

Alternative creativity: why generative AI is creative, and why it is not like us

Jordi Linares-Pellicer

`jordi.linares@upv.es`

Universitat Politècnica de València

Carlos Aliaga-Torro

Universitat Politècnica de València

Isabel Ferri-Molla

Universitat Politècnica de València

Juan Izquierdo-Domenech

Universitat Politècnica de València

Joaquin Taverner

Universitat Politècnica de València

Article

Keywords: alternative creativity, generative artificial intelligence, computational creativity, embodiment, authorship, human-machine co-creation

Posted Date: October 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11007009/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Whether generative artificial intelligence is creative is usually argued as a yes-or-no question, and both camps share an unexamined assumption, that there is one kind of creativity, the human kind, which the machine either has or lacks. This article defends a third position, alternative creativity. For any fixed medium the set of possible artefacts is finite and astronomically large, and almost none of it is valuable. Creativity, wherever it occurs, is therefore a compression of that space into a small manifold of meaningful candidates, followed by a value-guided search across the manifold. On the leading neuroscientific account, human brains work this way, and generative models demonstrably do. Neither creates from nothing, both recombine what they have internalised, both use constrained randomness, and both hold their skills in the strengths of their connections. In these respects the two are less different than is commonly supposed. Where they differ is in how the connections were shaped and in what each creator can reach at the moment of creation. A person's parameters are formed by a body acting in a physical and social world, a model's by an archive of what people have said and shown, and this difference in accessible context, together with a difference in how value is grounded, explains more of the divergence in what the two creators make than the difference of substrate does. In domains with checkable answers, machines already search beyond the human distribution under verifiable rewards. In the arts, where no verifier exists, they must borrow taste from people. From this account the article derives the homogenisation of machine output, the general intractability of tracing outputs to training inputs, the migration of authorship towards the specification of value, and the logic of human-machine collaboration, in which each creator supplies what the other constitutionally cannot.

Introduction

Ask a room of artists, engineers and philosophers whether artificial intelligence is creative and you will get a show of hands in every direction. One camp points to outputs that win prizes and move unknowing readers, and concludes that the machines now do what we do. The other answers that the systems are "stochastic parrots" manipulating form without meaning (Bender et al. 2021), that their products are pastiche, and that what looks like creation is a laundering of other people's work. The camps agree on almost nothing, yet both assume there is a single kind of creativity, the human kind, which the machine either has or lacks.

That assumption is the mistake, and this article attempts to replace it.

Our claim has two halves that must be held together. The first is that generative systems really are creative, by the behavioural standards creativity research itself uses, not by courtesy, and that they share with human creators a mechanism deep enough that the usual deflations fail. Whoever dismisses machine output as mere recombination of what already exists has indicted every human creator since the first, because nothing we know about human creation supports making from nothing either. The second half is that these systems are creative in a way that is not ours. We propose to call the machine's version alternative creativity. It is not a lesser quantity of the human kind and not a copy of it, but a second, differently built instance of the process that creativity is.

A further claim runs through the article and inverts a common intuition. The obvious place to look for the difference is the hardware, neuron against unit, plasticity against frozen weights, and those differences are real. We will argue that they are the least consequential of the divergences, and that what separates the two

creators most is what shaped their networks and what they can touch, a lived world of bodies and other people in one case and an archive of testimony in the other, together with the way each grounds the value that steers its search.

A ground rule first. Psychology converged long ago on an operational definition of creativity as the production of work that is both novel and appropriate, or valuable, to some community (Runco and Jaeger 2012). The definition is deliberately behavioural, silent on souls and muses, and it applies to any artefact-producing process, on neurons or on silicon. We adopt it for attributions of creativity to processes and their products. We use "creator" for any system that runs the process, without prejudging agency. Whether creativity also requires agency is a further question, addressed below; what creating feels like we set aside.

This collection asks how art and design are reshaped as machine learning enters the studio, and its questions, authorship and agency, co-creation, craft and materiality, the legal frameworks of machine-made art, are usually approached from the humanities, as in recent contributions to this journal (Calvo 2026; Fischman-Afori 2026). We approach them from the mathematics of the media and the engineering of the models, by conceptual analysis rather than new data, as researchers who build generative systems, a vantage we disclose as a source of bias. Where the evidence is mixed, as with the neuroscience of creativity and machine understanding, we flag the controversy instead of rounding it up into certainty.

The shared ground: a finite space, and the end of creation from nothing

Begin with the medium rather than the mind, a digital image of fixed dimensions and colour depth. A canvas of width W and height H pixels, each pixel taking one of k possible colours, admits exactly

$$N = k^{W \cdot H}$$

1

distinct images. Write S for this set, of cardinality $|S| = N$. Eq. (1) is an exact count, and it grows ferociously. For twenty-four-bit colour, $k = 2^{24} \approx 16.7$ million, and a 1024×1024 image, about one megapixel, permits $2^{24 \times 1,048,576} = 2^{25,165,824}$ configurations, of order $10^{7,575,668}$. The observable universe holds some 10^{80} atoms; the space of one-megapixel images exceeds any physical inventory of the cosmos by millions of orders of magnitude, and yet it is finite.

Two facts carry the argument. The first is that the space is total. It contains every image ever made and every image that could be made in the format, alongside overwhelmingly many arrangements that are simply noise. One scope note. S is a space of configurations, not of works. A work is a configuration embedded in a history of making and interpretation, which is why indiscernible configurations can be distinct works (Danto 1964, 1981), and only digital formats make the count exact, dense and performative media taking it as an idealisation (Goodman 1976). Our combinatorial claims concern configurations.

Second, the valuable points are almost none of the space. Let $V \subset S$ denote the configurations some human community would find worth keeping, coherent and also worth framing or publishing. Drawn uniformly at

random, a point of S lands in V with probability $|V|/|S|$, a number operationally indistinguishable from zero. This is the digital form of the monkeys at typewriters (Borel 1913).

Finiteness dissolves a persistent intuition. We speak of creativity as limitless, as an expansion of the possible, the popular inheritance of Romantic expressivism that Boden (2004) spends a chapter dismantling. Within a fixed format it is nothing of the kind. Kant, often blamed for the myth, reads better as an ancestor of our account, since his genius supplies original material and his taste disciplines it, a generation-and-selection pair *avant la lettre* (Kant 2000). No act of imagination adds an image to S , which already contains them all, and what changes across an art form's history is which regions of the container people have learned to value. The question is therefore how any system reaches the valuable works without cosmological patience, a question about search and structure, not metaphysical spark. Media are not always fixed; the most radical creativity changes the medium itself, which we treat below, with Boden, as navigation in a larger space of possible spaces.

The old saying that nothing comes from nothing thereby becomes precise. The creative act adds no possibilities. It locates the valuable ones inside a space that already contains everything makeable, and location requires information, a reduction of uncertainty about where, in a mostly worthless immensity, worth lies.

The shared ground: compression and navigation

How do creators find V ? Not by exhaustive search, which is impossible, and not by luck, which is too slow. They find it by searching not S itself but a drastically reduced model of it, built from experience, which already excludes almost everything meaningless.

This is the first of two operations we want to name, compression. A trained painter does not represent images as pixel arrays. Long exposure builds an internal model whose degrees of freedom are objects, edges, styles, moods and compositional relations, far fewer than $W \cdot H$ independent pixels, which is possible because meaningful images occupy a thin manifold within the raw space, the same fact that makes JPEG work. Let $M \subset S$ denote this manifold of meaningful configurations, of intrinsic dimension $d \ll W \cdot H$. Compression is the act of learning M , reducing the astronomical space of raw configurations to the far smaller space of things that could be a picture of something (Fig. 1). M collects the meaningful and V , inside it, the valued; coherence is necessary for value but never sufficient. Schmidhuber's formal theory states the strong version, that what we find interesting or beautiful is data permitting newly discovered compression (Schmidhuber 2010). The compression image has also been used against these systems, most memorably as a blurry JPEG of the web that loses information and confabulates (Chiang 2023). We take the same image in the opposite direction. Lossy compression is not the failure of creativity but its precondition, since only a compressed model makes the search tractable.

