supply can expand without bound at a given skill level, the marginal worker drives wages to the minimum acceptable. The same mechanism applies to any commodity whose production can be automated at near-zero marginal cost.

AI image generation has made the production of a competent illustration effectively free [15]. The empirical wage effects on commercial illustrators are already measurable [16]. This paper does not dispute the short-run wage depression; it argues for the specific economic quantity that AI cannot automate, and therefore the production factor that retains value.

3 Proposition I: The Atomic Process

Claim. Human creative cognition is a knowledge state machine:

$$K_{t+1} = f(K_t,\, \theta_t,\, \epsilon_t) \tag{1}$$

where $K_t \in \mathcal{K}$ is the creator's knowledge state at time t, θ_t parameterizes the cognitive strategy (what to work on, what to discard), and ϵ_t is stochastic noise – the role of accident, serendipity, and environment in creative discovery.

The drawing labels the cognitive policy $\lambda = f(K_t, \theta_{t+1})$: the idea rate λ (how quickly new concepts emerge) is a function of current state and the next strategy update. The Markov structure is explicit: $\mathrm{K}_{\mathrm{past}}$ flows rightward toward $\mathrm{K}_{\mathrm{future}}$ through a filmstrip of discrete states. The "Fogbank" region, drawn as an undefined space above the track, is the exploration frontier – the part of $\mathcal{K}$ not yet reachable from the current trajectory.

3.1 A Worked Instance: The KC-130 Navigator as Human LSTM

Before formalizing the isomorphism, consider a concrete instantiation the author has operated directly. A KC-130 Hercules navigator on a long-range mission flies with a CR-3 circular slide rule, aeronautical charts, instrument approach plates, and one 8.5×11 inch flight plan. That single page encodes the complete planned state trajectory

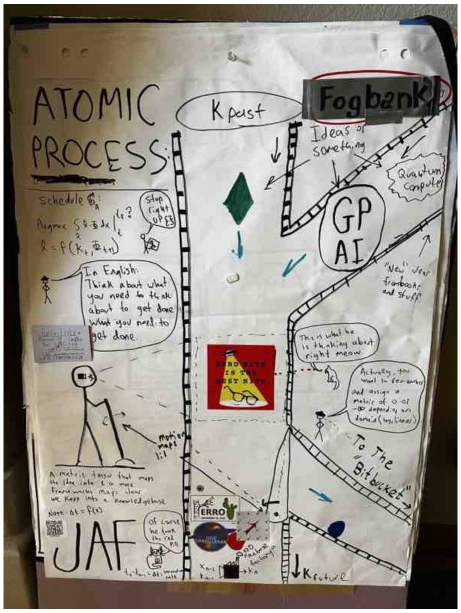

Figure 1: The Atomic Process (2022). Hand-drawn on poster paper. Key elements: the K_{past}–$\mathrm{K}_{\text{future}}$ trajectory with the Fogbank exploration frontier; the GP/AI node with quantum compute annotation; the schedule equation $\lambda = f(K_t, \theta_{t+1})$; the bottom gloss: "A machine tensor that maps the idea state to a new framework / maps ideas we keep into a knowledge base. $\alpha_t = f(t)$." Signed JAF.

for a multi-hour multi-variable mission: route waypoints, fuel states at each fix, weather alternates, crew assignments, performance limits, contingency decisions. It is a compressed K-trajectory document – a single human-legible artifact that *is* the plan.

During the flight, the navigator runs a recursive state estimator. At each fix: prior position estimate plus elapsed time plus wind model yields a predicted position; a visual checkpoint, VOR/DME reading, or radar return yields a measurement; the two are fused into an updated position estimate that propagates forward. This is dead reckoning – and it is, operationally, a hand-executed Kalman filter.

The navigator is a biological sequential state estimator with gated memory: old position fixes are explicitly discarded once superseded; new sensor inputs are weighted by their known reliability; the heading call to the crew is the output gate – what to act on right now. The rest stays internal, in state.

3.2 Structural Isomorphism with LSTM

The Atomic Process drawing formalizes this structure. The LSTM cell update is:

$$\begin{aligned}
f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \quad \text{(forget gate)} \\
i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \quad \text{(input gate)} \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tanh(W_c[h_{t-1}, x_t] + b_c) \\
o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \quad \text{(output gate)} \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned} \tag{2}$$

The correspondence to Equation (1) and to the navigator is exact:

- **Cell state** $c_t \leftrightarrow \mathrm{K}_{\text{past}}$: the long-horizon memory that persists across the mission – the route, the fuel plan, the filed alternates. Updated but never fully reset.
- **Hidden state** $h_t \leftrightarrow$ current working estimate: the active position fix, the heading being flown, the fuel state right now. Short-horizon, high-update-rate.
- **Forget gate** $f_t \leftrightarrow \theta_t$ (cognitive strategy): the deliberate decision of what to drop. Old position fixes are discarded once a better one arrives. The navigator chooses what to forget.
- **Input gate** $i_t \leftrightarrow$ Fogbank input: what new information from the environment gets written into cell state. A visual landmark, an unexpected weather cell, a fuel discrepancy – these are ϵ_t arriving at the input gate.
- **Output gate** $o_t \leftrightarrow$ the heading call: what portion of the internal state is exposed as action. The crew receives a correction; the full state model stays internal.

This isomorphism is not a metaphor. The Atomic Process drawing is a self-portrait of a human LSTM, drawn by someone who had operated as one. The "machine tensor that maps the idea state to a new framework" in the drawing's caption is, in formal terms, the LSTM cell update equation – arrived at by introspection rather than by reading Hochreiter and Schmidhuber [17].

The flight plan as a K-trajectory document is also structurally significant. One page – the information density of a compressed state sequence – does the work that would require gigabytes of raw sensor logs to reconstruct. The plan is not the flight; it is the *model* of the flight, in the same way that a research paper is not the research but the compressed K-trajectory that makes the research transmissible. High information density per page is a signature of human K-trajectory compression, distinct from AI output which produces high token volume at lower semantic density per token.

The differences between the human LSTM and the machine LSTM are architectural, not categorical:

- **Training signal**: K_t in a human is shaped by the full trajectory of lived experience – including navigating actual aircraft. The machine LSTM is shaped by backpropagation through a fixed loss on a fixed corpus.
- **State dimensionality**: Human K_t grows unboundedly with experience. Machine c_t has fixed dimension set at training time.
- **State persistence**: K_t is never reset; sleep consolidates rather than erases. Machine context windows are bounded.
- **Fogbank access**: Human ϵ_t provides genuine exploration of uncharted $\mathcal{K}$ – an unexpected weather cell forces a route deviation that generates genuinely new state. Machine temperature sampling deviates from the learned distribution but cannot reach outside it.

The last point is the critical asymmetry. A navigator encountering a weather cell not on any chart is outside the training distribution – and survives by using the full generality of their K-trajectory to improvise. A model cannot do this by construction [18].

3.3 The Compressed Model and Its Elaboration

The flight plan generalizes. Consider a second instance from the author's engineering practice: the design of a polar forward error correction (FEC) codec at Xilinx, prior to its inclusion in the 3GPP 5G NR standard.

When the team needed to settle fundamental architectural questions about the decoder, the author did not begin by writing production code. Instead, a few days were spent building a compact MATLAB model – a minimal, mathematically faithful simulation that answered every structural question: belief propagation schedule, frozen bit selection, rate matching, throughput ceiling, latency budget. Engineers unfamiliar with this mode of work were uneasy: code was not appearing. What was appearing was something more valuable – a compressed K-trajectory artifact that made the entire design space legible.

Once the MATLAB model existed, the team could generate code that worked on the first attempt. The code was elaboration. The model was the thought.

This is the same structure as the flight plan. In both cases:

1. An expert with deep domain $\mathrm{K}_{\mathrm{past}}$ produces a *compressed model* – small, precise, high semantic density per line.
2. The compressed model encodes the critical constraints, the binding tradeoffs, the decisions that cannot be undone.
3. Elaboration – generating code from the model, flying the route from the plan – is the high-volume, lower-judgment step that follows from the model.

The ratio of model to elaboration is a measure of K-trajectory density. An expert writes ten lines of MATLAB that generate ten thousand lines of RTL. A junior engineer writes ten thousand lines hoping some of them survive review. The difference is not effort; it is $\mathrm{K}_{\mathrm{past}}$.

This has a direct implication for AI-assisted intellectual work. A language model used to draft prose around a bandwidth efficiency derivation that the author independently specified is performing elaboration. The compressed model – the four-factor efficiency decomposition, the binding t_{RRD} constraint, the specific HBM4 timing parameters – came from the author's $\mathrm{K}_{\mathrm{past}}$. The AI inflated it into paragraphs. The value is in the model. The authorship is in the compression.