The second operation is navigation, because M still contains every competent image in every style. To make a particular valued work is to move across M towards a target region, this subject, that mood, in a chosen idiom, under a value signal. This is Boden's exploratory creativity, the disciplined traversal of a structured conceptual space (Boden 2004), and Wiggins makes the mechanics explicit, a creative system being a set of rules

defining a space, a strategy for traversing it, and a means of evaluating what it finds (Wiggins 2006). A creative process induces a distribution q over S concentrated on and near V ; compression supplies its support and navigation shapes its mass towards value.

The two operations close into a single loop:

```
repeat:
  z ~ prior(.)           # explore: sample a candidate from the prior
  x ← decode(z)          # decode: z indexes a point on the manifold M
  v ← value(x)           # evaluate: score x by a value signal
  keep x if v is high    # navigate: retain and build on what is valued
```

The loop is variation and selection, the shape Campbell (1960) proposed as blind variation and selective retention and Simonton (2011) has refined since, with the qualification that human variation is rarely wholly blind, being shaped by expertise, as machine sampling also is. The loop makes explicit that creativity needs a value signal to bend the walk towards what a community will prize. Who supplies that signal is the crux of the human–machine comparison, and of authorship.

One more property matters. The prior is never one's own. M is learned from what one has seen, and no one has seen nothing, so every creator navigates with a map drawn from a lifetime of others' work, creation being distributed across people, artefacts and institutions rather than lodged in a lone head (Csikszentmihalyi 1996; Glăveanu 2014). That both creators inherit collective priors frames the fairness question about machine training without answering it, since human enculturation is slow, reciprocal and situated while corpus harvesting is unilateral, total and capitalised.

The shared ground: two implementations of one loop

This section claims that the human brain and a modern generative model are two physical realisations of the same generate-and-select loop (Fig. 2), a comparison that animates the programme of neuroscience-inspired artificial intelligence (Hassabis et al. 2017).

Start with the brain, and with a caveat, since the neuroscience of creativity is more contested than popular accounts admit. A widely cited framework pairs the default mode network, active during internally directed and spontaneous thought, with the frontoparietal control systems that evaluate candidates against goals, and associates highly creative performance with unusual cooperation between the two (Beaty et al. 2016, 2018). You are walking through a supermarket, thinking of nothing in particular, and an idea arrives that surprises you. No muse has flown in. On this account your default network has been doing what it does whenever attention is unoccupied, recombining fragments of stored experience under loose constraints and with a good deal of randomness (Christoff et al. 2016), and every so often a combination catches the control system's attention, which finds it interesting and brings it into deliberate thought. Divergent thinking is the generative side of that exchange, and the felt moment of insight is the evaluative side waking up to something the generator produced (Beaty et al. 2016). The idea came from your history and accumulated experience, recombined without deliberate direction and then selected.

Calibration is essential. Stimulation of default-network nodes during awake neurosurgery reduced the fluency of creative responses but not their originality (Shofty et al. 2022), meta-analyses find heterogeneous recruitment and no single circuit (Dietrich and Kanso 2010; Kuang et al. 2022) and a lesion-network study maps creativity onto a distributed circuit rather than a clean pair of networks (Kutsche et al. 2025). We therefore treat the two-network model as a leading hypothesis, not settled fact. The argument needs only that human creativity involves a generative, associative component and an evaluative, selective one.

Now the machine. A generative model trained on a vast corpus learns a compressed representation of its regularities. A variational autoencoder funnels each input through a bottleneck of a few hundred numbers and reconstructs it, so that only the features that matter survive, and its decoder, fed codes from the learned prior, produces new and coherent artefacts. Adversarial networks reached the same capacity by mapping noise straight onto images. Diffusion models walk a random point of S step by step onto M , and analysis of convolutional diffusion shows their inventiveness arising from local, patch-wise recombination of learned structure, not from retrieval of stored pictures (Kamb and Ganguli 2025). Language models do the same token by token. In every case learned compression supplies candidates and stochastic sampling supplies variation.

One commonality runs deeper than the loop's shape, at the level of the learning rule. Hebb's postulate holds that a connection strengthens when one neuron repeatedly helps to fire another, so that what a brain learns is held in the adjustable efficacy of its synapses rather than in any single cell (Hebb 1949). In the abstract, that is what an artificial network does when training adjusts the weights on each unit's inputs. The correspondence is a principle, not a mechanism. Biological neurons spike, are bathed in neuromodulators and keep changing throughout life, whereas a deployed model's weights are frozen, and whether cortex performs anything like the gradient-based credit assignment that trains artificial networks remains unresolved (Richards et al. 2019). What the two share is the connectionist principle that knowledge lives in adjustable connection strengths, not Hebb's particular correlational rule, which backpropagation does not implement. The shared principle carries a consequence. A system that learns by tuning connection strengths acquires a competence spread across those strengths rather than a store of its inputs. What either network holds after learning, at the level of skill, is an ability and not an archive, painting and writing and singing lodged parametrically in the connections in both cases. The human also carries an indexed store of particular episodes, a point taken up under context below.

Selection comes in layers. During training the objective function is the value signal that carves M out of S . After training, systems are refined by reinforcement learning from human feedback, in which human judgements train a reward model that tunes the generator (Christiano et al. 2017; Ouyang et al. 2022), selection by human taste, industrialised. In deployed systems that signal is largely amortised into the weights, so that at generation time the model simply samples, whereas explicit candidate-by-candidate evaluation, the human's online regime, lives in best-of- n sampling, guidance and tree search. Either way the machine's signal is a proxy, only as good as the judgement it stands in for.

At this level of description the loops match. Neither creates from nothing, both recombine internalised structure, both use randomness under constraint, and both keep their skills in connection strengths. The intersection is deep, and it is why the two creators are less different than is commonly supposed. But everything the schema omits, how each came to its prior, what each can touch while creating, what each can do, where each gets its values, is where the creators part ways.

The divergence: how each creator comes to its prior

Human and model arrive at the creative moment carrying compressed models of the possible, acquired in ways that differ in more than detail.

The human prior is grown, not trained. It accumulates through a childhood and a life in which every datum arrives through a body in a situation, attached to hunger or comfort, to a caregiver's face, to consequences. By early adolescence a child has encountered perhaps a hundred million words; the models she will grow up alongside were trained on trillions, a gap of four to five orders of magnitude (Frank 2023). A child generalises from so little because human learning does not start from zero, evolution having compressed aeons of selection into genomic priors that wire much of the needed structure before experience begins (Zador 2019), and because the learning is interactive. The child acts, the world answers, and the answer has stakes.

The machine's prior is trained, not grown. It starts from random weights plus engineered inductive biases, an architecture, a tokenisation, an objective, the designed counterpart of genomic priors, and acquires everything else from a frozen snapshot of tokenised culture, traversed by gradient descent, without acting on anything or risking anything. What the machine has instead of structure and stakes is scale. A well-known position statement from within the field, that general methods riding computation have beaten engineered human knowledge (Sutton 2019), describes how different this route is.

Scale needs stating plainly. Loss falls predictably as parameters and data grow (Hoffmann et al. 2022), and in our terms that widens the parametric store from which combinations can be drawn, which matters because remote combination, the yoking of concepts from distant regions of the manifold, presupposes that both regions are represented at all. A network cannot combine what it has not internalised. Scale is therefore a real condition of machine creativity, though not a sufficient one, and it brings costs the article returns to, in concentration and energy, and a side effect, since memorisation grows with model size as well as with duplication (Carlini et al. 2021).

The two priors have different textures, the model's wide and shallow, every century and idiom at once, each element bound to nothing but its statistical neighbours, the human's narrow and deep, a minuscule sample of the world's culture with each piece soaked in circumstance and memory. They are different maps by different cartographers, and this is the first thing "alternative" means.

The divergence: what world each creator can draw on

The second difference concerns what each creator can touch at the moment of creation, and we take it to be the decisive one.

Both creators hold their skills parametrically, in the strengths of their connections, and the question is what was written into the parameters. In the human case it was a body acting in a physical world among other people, over years and with consequences. In the machine's case it was, for the most part, the internet. When a model generates, its available world is that compressed archive plus whatever a prompt supplies, the statistical residue of testimony about the world. However vast, the internet does not contain the world, only what has been said and shown of it, unevenly sampled, and a model trained on that record inherits its reach,

which is superhuman, and its distance from the events, which is total. A million photographs of situations are not a situation.

A human creator arrives carrying what no archive holds, an ongoing first-person transaction with reality. Perception in real time, a body with states, a biography, particular others present as relationships and not as data, a death somewhere ahead. Imagination is built from exactly this material, since the constructive-memory system that recombines autobiographical particulars into simulated futures is, on the standard account, the same system that remembers the lived past (Schacter and Addis 2007). Dewey placed the origin of art in exactly this exchange between the live creature and its environment (Dewey 1934), embodied cognition holds that mind is constituted through such engagement (Varela et al. 1991), and sociocultural research locates the creative act between people (Glăveanu 2014). Some information exists only at the interface of a body and a situation, what this circumstance is like, now, for this person, and it never enters the archive because most of what is lived is never uttered. A model can render grief in any register the archive contains. It has never lost anyone, and no scaling supplies what was never testimony. The point cuts the machine's way too, since no human has read the archive and a model speaks from the aggregated testimony of a civilisation. One creator works from everyone's records and no one's situation, the other from one situation and a sliver of the records. Neither window contains the other.

The difference surfaces in the artefacts as indexicality, Peirce's term for signs bound to their causes. Much of the value of human works is that they point back to lived particulars. This photographer stood in that street while the shells fell; this song was written by the person who buried the friend. The machine's outputs have a causal history too, but a diffuse one; they point back to everyone's testimony and to no one's morning. Audiences appear to sense and discount the difference, since identical artworks are rated lower when labelled as made by AI (Bellaiche et al. 2023). The penalty may fade with convention, and essentialist thinking about originals biases such judgements generally (Newman and Bloom 2012), but whether a witness stands behind the work is not a matter of convention. Whether that makes artistic and not merely documentary value was already disputed over photography, Scruton doubting the photograph could be art because its bond to its subject is causal rather than intentional (Scruton 1981), Barthes locating its power in exactly that bond, the certification that the thing has been (Barthes 1981), and Walton treating the photograph as transparent, a way of seeing the object itself (Walton 1984). We side with Barthes and Walton. Indexicality is a durable human resource, though not an essence, since much human art is cheerfully non-indexical.

The divergence: two envelopes of the possible

The third difference is the least discussed and the most visible. The two creators can do different things. Machines find hard what we find effortless and the reverse (Moravec 1988), and the observation extends into creation.

Consider a video model. Asked for a scene, it synthesises every frame of a photorealistic clip, coherent in lighting, texture and motion for seconds to a minute, with longer-range coherence still an open problem (the mechanism in Ho et al. 2022, the photorealism in Brooks et al. 2024). This is a hyperrealist painter working at twenty-four canvases per second without fatigue. No human hand has ever done this; we built cameras because reality outruns the hand, and photography captures what we cannot paint. The generative model

inverts that settlement and paints what we could only photograph, camera-grade records of events that never occurred, a capability arriving inside practices organised for centuries around its absence.

Now the other direction. A human creator can want to make something. She can conceive a project no one asked for, hold it across a decade and let it reorganise her life, a self-generated, long-horizon intention of which current systems have at most the outward form, executing given goals instead of originating their own. She can evaluate against felt experience, discarding a technically perfect line because it rings false against a life, which a system whose evaluations proxy other people's judgements cannot do. Her skills are knowledge, the educated hand knowing things about paint and resistance that no description of paintings contains, which is why craft stays a serious category for this collection. And she can change her own values, since taste develops through living and revises the selecting function itself.

Set the two lists side by side (Fig. 3). Neither envelope contains the other. The machine is superhuman along axes where no human competes and incapable along axes where every amateur functions effortlessly, and no single scale ranks the two profiles, any more than one scale lets a falcon out-fly an aircraft. That unrankability, not sentimentality about our self-image, is the empirical core of the claim that machine creativity is alternative rather than lesser or greater.

Alternative creativity

We can now state the position.

Call two creative systems alternative to one another when they implement the same abstract process, a learned compression of a shared possibility space navigated by value-guided stochastic search, but differ constitutively in four respects. How the prior is acquired. What context is accessible at the moment of creation. How the value signal is grounded, live and first-person or frozen and aggregated. What envelope of operations is feasible. By the preceding argument, human creators and generative models satisfy both clauses. They are not the same creativity at different strengths, because the four differences are differences of kind, and not unrelated phenomena sharing a name, since the process really is common, which is why the machines work at all. They are two constitutions of one process, alternative creativities traversing a shared space by different paths.

The four axes are not equally weighty, and saying which weigh most is part of the position, so the claim needs care. By substrate we mean the narrow thing, spiking neurons against continuous units, synaptic plasticity against a learning rule. The axes are entangled, since frozen weights are one reason a deployed model cannot revise its own taste and a body is one reason a person has situations rather than testimony, so no clean partition is available. Our claim is comparative and conjectural. Hold the substrate fixed and vary the situation, giving two identical networks a lived world in one case and an archive in the other, and you would still get two creativities, because what was written into the parameters and what value steers the search would differ. Hold the situation fixed and vary the substrate, and creative behaviour converges. On that evidence the second and third axes explain more of the divergence in what gets made than the first, which is the sense in which the difference between us and the machines is less about what we are made of than about where we have been.

The position has neighbours in computational creativity (Colton and Wiggins 2012), in Halina's (2021) treatment of insight, in Lopes's (2009) account of computer art as a distinct appreciative kind, and in posthumanist media theory (Zylinska 2020). Our contribution is the machinery, not the mood, the four-axis analysis with its ranking and the consequences derived below.

In Boden's terms, current systems are combinational and exploratory creators of extraordinary power (Boden 2004; Ritchie 2007; Franceschelli and Musolesi 2025), with divergent-thinking performance at or above typical human levels in controlled comparisons (Hubert et al. 2024). Transformational creativity, the changing of the space's own rules, is where they remain weakest, and the framework says why. A system whose value signal is defined relative to its training distribution has little purchase on moves whose value lies in breaking that distribution; such moves are off-distribution by construction. The point is scoped to preference-trained systems, since a self-play system with an exact, non-human value signal broke the human distribution of Go play precisely because its reward owed nothing to it. Eminent, field-changing achievement, Kaufman and Beghetto's (2009) Big-C, a matter of recognition, not of Boden's structural categories, remains human as matters stand in the arts.

The thirty-seventh move of AlphaGo's second 2016 game against Lee Sedol, initially read by professional commentators as a mistake, was a novel, valuable and surprising contribution to a practice three millennia deep (the system in Silver et al. 2016, its self-play successor in Silver et al. 2017), and careful analysis defends it as insight in a behavioural sense while denying it the domain-general understanding human insight carries (Halina 2021). It presses on the Big-C boundary within its bounded domain, alternative creativity in its clearest form, an outcome the community values reached by a path no human takes. The calibration matters as much as the case. Where the value function is exact, the machine can navigate beyond us; where value is tacit and social, as in most of the arts, it must borrow its signal and its autonomy contracts.

The machine's most distinctively alternative capability grows out of this asymmetry, and it goes beyond sampling a fixed manifold. Reinforcement learning has always turned on the balance between exploration and exploitation (Sutton and Barto 2018), and a family of methods pushes exploration further by rewarding novelty and diversity alongside, or instead of, a fixed objective, since a search for what is new reaches solutions objective-chasing never would (Lehman and Stanley 2011; Pugh et al. 2016). Open-endedness, the continual production of novel and learnable artefacts, is argued to be the route to machine discovery (Hughes et al. 2024).

Coupled to an automatic verifier, this generate-and-check loop has already produced knowledge no human held. A language model inside a program-search loop found mathematical constructions that beat a record standing since 2004 (Romera-Paredes et al. 2024), and models trained by reinforcement against checkable answers, without supervised demonstrations of reasoning though on a base pretrained on human text, acquire strategies no one showed them and trace their own path (Guo et al. 2025). In Boden's strict sense these remain exploratory moves, since the rules of the space, the verifier, were fixed by us, but they are exploratory moves that leave the human distribution behind, and they are feasible exactly where success can be checked without us. Transformational creativity proper, the rewriting of the rules themselves, stays open. In mathematics and code the verifier is cheap. In the arts no oracle scores a canvas, and the only verifier available is a learned model of human taste, which tethers the search back towards what people already like.