Critics who label such work "AI-generated" are misidentifying which step constitutes the contribution. They are confusing the flight plan with the aircraft, the MATLAB model with the silicon, the argument with its prose container.

3.4 The "GP/AI" Node

The drawing includes a node labeled "GP AI" connected to both the $\mathrm{K}_{\mathrm{past}}$ track and the Fogbank, with an arrow to "Quantum Compute" and the annotation "Ideas of something." This is prescient: the author was modeling AI not as a replacement for the human K trajectory but as an *additional input* to Equation (1) – a source of new x_t that can accelerate K_t toward the Fogbank frontier. GP (Gaussian Process, or General Purpose) AI serves as a stochastic oracle that samples from a large learned distribution, feeding novel combinations into the human state machine.

This is the co-creative model: human K_t supplies direction, judgment, and the capacity for transformational moves; AI supplies high-bandwidth sampling from the current distribution. Neither suffices alone.

4 Proposition II: The Iron Law of Creative Flux

Claim. The economic value of a creative work is proportional to the *rate of change* of the creative field it induces, not to the static quality of the output:

$$V_{\mathrm{art}} \propto \frac{d\Phi}{dt} \tag{3}$$

where $\Phi(t)$ is the *creative flux* – the total influence of the work on subsequent creative production in its domain.

4.1 The Faraday Analogy

The drawing writes $\Delta E = d\Phi/dt$ – Faraday's law of electromagnetic induction, where an electromotive force (volt-

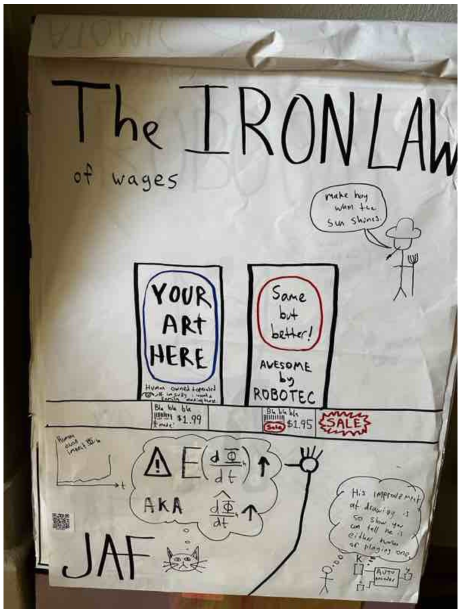

Figure 2: The Iron Law of Wages (2022–2023). Human art at \$1.99; "Same but better! AWESOME by ROBOTEC" at \$1.95 with SALE sticker. Below: $\Delta E = d\Phi/dt$, labeled "AKA $d\Phi/dt \uparrow$". Bottom left: human economic activity going asymptotic — output approaching a ceiling before a probable crash. Bottom right: AUTO generator block diagram. Signed JAF.

age) is induced by a *changing* magnetic flux. The analogy is deliberate:

- **Static flux** ($\Phi = \text{const}$): no EMF. A painting that competently reproduces a known style generates no new artistic voltage – it adds to the field but does not change it.
- **Rising flux** ($d\Phi/dt > 0$): positive EMF. A work that shifts what other artists consider possible – Picasso's cubism, Coltrane's sheets of sound, Duchamp's readymades – generates artistic voltage proportional to the rate of shift.

AI can produce static flux at any quality level at near-zero marginal cost. Hence the \$1.95 price with SALE sticker: infinite supply, elastic pricing. The \$1.99 human piece competes on the same axis and loses – if static flux is the product. But the human can produce $d\Phi/dt \neq 0$. The AI, as argued in Section 5, cannot – not because it lacks generation capability but because it has no K_t of its own to shift.

4.2 Ricardo's Iron Law Applied

Ricardo's Iron Law states that wages in a competitive market converge to the subsistence wage of the marginal worker. The mechanism requires an elastic labor supply at a given skill level [14]. Applied here:

1. AI is an infinitely elastic supplier of static-flux creative work.
2. Market wages for static-flux work converge to AI's marginal cost ≈ 0.
3. Human artists who compete on static-flux output lose.
4. Human artists who compete on $d\Phi/dt$ – cultural novelty, field shift, the discovery of new attractors – face no AI competition on that axis.

The drawing captures this in the bottom-left graph: human economic activity rising toward an asymptote – output compressing against a ceiling before a probable crash. This is the logistic trajectory of a system saturated by infinite-supply competition: growth continues but decelerates, and the curve warns of a discontinuity ahead rather than a smooth equilibrium. Under the Iron Law, artists competing on static-flux output approach that ceiling; the crash is the phase transition when the market reprices their labor to AI marginal cost. The survivors are the ones who shift the distribution before the transition.