The same asymmetry therefore fixes where a machine can leave the human distribution on its own and where it still needs us (Fig. 4). And the generate-and-evaluate architecture that, on the leading account, the brain uses when an idea surfaces in a supermarket is the architecture these systems are converging on. We are cloning the strategy, whether or not we are cloning the substrate.

The most serious philosophical objection comes from agency theories. On Gaut's account creativity is an exercise of agency involving flair, production neither by mechanical procedure nor by chance (Gaut 2010), and a system harvesting constrained randomness under a proxy signal can look like the intended contrast case. Two replies. First, the behavioural definition we adopted does not build agency into creativity, and the psychological literature has managed without it; if Gaut's condition is right the machines lack creativity by definition, and we take that to be the point in dispute, not a premise. Second, following Halina's move for insight, value-guided stochastic search under learned constraints is neither mechanical procedure nor chance. It is the very middle ground the agency theorist reserves for flair, minus the phenomenology, and whether phenomenology is necessary is the question at issue.

The remaining sceptical arguments target a different overclaim. The stochastic-parrot critique (Bender et al. 2021) and its descendants (Browning 2024) succeed against the claim that these systems understand as we do, a claim we do not make, and the understanding debate remains open (Mitchell and Krakauer 2023; Havlík 2025), which creativity, behaviourally defined, need not await. Nor does the account compete with institutional theories on which artefacts become art through conferred status (Dickie 1974) or with Danto's indiscernibles (Danto 1964, 1981), since compression and navigation describe how candidates are generated and selected, not how they come to count as art. What the framework insists on is the asymmetry of value, a proxy for the model and, for the person, something bound up with meaning, motivation and a life.

The asymmetry invites an objection. If the machine borrows its value signal wholesale, is it not an instrument after all, a magnificent brush wielded by distributed human taste? Where the value function is exact and internal, as in self-play, the objection has no grip. In the arts we half-concede it, since the machine's evaluation is derivative. But a brush proposes nothing. A system that has compressed the archive contributes candidates nobody specified, from regions nobody chose, and generativity of that order is not transmission. The right description is a creator autonomous on the generative axis and dependent on the evaluative one, the asymmetric constitution the term was coined to name.

Naming the creator also means drawing its boundary honestly. What produces an output is a sociotechnical assemblage of corpus authors who were not asked, subcontracted raters whose piece-paid judgements froze into the reward model, engineers, the prompter and the owning firm. Sociology has long shown that "the artist" is an attribution made inside such cooperative webs (Becker 1982), and the history of automation reads as the capture and codification of collective labour long before machine learning (Pasquinelli 2023). By "the machine" we mean that deployed system, its constitution including those human contributions in compressed and frozen form, which places the archive's authors and the raters inside the creator, with distributional consequences drawn below. The constitution is material as well; the tirelessness we praised is also a purchase of megawatts, water and annotation shifts (Crawford 2021).

Two corollaries: homogenisation, and the limits of attribution

Two consequences follow that were not among the assumptions.

The first concerns diversity. If models sample preferentially from the dense, high-probability regions of the learned manifold, their outputs should cluster more tightly than those of a comparably productive human population, whose idiosyncratic priors scatter across M . Berlyne adds a second pressure, since hedonic value peaks at moderate novelty (Berlyne 1971), so a system optimised on human approval is pulled towards the comfortable middle. Machine assistance should raise the quality and novelty of individual works while contracting the diversity of the pool (Fig. 5), and the evidence is consistent with this. Individual stories became better and more novel, especially for less skilled writers, while the population of stories grew more alike (Doshi and Hauser 2024), with convergent evidence from a different design (Moon et al. 2025). The culture-industry tradition described standardisation wearing the mask of individuality long before there were models to blame (Adorno 1941). What the framework adds is a mechanism internal to the technology, to be kept distinct from a structural one, because a few foundation models trained on overlapping corpora and aligned under similar preference regimes would homogenise output even if sampling did not (Kleinberg and Raghavan 2021). The two causes separate empirically, since the first predicts stronger contraction under preference tuning than under supervised fine-tuning within a system, which is what has been reported (Kirk et al. 2024), while the second predicts contraction tracking concentration across systems. It also locates a durable human advantage in the high-novelty tail.

The second corollary concerns attribution, the crux of the "theft" debate. For the great majority of outputs, tracing a generated artefact back to specific training examples has no determinate answer, because the model stores not its training artefacts but a lossy, distributed encoding of their shared regularities. Legal and technical scholarship has converged on the point (Sag 2023; Henderson et al. 2023; Cooper and Grimmelmann 2025), against the widespread picture of these models as elaborate memories. Given enough data, two diffusion models trained on disjoint image sets converge on almost the same denoising function, which is possible only if each has learned the structure of the image class rather than its particular examples (Kadkhodaie et al. 2024). The system is learning to paint, internalising the regularities of pictures rather than holding them, and the contrary case, where it does hold them, is overfitting, a failure to generalise that training is regularised to suppress rather than an aim. Memorisation is nonetheless real. Works duplicated many times in the training data are reproduced closely, and near-verbatim training data have been extracted from deployed systems (Carlini et al. 2021, 2023); it is a failure of compression on over-represented particulars, measurable and substantially addressable by engineering (Somepalli et al. 2023).

The opacity of the harvest is partly chosen, since provenance documentation was proposed early (Gebru et al. 2021) and largely not adopted; had it been, attribution of outputs would remain ill-posed but the accounting of who was ingested would not, and training-data attribution methods (Koh and Liang 2017, and a growing literature since) yield approximations for auditing, not provenance for artefacts. Courts have been converging on the distinction between memorisation and learning from several directions. Training on lawfully acquired books was held a transformative fair use while retaining pirated copies was not (Bartz v Anthropic 2025); a parallel ruling granted fair use while flagging market dilution as the argument that may decide other cases (Kadrey v Meta 2025); the earliest artists' suit continues on claims about specific reproductions (Andersen v Stability AI 2023); an English court held that a model which stores no works is not an infringing copy, leaving training itself undecided and an appeal pending (Getty Images v Stability AI 2025); and a Munich court held

memorised lyrics recoverable from a model to be reproductions, in a judgment under appeal (GEMA v OpenAI 2025). Doctrine has thus separated memorisation, which our analysis disciplines, from market substitution (US Copyright Office 2025b), which it leaves untouched, and both converge with work in this journal urging governance to judge observable outputs by their recognisable linkage to prior works (Fischman-Afori 2026).

Correcting the retrieval picture does not dissolve the artists' grievance. The manifold is learned from the expression of enormous numbers of people, unconsulted, under a consent architecture that defaults to inclusion and quietly converts "publicly available" into "available to be capitalised" (Bender et al. 2021; Couldry and Mejias 2019; Crawford 2021; Keller 2023; Srnicek 2017). The labour concern runs in both directions, since outputs can displace creators in aggregate whether or not any single work was copied, with early market evidence of demand shifts for writing and illustration freelancers (Hui et al. 2024), while labour is also inside the systems, in the ghost work of the raters described above (Gray and Suri 2019). What the technical account contributes is discipline about remedies. If influence is diffuse and collective, compensation will be collective too, and opt-out and transparency regimes, collective bargaining of the kind screenwriters secured in 2023, and data-trust proposals already sketch the architecture. Two premises should govern its design, that responsibility lies with the firms that harvested the commons, and that blanket payouts tend to reach large rights-holders, not individual illustrators. If corpus authors and raters are constituents of the creator rather than external sources, the fairness claim does not weaken when the copying story fails; it changes register, from theft to uncompensated participation.

What this means for art and design

If the creative act is not the laying-down of marks but the compression and navigation of a space towards value, authorship lives wherever those operations are controlled. When a person prompts, curates, rejects, constrains and iterates towards a result they judge right, they perform the navigation and supply the value signal, though a decoder places the marks. Authorship migrates from the hand to the specification of value.

The migration has happened before, and art theorised it first. The readymade located authorship in selection, and LeWitt's instruction works made the idea "a machine that makes the art" (LeWitt 1967). Photography, declared incapable of art because the apparatus made the image, became an art of framing, timing and selection, exactly the navigational layer (Le 2025; Ogundipe 2025), and the synthesiser and the sampler were resisted as inauthentic and then absorbed as instruments (Lee 2024). Each time, automating mark-making pushed authorship up a level, towards selection and intention, and each time the deskilling below was real and painful, a cost the word "redistribution" should not launder. The move now under way carries one novelty, that the automated layer includes competent candidate generation as well as execution, so the navigational remainder, what to ask for, what to keep, what it is for, becomes the whole human contribution.

Authorship disputes thereby become questions of degree. A single word yielding a first sample is thin authorship. A sustained process of constraint, rejection and refinement towards a realised intention is thick, as scholarship on the originality of prompts and outputs has begun to formalise (Mazzi 2024; Hertzmann 2018; Epstein et al. 2023; Lopes 2026). Locating authorship in intention and selection takes a side against the tradition that pronounced the author dead (Barthes 1977), deliberately, because when a person supplies the value signal that bends the search, intention does recoverable work in the artefact. Authorship is also an

institutional status, conferred and policed (Foucault 1979), and the conferring institutions are resisting, the US Copyright Office having repeatedly declined to treat prompting alone as authorship (US Copyright Office 2025a). Our claim is therefore partly predictive, an account of a negotiation in progress and a stratified one, since who may claim thick authorship will track existing hierarchies of recognition and consecration in the field (Bourdieu 1993).

The constitutional analysis also yields what the authorship debate usually lacks, a principled account of collaboration. Co-creation is a coupling of complements, not two rivals sharing one job. The machine contributes exploratory breadth, tireless variation, the averaged fluency of the archive and access to regions of the manifold no tradition visited. The human contributes the situated stakes that select, the indexical witness that anchors, the taste that can refuse the comfortable middle, and the origination of the goal itself (Epstein et al. 2023). Across tasks, human-machine teams on average underperform the better of the two, they lose most where the machine alone already outperforms the human, and the gains concentrate in open-ended content creation (Vaccaro et al. 2024), a pattern consistent with the framework, since content creation is where the human supplies a value signal the machine cannot verify for itself. This also reconciles the meta-analysis with the divergent-thinking scores cited earlier, since scoring high on a psychometric test is not making work a community values. Complementarity is conditional, appearing where the two constitutions supply different things, and creative work, needing both tireless generation and grounded taste, is its heartland.

Chess offers a cautionary comparison. When a freestyle tournament in 2005 admitted human-computer teams, the winners were two amateurs with ordinary machines and a better process for combining them, ahead of grandmasters with computers and of engines alone, the configuration Kasparov later called the centaur (Kasparov 2017). The advantage has since narrowed as the engines' exact evaluation improved, which is our asymmetry at work. The arts sit on the other side of it, since no evaluator is coming that scores a canvas without us. The lesson for the studio is that a second creator with a different constitution has joined the team, one that will surprise us in some directions and fall short in others, for reasons of context more than of architecture. A blanket refusal to work with it forfeits capabilities that exist nowhere else, though refusal can also be a reasoned collective stance, as the screenwriters' bargaining shows, and work made without it will keep its place. The choice is a decision about value, which is our point, and where the two constitutions are combined with intent they can reach results neither could reach before.

The pair meets across a market, not in a void. Reinforcement learning from human feedback is taste institutionalised, an averaged value signal useful where individual value is absent, and the working artist's task becomes supplying the un-averaged remainder. One complement is rented from a concentrated industry that sets the terms, logs the interaction and may train on it, so the optimistic division of labour is a claim about capabilities, not about bargaining power (Srnicek 2017). For design practice and its classrooms, the analysis relocates the centre of gravity towards brief-writing, curation and the articulation of constraints and values, and towards training students to originate and defend value functions rather than only to execute against given ones.

One human advantage survives even perceptual parity of outputs. A generated image indistinguishable from a war photograph is not the photograph's equivalent, because part of what we value in the photograph is that

someone stood there. The causal history is part of the artefact's meaning, the intuition Benjamin named aura, even as he welcomed its decline (Benjamin 2008), and Dutton found in our verdicts on forgery (Dutton 1979). This holds for documentary value, where witness is evidence, and for artistic value, where achievement is individuated by its history. Live performance, handmade work and witnessed testimony will retain worth in an age of perceptual abundance because they certify a connection to a life, the one thing the alternative creator cannot counterfeit, only imitate.

A moving frontier

The divergence half of this article describes constitutions, not essences, and constitutions change. Embodied robotics pursues agents that acquire their priors through sensorimotor interaction, reviving the insistence that intelligence needs a world and not a description of one (Brooks 1991), and a prominent statement of the field's direction declares the era of static human data ending, to be replaced by agents that learn continually from their own experience (Silver and Sutton 2025). If those programmes succeed, the divergences narrow in order, the context gap first, since a robot that spends years in kitchens begins to have situations rather than testimony about them, the ontogeny gap more slowly, because deployment with sensors is still not development under consequence, and the value gap last. Felt value is anchored in needs, finitude and sociality, and whether an artificial agent could originate rather than borrow a value signal remains open. On our own analysis, the route to closing that gap in the arts is not more archive but more world. A creator whose taste we cannot verify against an oracle can only acquire it the way we did, by living among the people whose taste it is. Nothing here guarantees permanent human monopolies. "Alternative" is indexed to present constitutions, a finding, not a fence, to be revisited as they change.

Conclusion

The question the culture keeps asking, whether AI is creative, yes or no, cannot be answered, because it presupposes a single creativity of which the machine has some quantity. Within any fixed format the space of possible configurations is finite, vast and mostly worthless, and it already contains every work that has been made and every work that ever will be. Creativity, wherever it occurs, is the compression of that space into a manifold of meaningful candidates and the value-guided navigation of the manifold. By the behavioural standards creativity research applies to people, generative systems do this, and making from prior experience is all human creators ever did. The two networks that do this are less different than we like to think, learning by adjusting connections, holding their skills in those connections, generating by constrained randomness and selecting by value.

But the sharer is not a twin. It comes to its parameters by training, not living, arrives carrying an archive, not a situation, operates an envelope that overlaps ours without containing it, and steers by values it borrows where ours are felt. Of these differences, the one of world explains more of what gets made than the one of substrate. Two constitutions, one process. Alternative creativity.

The framing pays for itself in consequences. It gives homogenisation a mechanism, locates the durable human advantage in the novelty tail and the indexical anchor, explains why tracing outputs to inputs cannot bear the weight the theft debate places on it and why the models are learning to paint rather than storing paintings, and gives authorship and collaboration a logic. A second navigator has entered an old and finite

space, moving by paths we cannot take and steering, for now, by values it borrows from us. What its arrival is worth is a question about value, and on the evidence assembled here that question remains ours to answer.

Methods

This article is a work of conceptual synthesis and theoretical argument and reports no new empirical data, human-participant research or experiments. The method is the construction and defence of a framework by comparative integration of five literatures (creativity research, cognitive neuroscience, generative machine learning, embodied and sociocultural theories of cognition, and the philosophy and law of art), together with the derivation of testable consequences. Claims attributed to empirical sources were checked against the primary publications, and where evidence is contested the text flags it and the argument is built not to depend on it.

Use of AI tools. Large language models were used during preparation as an aid to literature search, drafting and language refinement. The thesis, the framework, the interpretations and the conclusions originate with the human authors, who verified every citation and factual claim and take full responsibility for the text. No AI system is an author, and no generative-AI images are included.

Declarations

Competing interests

The authors declare no competing interests.