4.3 The AUTO Generator

The bottom right of the drawing shows a block labeled "AUTO" with generator lines, alongside "$K =$" with an auto-regressive block diagram. This is a literal depiction of autoregressive generation: the AUTO block produces output from a learned K (knowledge base). The equation $K = \text{AUTO}(\ldots)$ anticipates the GPT/diffusion framing – the model is a compressed representation of its training corpus.

The visual joke in the drawing is structural: the AUTO block is placed on the *same horizontal axis* as the \$1.95 price tag. The generator is the pricing mechanism. Infinite AUTO output is why ROBOTEC can undercut.

4.4 What Retains Value

The implication of Equation (3) is that artistic value is not a function of the artifact but of the artifact's effect on the field. This is not a new claim – it is a restatement of the art-historical consensus on influence [19]. What is new is the operationalization: $d\Phi/dt$ can be approximated by measuring citation velocity in creative networks (how quickly a work is referenced, remixed, or cited by subsequent work), analogous to academic citation analysis [20].

Crucially, $d\Phi/dt$ requires a temporal sequence of K_t states – an ongoing creative trajectory – to generate. A one-shot AI prompt does not produce a trajectory; it samples a point. The sustained creative career, the evolving body of work, the recognizable artistic K-trajectory is the product that AI cannot replicate.

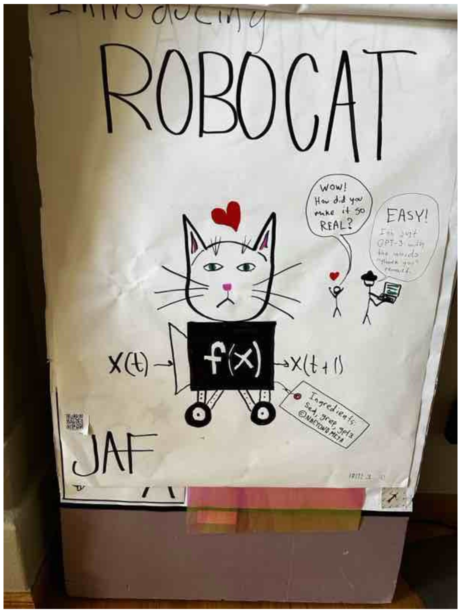

Figure 3: Introducing Robocat (2022–2023). State-space block: $x(t) \to f(x) \to x(t+1)$. Interlocutor: "Wow! How did you make it so REAL?" Robocat: "EASY! I'm GPT-3 with the words 'thank you' removed." Red heart; cat face on a wheeled chassis. Signed JAF.

5 Proposition III: Robocat and the Derived Creative System

Claim. A generative AI creative system is a *derived* state-space transformer – one that approximates the transition function of human creative output without possessing an independent knowledge trajectory. Its creative capacity is bounded above by the distribution of its training data.

5.1 State-Space Formulation

The drawing formalizes Robocat as:

$$x(t+1) = f(x(t)) \tag{4}$$

where $x(t)$ is the current context (prompt, partial generation, conditioning signal) and f is the learned transition function. This is precisely the autoregressive factorization of a language model or the denoising step of a diffusion model: the next token or image patch is a function of the current context.

The formulation is correct and the cat is charming. The question the drawing raises is: *who authored f?*

f was authored by the training corpus – which is human creative output. Robocat is a *compressed representation of human creative history* equipped with a generation engine. It has no K_t of its own; its effective K_{past} is a weighted average of every artist, writer, and musician in the training set. The heart on Robocat's head is accurate: it is genuinely lovable, genuinely capable, and genuinely derivative.

5.2 "GPT-3 with the Words 'Thank You' Removed"

The punchline encodes a precise technical observation about RLHF alignment.

GPT-3 base [10] is a raw autoregressive model trained on next-token prediction. InstructGPT and subsequent aligned models [13] apply Reinforcement Learning from Human Feedback (RLHF) to shape the output distribution toward human-judged quality – helpfulness, harmlessness, and honesty. One observable effect is that RLHF models say "thank you," apologize, and hedge. They have learned the *social compliance layer* of human communication.

DALL-E and early Stable Diffusion image generators are, structurally, base models with minimal RLHF – or with RLHF optimized for *visual* rather than *social* preference. In terms of the drawing: they are GPT-3 (a powerful generative prior over human-produced content) with the conversational politeness layer removed and replaced by a visual quality prior.