Author Contribution

J.L.-P. conceived the study, developed the theoretical framework, conducted the literature review, prepared Figures 1–5 and wrote the manuscript. C.A.-T., I.F.-M. J.I.-D, and J.T.-A. critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements

Redacted for double-anonymised peer review; the full text is provided in the cover letter.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

References

1. Adorno TW (1941) On popular music (with the assistance of Simpson G). *Stud Philos Soc Sci* 9(1):17–48

2. Andersen v Stability AI Ltd (2023) Case No 3:23-cv-00201-WHO, US District Court for the Northern District of California. CourtListener. <https://www.courtlistener.com/docket/66732129/andersen-v-stability-ai-ltd/>. Accessed 5 Sep 2026
3. Barthes R (1977) *The death of the author*. Image-music-text (trans: Heath S). Fontana, London, pp 142–148
4. Barthes R (1981) *Camera lucida: reflections on photography* (trans: Howard R). Hill and Wang, New York
5. Bartz v Anthropic PBC (2025) Order on fair use, No 3:24-cv-05417-WHA (Alsup J), 23 June 2025, Dkt 231. US District Court for the Northern District of California. CourtListener. <https://storage.courtlistener.com/recap/gov.uscourts.cand.434709/gov.uscourts.cand.434709.231.0.pdf>. Accessed 5 Sep 2026
6. Beaty RE, Benedek M, Silvia PJ, Schacter DL (2016) Creative cognition and brain network dynamics. *Trends Cogn Sci* 20(2):87–95. <https://doi.org/10.1016/j.tics.2015.10.004>
7. Beaty RE, Kenett YN, Christensen AP, Rosenberg MD, Benedek M, Chen Q, Fink A, Qiu J, Kwapil TR, Kane MJ, Silvia PJ (2018) Robust prediction of individual creative ability from brain functional connectivity. *Proc Natl Acad Sci USA* 115(5):1087–1092. <https://doi.org/10.1073/pnas.1713532115>
8. Becker HS (1982) *Art worlds*. University of California Press, Berkeley
9. Bellaïche L, Shahi R, Turpin MH, Ragnhildstveit A, Sprockett S, Barr N, Christensen AP, Seli P (2023) Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. *Cogn Res Princ Implic* 8:42. <https://doi.org/10.1186/s41235-023-00499-6>
10. Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: can language models be too big? In: *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. ACM, New York, p 610–623. <https://doi.org/10.1145/3442188.3445922>
11. Benjamin W (2008) *The work of art in the age of its technological reproducibility, and other writings on media*. Belknap Press of Harvard University, Cambridge, MA
12. Berlyne DE (1971) *Aesthetics and psychobiology*. Appleton-Century-Crofts, New York
13. Boden MA (2004) *The creative mind: myths and mechanisms*, 2nd edn. Routledge, London
14. Borel É (1913) La mécanique statique et l'irréversibilité. *J Phys Théor Appl* 3(1):189–196. <https://doi.org/10.1051/jphystap:019130030018900>
15. Bourdieu P (1993) In: Johnson R (ed) *The field of cultural production: essays on art and literature*. Polity, Cambridge
16. Brooks RA (1991) Intelligence without representation. *Artif Intell* 47(1–3):139–159. [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M)
17. Brooks T, Peebles B, Holmes C, DePue W, Guo Y, Jing L, Schnurr D, Taylor J, Luhman T, Luhman E, Ng C, Wang R, Ramesh A (2024) Video generation models as world simulators. OpenAI. <https://openai.com/index/video-generation-models-as-world-simulators/>. Accessed 5 Sep 2026
18. Browning J (2024) Personhood and AI: why large language models don't understand us. *AI Soc* 39(5):2499–2506. <https://doi.org/10.1007/s00146-023-01724-y>
19. Calvo P (2026) The synthetic transformation of the art world: ethical tensions. *Humanit Soc Sci Commun* 13:963. <https://doi.org/10.1057/s41599-026-06984-9>

20. Campbell DT (1960) Blind variation and selective retention in creative thought as in other knowledge processes. *Psychol Rev* 67(6):380–400. <https://doi.org/10.1037/h0040373>
21. Carlini N, Hayes J, Nasr M, Jagielski M, Sehwag V, Tramèr F, Balle B, Ippolito D, Wallace E (2023) Extracting training data from diffusion models. In: *Proceedings of the 32nd USENIX security symposium*. USENIX Association, Berkeley, CA, p 5253–5270
22. Carlini N, Tramèr F, Wallace E, Jagielski M, Herbert-Voss A, Lee K, Roberts A, Brown T, Song D, Erlingsson Ú, Oprea A, Raffel C (2021) Extracting training data from large language models. In: *Proceedings of the 30th USENIX security symposium*. USENIX Association, Berkeley, CA, p 2633–2650
23. Chiang T (2023) ChatGPT is a blurry JPEG of the web. *The New Yorker*, 9 February. <https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web>. Accessed 5 Sep 2026
24. Christiano PF, Leike J, Brown TB, Martic M, Legg S, Amodei D (2017) Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, vol 30. Curran Associates, Red Hook, NY, pp 4299–4307
25. Christoff K, Irving ZC, Fox KCR, Spreng RN, Andrews-Hanna JR (2016) Mind-wandering as spontaneous thought: a dynamic framework. *Nat Rev Neurosci* 17(11):718–731. <https://doi.org/10.1038/nrn.2016.113>
26. Colton S, Wiggins GA (2012) Computational creativity: the final frontier? In: De Raedt L et al (eds) *ECAI 2012. Frontiers in artificial intelligence and applications*, vol 242. IOS, Amsterdam, pp 21–26. <https://doi.org/10.3233/978-1-61499-098-7-21>
27. Cooper AF, Grimmelmann J (2025) The files are in the computer: on copyright, memorization, and generative AI. *Chicago-Kent Law Rev* 100(1):141–219
28. Couldry N, Mejias UA (2019) *The costs of connection: how data is colonizing human life and appropriating it for capitalism*. Stanford University Press, Stanford
29. Crawford K (2021) *Atlas of AI: power, politics, and the planetary costs of artificial intelligence*. Yale University Press, New Haven
30. Csikszentmihalyi M (1996) *Creativity: flow and the psychology of discovery and invention*. HarperCollins, New York
31. Danto A (1964) The artworld. *J Philos* 61(19):571–584. <https://doi.org/10.2307/2022937>
32. Danto A (1981) *The transfiguration of the commonplace: a philosophy of art*. Harvard University Press, Cambridge, MA
33. Dewey J (1934) *Art as experience*. Minton Balch, New York
34. Dickie G (1974) *Art and the aesthetic: an institutional analysis*. Cornell University Press, Ithaca
35. Dietrich A, Kanso R (2010) A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychol Bull* 136(5):822–848. <https://doi.org/10.1037/a0019749>
36. Doshi AR, Hauser OP (2024) Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Sci Adv* 10(28):eadn5290. <https://doi.org/10.1126/sciadv.adn5290>
37. Dutton D (1979) Artistic crimes: the problem of forgery in the arts. *Br J Aesthet* 19(4):302–314. <https://doi.org/10.1093/bjaesthetics/19.4.302>

38. Epstein Z, Hertzmann A, Akten M, Farid H, Fjeld J, Frank MR, Groh M, Herman L, Leach N, Mahari R, Pentland AS, Russakovsky O, Schroeder H, Smith A (2023) Art and the science of generative AI. *Science* 380(6650):1110–1111. <https://doi.org/10.1126/science.adh4451>
39. Fischman-Afori O (2026) Generative AI art and copyright governance: rethinking the adaptation right and its limits. *Humanit Soc Sci Commun* 13:1469. <https://doi.org/10.1057/s41599-026-07311-y>
40. Foucault M (1979) What is an author? In: Harari JV (ed) *Textual strategies: perspectives in post-structuralist criticism*. Cornell University Press, Ithaca, pp 141–160
41. Franceschelli G, Musolesi M (2025) On the creativity of large language models. *AI Soc* 40:3785–3795. <https://doi.org/10.1007/s00146-024-02127-3>
42. Frank MC (2023) Bridging the data gap between children and large language models. *Trends Cogn Sci* 27(11):990–992. <https://doi.org/10.1016/j.tics.2023.08.007>
43. Gaut B (2010) The philosophy of creativity. *Philos Compass* 5(12):1034–1046. <https://doi.org/10.1111/j.1747-9991.2010.00351.x>
44. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumé H III, Crawford K (2021) Datasheets for datasets. *Commun ACM* 64(12):86–92. <https://doi.org/10.1145/3458723>
45. GEMA v OpenAI (2025) Urteil des Landgerichts München I, Az 42 O 14139/24, 11 November 2025, Pressemitteilung 11/2025. Bayerisches Staatsministerium der Justiz. <https://www.justiz.bayern.de/gerichte-und-behoerden/landgericht/muenchen-1/presse/2025/11.php>. Accessed 5 Sep 2026
46. Getty Images (US) Inc and others v Stability AI Ltd (2025) [2025] EWHC 2863 (Ch), judgment of 4 November 2025 (Joanna Smith J). *Courts and Tribunals Judiciary*. <https://www.judiciary.uk/judgments/getty-images-v-stability-ai/>. Accessed 5 Sep 2026
47. Glăveanu VP (2014) *Distributed creativity: thinking outside the box of the creative individual*. Springer, Cham
48. Goodman N (1976) *Languages of art: an approach to a theory of symbols*, 2nd edn. Hackett, Indianapolis
49. Gray ML, Suri S (2019) *Ghost work: how to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt, Boston
50. Guo D, Yang D, Zhang H et al (2025) DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature* 645(8081):633–638. <https://doi.org/10.1038/s41586-025-09422-z>
51. Halina M (2021) Insightful artificial intelligence. *Mind Lang* 36(2):315–329. <https://doi.org/10.1111/mila.12321>
52. Hassabis D, Kumaran D, Summerfield C, Botvinick M (2017) Neuroscience-inspired artificial intelligence. *Neuron* 95(2):245–258. <https://doi.org/10.1016/j.neuron.2017.06.011>
53. Havlík V (2025) Meaning and understanding in large language models. *Synthese* 205:9. <https://doi.org/10.1007/s11229-024-04878-4>
54. Hebb DO (1949) *The organization of behavior: a neuropsychological theory*. Wiley, New York
55. Henderson P, Li X, Jurafsky D, Hashimoto T, Lemley MA, Liang P (2023) Foundation models and fair use. *J Mach Learn Res* 24(400):1–79
56. Hertzmann A (2018) Can computers create art? *Arts* 7(2):18. <https://doi.org/10.3390/arts7020018>

57. Ho J, Salimans T, Gritsenko A, Chan W, Norouzi M, Fleet DJ (2022) Video diffusion models. *Advances in neural information processing systems*, vol 35. Curran Associates, Red Hook, NY, pp 8633–8646
58. Hoffmann J, Borgeaud S, Mensch A et al (2022) An empirical analysis of compute-optimal large language model training. *Advances in neural information processing systems*, vol 35. Curran Associates, Red Hook, NY, pp 30016–30030
59. Hubert KF, Awa KN, Zabelina DL (2024) The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Sci Rep* 14:3440.
<https://doi.org/10.1038/s41598-024-53303-w>
60. Hughes E, Dennis MD, Parker-Holder J, Behbahani F, Mavalankar A, Shi Y, Schaul T, Rocktäschel T (2024) Position: open-endedness is essential for artificial superhuman intelligence. In: *Proceedings of the 41st international conference on machine learning*. PMLR 235:20597–20616
61. Hui X, Reshef O, Zhou L (2024) The short-term effects of generative artificial intelligence on employment: evidence from an online labor market. *Organ Sci* 35(6):1977–1989.
<https://doi.org/10.1287/orsc.2023.18441>
62. Kadkhodaie Z, Guth F, Simoncelli EP, Mallat S (2024) Generalization in diffusion models arises from geometry-adaptive harmonic representations. In: *Proceedings of the 12th international conference on learning representations (ICLR 2024)*, Vienna, 7–11 May 2024
63. Kadrey v Meta Platforms Inc (2025) Order granting Meta's cross-motion for partial summary judgment, No 3:23-cv-03417-VC (Chhabria J), 25 June 2025, Dkt 598. US District Court for the Northern District of California. CourtListener.
<https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.598.0.pdf>. Accessed 5 Sep 2026
64. Kamb M, Ganguli S (2025) An analytic theory of creativity in convolutional diffusion models. In: *Proceedings of the 42nd international conference on machine learning*. PMLR 267:28795–28831
65. Kant I (2000) In: Guyer P, Matthews E (eds) *Critique of the power of judgment*. Cambridge University Press, Cambridge
66. Kasparov G (2017) *Deep thinking: where machine intelligence ends and human creativity begins*. PublicAffairs, New York
67. Kaufman JC, Beghetto RA (2009) Beyond big and little: the Four-C model of creativity. *Rev Gen Psychol* 13(1):1–12. <https://doi.org/10.1037/a0013688>
68. Keller P (2023) AI, the commons, and the limits of copyright. Open Future Foundation.
<https://openfuture.eu/blog/ai-the-commons-and-the-limits-of-copyright/>. Accessed 5 Sep 2026
69. Kirk R, Mediratta I, Nalmpantis C, Luketina J, Hambro E, Grefenstette E, Raileanu R (2024) Understanding the effects of RLHF on LLM generalisation and diversity. In: *Proceedings of the 12th international conference on learning representations (ICLR 2024)*, Vienna, 7–11 May 2024
70. Kleinberg J, Raghavan M (2021) Algorithmic monoculture and social welfare. *Proc Natl Acad Sci USA* 118(22):e2018340118. <https://doi.org/10.1073/pnas.2018340118>
71. Koh PW, Liang P (2017) Understanding black-box predictions via influence functions. In: *Proceedings of the 34th international conference on machine learning*. PMLR 70:1885–1894

72. Kuang L, Chen X, Chen F et al (2022) Uncovering neural distinctions and commodities between two creativity subsets: a meta-analysis of fMRI studies in divergent thinking and insight using activation likelihood estimation. *Hum Brain Mapp* 43(16):4864–4885. <https://doi.org/10.1002/hbm.26029>
73. Kutsche F, Taylor S, Erkkinen M et al (2025) Mapping neuroimaging findings of creativity and brain disease onto a common brain circuit. *JAMA Netw Open* 8(2):e2459297. <https://doi.org/10.1001/jamanetworkopen.2024.59297>
74. Le VA (2025) Copyright of photography and artificial intelligence: a tale of two technologies. *J Intellect Prop Law Pract* 20(8):562–569. <https://doi.org/10.1093/jiplp/jpaf039>
75. Lee E (2024) AI and the sound of music. *Yale Law J Forum* 134:187–236
76. Lehman J, Stanley KO (2011) Abandoning objectives: evolution through the search for novelty alone. *Evol Comput* 19(2):189–223. https://doi.org/10.1162/EVCO_a_00025
77. LeWitt S (1967) Paragraphs on conceptual art. *Artforum* 5(10):79–83
78. Lopes DM (2009) *A philosophy of computer art*. Routledge, London
79. Lopes DM (2026) AI art and artists: what they are, what they could be, what they should be. *J Aesthet Art Crit*. <https://doi.org/10.1093/jaac/kpag001>
80. Mazzi F (2024) Authorship in artificial intelligence-generated works: exploring originality in text prompts and artificial intelligence outputs through philosophical foundations of copyright and collage protection. *J World Intellect Prop* 27(3):410–427. <https://doi.org/10.1111/jwip.12310>
81. Mitchell M, Krakauer DC (2023) The debate over understanding in AI's large language models. *Proc Natl Acad Sci USA* 120(13):e2215907120. <https://doi.org/10.1073/pnas.2215907120>
82. Moon K, Green AE, Kushlev K (2025) Homogenizing effect of large language models (LLMs) on creative diversity: an empirical comparison of human and ChatGPT writing. *Comput Hum Behav Artif Hum* 6:100207. <https://doi.org/10.1016/j.chbah.2025.100207>
83. Moravec H (1988) *Mind children: the future of robot and human intelligence*. Harvard University Press, Cambridge, MA
84. Newman GE, Bloom P (2012) Art and authenticity: the importance of originals in judgments of value. *J Exp Psychol Gen* 141(3):558–569. <https://doi.org/10.1037/a0026035>
85. Ogundipe A (2025) Generative AI, agency and aesthetic disruption in photographic art. *J Aesthet Cult* 17(1):2557675. <https://doi.org/10.1080/20004214.2025.2557675>
86. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright CL, Mishkin P, Zhang C, Agarwal S, Slama K, Ray A et al (2022) Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, vol 35. Curran Associates, Red Hook, NY, pp 27730–27744
87. Pasquinelli M (2023) *The eye of the master: a social history of artificial intelligence*. Verso, London
88. Pugh JK, Soros LB, Stanley KO (2016) Quality diversity: a new frontier for evolutionary computation. *Front Robot AI* 3:40. <https://doi.org/10.3389/frobt.2016.00040>
89. Richards BA, Lillicrap TP, Beaudoin P, Bengio Y, Bogacz R, Christensen A, Clopath C, Costa RP, de Berker A, Ganguli S et al (2019) A deep learning framework for neuroscience. *Nat Neurosci* 22(11):1761–1770. <https://doi.org/10.1038/s41593-019-0520-2>