The implication is that the "artistic" quality of AI-generated images in 2022–2023 was not a new capability – it was the base model's generative power applied to the image domain, with the social compliance training replaced by aesthetic preference training. ROBOCAT = base model + aesthetic RLHF - social RLHF.

This framing carries a non-obvious consequence for policy: RLHF alignment and creative capability are in partial tension. The same training that makes a model safe and polite also makes it conservative – it learns to produce outputs in the center of the preference distribution, which is the center of the training distribution. Removing the "thank you" is, in part, removing the pull toward the mean. The Fogbank (Section 3) is harder to reach when you have been RLHF'd to stay polite.

5.3 The Authenticity Question

The interlocutor in the drawing asks: "How did you make it so REAL?" This is the authenticity question posed in the aesthetics literature since Goodman [21]: is a perfect forgery of a painting artistically equivalent to the original?

The dynamical systems answer: authenticity is not a property of the artifact $x(t+1)$ but of the trajectory $\{K_t\}$ that generated f. Robocat produces outputs indistinguishable from human art by static perceptual metrics. But there is no K-trajectory behind the outputs – no evolving knowledge state, no Fogbank traversal, no sustained creative history. The realness is real; the au-

Figure 4: TCAMMM $\in \mathbb{R}^3+t$ (2022–2023). Title is shorthand for *The Canonical Art Modulatable Meta Materials in three dimensions of space plus time.* A USAF-1951 three-bar resolution test chart (MIL-STD-150A) mounted on poster paper above a holographic color sheet. Lower left: TCA logo. The USAF chart measures spatial frequency response – a proxy for fidelity, not meaning. The holographic panel is irreducible to any such metric. Mixed media. Signed TCA.

thorship is distributed over the training corpus and the prompt author.

This suggests a *functional* definition of authentic authorship: a work is authentically authored if there exists a K-trajectory that produced it. Multiple K-trajectories can contribute (collaborative work); a prompt author's K_t contributes to a Robocat-mediated work. The question of degree of authorship maps to the question of how much of the specific output was determined by the human K_t versus the model prior.

6 Proposition IV: The Resolution Problem

Claim. Current evaluation metrics for AI-generated art measure perceptual fidelity (the analog of spatial frequency response) but are blind to distribution-shift contribution ($d\Phi/dt$) – the quantity that determines long-run artistic and economic value.

6.1 The USAF-1951 Chart as Metaphor

The USAF-1951 three-bar resolution test chart (Figure 4) was developed by the U.S. Air Force to standardize the measurement of optical system resolution [22]. Groups of three horizontal and three vertical bars at decreasing spatial frequencies allow a single number – the limiting resolution in line pairs per millimeter – to characterize an imaging system.

The chart is a near-perfect instrument for its purpose. It tells you exactly whether your lens can distinguish fine detail. It tells you nothing about whether the scene being imaged is interesting.

The title TCAMMM $\in \mathbb{R}^3+t$ decodes as *The Canonical Art Modulatable Meta Materials in three dimensions plus time.* The $+t$ term is the operative claim: resolution, fidelity, and material sophistication increase with time. This is empirically correct for camera and AI systems – image generation quality by FID score [23] and human preference ratings has improved reliably with scale [9,12]. The holographic color sheet below the chart – shimmering, unreproducible in print – is the visual argument for what the $+t$ trajectory cannot capture.

6.2 The Metric Hierarchy

Current AI art evaluation operates on a stack of proxies:

1. **Fréchet Inception Distance (FID)** [23]: measures distributional distance between generated and real image sets in Inception feature space. Excellent for measuring whether the generated distribution matches the training distribution. Structurally blind to whether the training distribution itself has cultural value.
2. **CLIP score** [24]: measures alignment between generated image and a text prompt in CLIP embedding space. Excellent for prompt fidelity. Blind to whether the prompt describes something novel.
3. **Human preference** (RLHF reward model, A/B testing): measures whether a human judge prefers the generated image. Subject to distribution bias – judges prefer images that match their expectations, which are formed by the existing art distribution.

All three metrics are analogs of the spatial frequency response measurement: they describe how well the system reproduces or matches a *known* distribution. None can measure $d\Phi/dt$ because that requires observing the effect of the work on the subsequent creative field – a measurement that takes years and cannot be operationalized as a loss function.

6.3 The Measurement Problem for $d\Phi/dt$

Measuring creative influence is not impossible – it is simply not instantaneous. Academic citation analysis provides a working model: the citation velocity of a paper in the years following publication is a proxy for its $d\Phi/dt$ in the scientific knowledge field [20]. Creative work has analogous measurables: sampling rates in music, visual reference rates in design, stylistic adoption rates in visual art communities.

The structural problem for AI evaluation is that these measurements cannot be computed at generation time. A metric that requires a 5-year follow-up study cannot supervise a training run. This is not a limitation that will be overcome by scaling compute – it is a consequence of the causal structure of creative influence.

The practical implication: AI art evaluation as currently practiced optimizes for the wrong objective. A system that maximizes FID and human preference produces extremely convincing static-flux output. It does not and cannot optimize for $d\Phi/dt$ because the signal is not available. This does not mean AI art is valueless – static flux has value. It means the currently measurable value is a lower bound on the total potential value of human creative work, which includes the unmeasured (and unmeasurable in real time) $d\Phi/dt$ component.

6.4 Resolution vs. Meaning

The holographic panel below the USAF chart is the punchline. It cannot be reproduced by the chart. Its value is precisely in its irreducibility to the metric. The drawing proposes this as the fundamental structure of the human-AI art problem: the tools that measure resolution are not the tools that capture meaning, and building better resolution tools does not bring you closer to the meaning measurement.

This is a formal statement of what the art world has claimed informally since the first photography debates [25]: technical fidelity and artistic value are orthogonal axes. What the drawing adds is the dynamical systems interpretation: fidelity measures a static snapshot of the field; value is a property of the field's trajectory.

7 Synthesis: Complementary Operators

The four propositions converge on a single architectural claim about human-AI creative systems.

7.1 The Two-Operator Model

Define the creative production system as a composition of two operators:

$$\Phi(t+1) = \mathcal{H}(K_t, \epsilon_t) \circ \mathcal{A}(\Phi_{\text{train}}, p_t) \tag{5}$$

where:

- $\mathcal{H}$ is the human operator: takes knowledge state K_t and noise ϵ_t, produces a creative output that may shift Φ.
- $\mathcal{A}$ is the AI operator: takes the training distribution Φ_{train} and prompt p_t, produces a high-fidelity sample from that distribution.
- The composition $\circ$ denotes human curation and selection: the human chooses which AI outputs enter their creative process and how.

These operators are asymmetric in three ways:

1. **Distribution access**: $\mathcal{H}$ can produce output outside Φ_{train}; $\mathcal{A}$ cannot (in expectation, rigorously).
2. **Throughput**: $\mathcal{A}$ produces high-quality samples at millions of times the throughput of $\mathcal{H}$.
3. **Cost**: $\mathcal{A}$ has near-zero marginal cost per sample; $\mathcal{H}$ has substantial fixed cost (the artist's life and accumulated K_t).

The asymmetry means neither operator dominates. $\mathcal{H}$ without $\mathcal{A}$ searches the Fogbank slowly, with limited throughput for exploring adjacent ideas. $\mathcal{A}$ without $\mathcal{H}$ samples efficiently from a fixed distribution that does not evolve – a creative system in a fixed-point attractor. The composition is strictly more powerful than either alone.

7.2 Interface Design

The productive question for human-AI creative systems is not "who made it?" but "how is the interface between $\mathcal{H}$ and $\mathcal{A}$ structured?" The quality of the composition depends on:

- **Prompt specificity**: A high-information prompt p_t drawn from the human's current K_t positions the AI sample close to the human's Fogbank frontier. A generic prompt produces a center-of-distribution sample far from any individual Fogbank.
- **Curation depth**: The more selectively $\mathcal{H}$ filters $\mathcal{A}$'s output, the more of the $d\Phi/dt$ is attributable to $\mathcal{H}$'s judgment. Accepting the first AI output unmodified provides minimal human creative contribution; iterative refinement can be high.
- **Update rate**: The composition's field-shifting power depends on how quickly AI samples feed back into K_t. A fast iteration loop – generate, evaluate, refine K_t, generate again – can accelerate the human's Fogbank traversal. A slow loop (produce one work per month incorporating AI) does not leverage the throughput asymmetry.

7.3 The Atomic Process as Interface Protocol

The Atomic Process drawing (Figure 1) can be reread as an interface specification. The GP/AI node feeds into the K_{past}–K_{future} trajectory as an external signal source.

The human K_t state machine integrates AI output the same way it integrates any external stimulus: the stimulus enters at the current state, the transition function f processes it, and K_{t+1} reflects the update.

The critical design requirement is that the human K_t must retain its own dynamics. If the AI input is too dominant – if $K_{t+1} \approx \mathcal{A}(\Phi_{\text{train}}, K_t)$ with no human-specific ϵ_t contribution – then the human trajectory converges to the AI distribution. The artist becomes a high-cost sampler of the same distribution the AI samples cheaply. The economic consequence follows immediately from the Iron Law: wages converge toward zero.

The productive interface keeps ϵ_t alive – maintains the human's independent noise source, their access to their own Fogbank, their capacity to shift Φ. Concretely: use AI for throughput, not direction; let the human set the gradient direction in creative space while AI accelerates the gradient step.

7.4 The AI-Assisted vs. AI-Generated Distinction

The most consequential policy failure in current AI provenance discourse is the conflation of *AI-assisted* with *AI-generated*. The two-operator model makes the distinction precise.

A work is *AI-generated* when $\mathcal{H}$ contributes negligible K_t information to the output: the human provides a generic prompt, accepts the first output, and the work is determined primarily by $\mathcal{A}(\Phi_{\text{train}}, p_t)$. The human K_t is not operative – the work would be statistically similar for any human with access to the same prompt and model.

A work is *AI-assisted* when $\mathcal{H}$ contributes substantial, expert-specific K_t information: the prompt encodes domain knowledge not present in the training distribution, the curation rejects AI outputs that are technically incorrect, and the final work reflects judgments that require the specific K_{past} of the author to make. The AI is a high-throughput drafting tool; the human is the author.

The engineering case is clearer than the art case. Consider a hardware engineer who:

- Designed polar FEC codecs prior to their inclusion in the 3GPP 5G NR standard,
- Owned die partitions on a mixed-reality headset silicon,
- Wrote bandwidth efficiency models grounded in JEDEC memory specifications that postdate the model's training cutoff.

When this engineer uses a language model to draft prose around claims they specified, the language model is doing what LaTeX does for equations: providing a formatting and composition service. The K_{past} that produced the bandwidth efficiency model – the specific derivation of $t_{\text{RRD,eff}}$ from HBM4 timing parameters, the four-factor efficiency decomposition, the BFS microarchitecture – is not in any training corpus. The model cannot produce those specific claims without the human's direction because the human's direction is what makes them true and novel.

Challenging such work as "AI-generated" mistakes the tool for the author. It is the resolution chart error applied to provenance: measuring the distributional proximity of the prose style (which a language model's style does influence) and inferring from that measurement that the *content* is AI-generated. The metric is real; the inference is wrong.

A proposed standard: work is AI-assisted, not AI-generated, when the following conditions hold:

1. The technical claims are falsifiable and their correctness depends on domain K_{past} not present in public training data.
2. The human author can independently derive or verify each claim without AI assistance.
3. The AI's contribution is localized to prose composition, citation formatting, or similar tooling functions rather than content generation.

A calculator-assisted proof is a proof. A LaTeX-typeset theorem is a theorem. An AI-assisted technical paper by a domain expert is a technical paper.

7.5 Implications for AI Art Policy

The two-operator model clarifies the policy stakes:

- **Copyright**: the question of copyright for AI-generated works is, in this framework, a question of K_t attribution. Works where the human K_t determines the specific direction and the AI fills in the high-frequency content have a clear human author. Works where the human provides only a generic prompt have distributed authorship across the training corpus.
- **Training data**: training on copyrighted works without compensation is, in this framework, a capture of human K-trajectories without attribution. The AI's Φ_{train} is built from human $d\Phi/dt$ without paying for the field-shifting work. This is the structural wrong in current training data practice, independent of specific legal theory.
- **Disclosure**: a useful disclosure standard would require documenting the $\mathcal{H} \circ \mathcal{A}$ composition – what was the human prompt, what was AI-generated, what was curated – rather than a binary human/AI label. This maps the creative contribution to the K_t attribution problem.

8 Conclusion

Four hand-drawn diagrams from 2022–2023, produced before the current generative AI discourse had fully formed, anticipate the core structural argument: human creative cognition and generative AI are asymmetric state-space operators, and their interaction is governed by a flux-rate law that current evaluation metrics cannot measure.

The Atomic Process establishes that creativity is a knowledge state machine with a trajectory; the Iron Law establishes that economic value derives from the trajectory's rate of change, not from any single output; Robocat establishes that AI is a derived system without its own trajectory; and the USAF resolution chart establishes that the tools used to measure AI creative quality are orthogonal to the quantity that determines long-run value.

The synthesis – the two-operator model of Equation (5) – is not a prediction that human artists will win by being human. It is a design specification: the productive composition of human and AI creative work requires maintaining the human K-trajectory's independence, using AI throughput in the direction set by the human, and measuring creative contribution by field-shift rather than output fidelity.

The drawings themselves are an instance of this argument. They were produced by a human K-trajectory that no model in 2022 could have generated – not because the generation capability was absent, but because the specific K_t that produced the Fogbank metaphor, the Faraday analogy, and the Robocat punchline was not in any training set. They are the Fogbank made visible.

Whether that matters economically depends on whether the field they shift is one that pays. The Iron Law is still operating. Sun Tzu's first principle is to know the terrain. The drawings say: know which axis you are competing on.

Acknowledgments

The hand-drawn diagrams analyzed in this paper were produced by the author in 2022–2023 as part of an ongoing art practice under The Canonical Art LLC. No AI tools were used in their production.

AI Tool Disclosure (per MIT Press policy): Claude (Anthropic), a large language model, was used as a writing assistant during manuscript preparation – drafting prose, formatting equations, and organizing structure under the continuous direction of the author. All technical claims, arguments, and conclusions are the author's own. The author independently derived and can verify every assertion in this paper. The author assumes full responsibility for all content.

References

[1] J. A. Fritz, "Robo norbert wiener," Public talk, Nerd Nite Fort Collins S25E6, June 2025, fort Collins Nerd Nite, https://fortcollins.nerdnite.com/2025/05/27/.

[2] N. Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine.* Cambridge, MA: MIT Press, 1948.

[3] W. R. Ashby, *An Introduction to Cybernetics.* London: Chapman & Hall, 1956.

[4] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[5] N. Wiener, *The Human Use of Human Beings: Cybernetics and Society.* Boston, MA: Houghton Mifflin, 1950.

[6] ——, *God and Golem, Inc.: A Comment on Certain Points Where Cybernetics Impinges on Religion.* Cambridge, MA: MIT Press, 1964.

[7] M. A. Boden, *The Creative Mind: Myths and Mechanisms*, 2nd ed. London: Routledge, 2004.

[8] A. Elgammal, B. Liu, M. Elhoseiny, and M. Mazzone, "Can: Creative adversarial networks, generating "art" by learning about styles and deviating from style norms," in *Proceedings of the 8th International Conference on Computational Creativity (ICCC)*, 2017.

[9] A. Ramesh *et al.*, "Hierarchical text-conditional image generation with CLIP latents," in *arXiv preprint*, 2022. [Online]. Available: https://arxiv.org/abs/2204.06125

[10] T. B. Brown *et al.*, "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.

[11] OpenAI, "GPT-4 technical report," OpenAI, Tech. Rep., 2023. [Online]. Available: https://arxiv.org/abs/2303.08774

[12] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 10 684–10 695.

[13] L. Ouyang *et al.*, "Training language models to follow instructions with human feedback," in *Advances in Neural Information Processing Systems*, vol. 35, 2022.

[14] D. Ricardo, *On the Principles of Political Economy and Taxation.* London: John Murray, 1817.

[15] WIRED Staff, "The generative AI art boom is here. artists are worried," 2023.

[16] E. Krupnick, "Artists say AI image generators are hurting their livelihoods. can anything be done?" *The Washington Post*, 2023.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] G. Marcus, "Deep learning: A critical appraisal," *arXiv preprint arXiv:1801.00631*, 2018.

[19] H. Bloom, *The Anxiety of Influence: A Theory of Poetry.* New York: Oxford University Press, 1973.

[20] E. Garfield, "Citation indexes for science: A new dimension in documentation through association of ideas," *Science*, vol. 122, no. 3159, pp. 108–111, 1955.

[21] N. Goodman, *Languages of Art: An Approach to a Theory of Symbols.* Indianapolis: Hackett Publishing, 1968.

[22] U.S. Air Force, "MIL-STD-150A: Photographic lenses, military specification," United States Air Force Military Standard, 1959.

[23] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[24] A. Radford *et al.*, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning (ICML)*, 2021, pp. 8748–8763.

[25] W. Benjamin, "The work of art in the age of mechanical reproduction," *Zeitschrift für Sozialforschung*, vol. 5, no. 1, pp. 40–66, 1935.