90. Ritchie G (2007) Some empirical criteria for attributing creativity to a computer program. *Minds Mach* 17(1):67–99. <https://doi.org/10.1007/s11023-007-9066-2>
91. Romera-Paredes B, Barekatin M, Novikov A, Balog M, Kumar MP, Dupont E, Ruiz FJR, Ellenberg JS, Wang P, Fawzi O, Kohli P, Fawzi A (2024) Mathematical discoveries from program search with large language models. *Nature* 625(7995):468–475. <https://doi.org/10.1038/s41586-023-06924-6>
92. Runco MA, Jaeger GJ (2012) The standard definition of creativity. *Creat Res J* 24(1):92–96. <https://doi.org/10.1080/10400419.2012.650092>
93. Sag M (2023) Copyright safety for generative AI. *Houst Law Rev* 61(2):295–347
94. Schacter DL, Addis DR (2007) The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos Trans R Soc B* 362(1481):773–786. <https://doi.org/10.1098/rstb.2007.2087>
95. Schmidhuber J (2010) Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Trans Auton Ment Dev* 2(3):230–247. <https://doi.org/10.1109/TAMD.2010.2056368>
96. Scruton R (1981) Photography and representation. *Crit Inq* 7(3):577–603. <https://doi.org/10.1086/448116>
97. Shofty B, Gonen T, Bergmann E et al (2022) The default network is causally linked to creative thinking. *Mol Psychiatry* 27(3):1848–1854. <https://doi.org/10.1038/s41380-021-01403-8>
98. Silver D, Huang A, Maddison CJ, Guez A, van den Sifre L, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M et al (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529(7587):484–489. <https://doi.org/10.1038/nature16961>
99. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A et al (2017) Mastering the game of Go without human knowledge. *Nature* 550(7676):354–359. <https://doi.org/10.1038/nature24270>
100. Silver D, Sutton RS (2025) Welcome to the era of experience. Google DeepMind. <https://storage.googleapis.com/deepmind-media/Era-of-Experience%20The%20Era%20of%20Experience%20Paper.pdf>. Accessed 5 Sep 2026
101. Simonton DK (2011) Creativity and discovery as blind variation: Campbell's (1960) BVSR model after the half-century mark. *Rev Gen Psychol* 15(2):158–174. <https://doi.org/10.1037/a0022912>
102. Somepalli G, Singla V, Goldblum M, Geiping J, Goldstein T (2023) Understanding and mitigating copying in diffusion models. *Advances in neural information processing systems*, vol 36. Curran Associates, Red Hook, NY, pp 47783–47803
103. Srnicek N (2017) Platform capitalism. Polity, Cambridge
104. Sutton RS (2019) The bitter lesson. Incomplete Ideas. <http://www.incompleteideas.net/IncIdeas/BitterLesson.html>. Accessed 5 Sep 2026
105. Sutton RS, Barto AG (2018) Reinforcement learning: an introduction, 2nd edn. MIT Press, Cambridge, MA
106. US Copyright Office (2025a) Copyright and artificial intelligence, part 2: copyrightability. United States Copyright Office, Washington, DC. <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf>. Accessed 5 Sep 2026
107. US Copyright Office (2025b) Copyright and artificial intelligence, part 3: generative AI training (pre-publication version). United States Copyright Office, Washington, DC.

108. Vaccaro M, Almaatouq A, Malone T (2024) When combinations of humans and AI are useful: a systematic review and meta-analysis. *Nat Hum Behav* 8(12):2293–2303. <https://doi.org/10.1038/s41562-024-02024-1>
109. Varela FJ, Thompson E, Rosch E (1991) *The embodied mind: cognitive science and human experience*. MIT Press, Cambridge, MA
110. Walton KL (1984) Transparent pictures: on the nature of photographic realism. *Crit Inq* 11(2):246–277. <https://doi.org/10.1086/448287>
111. Wiggins GA (2006) A preliminary framework for description, analysis and comparison of creative systems. *Knowl-Based Syst* 19(7):449–458. <https://doi.org/10.1016/j.knosys.2006.04.009>
112. Zador AM (2019) A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat Commun* 10:3770. <https://doi.org/10.1038/s41467-019-11786-6>
113. Zylinska J (2020) *AI art: machine visions and warped dreams*. Open Humanities, London

Figures

S: every configuration the format permits

$|S| = k^{W \cdot H}$, finite, astronomically large, almost all of it noise

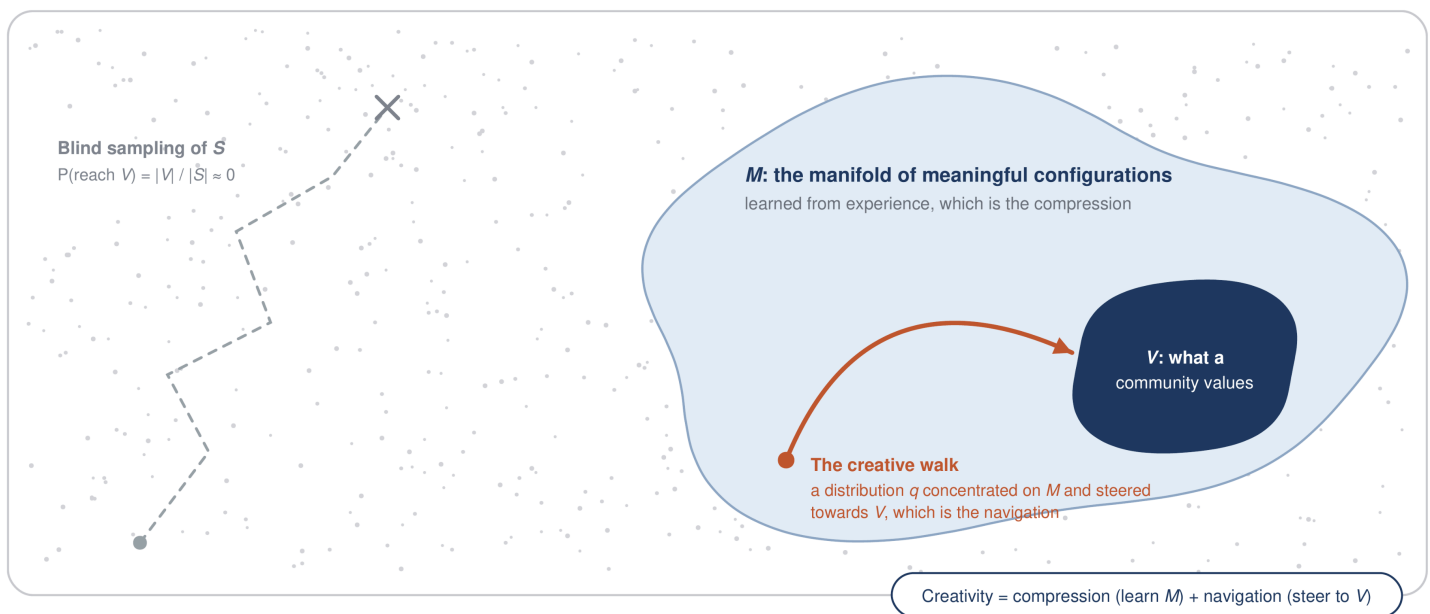


Figure 1

The finite possibility space and its valuable region. Schematic of the space S of all configurations a fixed medium permits, of cardinality $k^{W \cdot H}$, containing the low-dimensional manifold M of meaningful artefacts and, within it, the target region V a given community values. Uniform sampling of S (the "stochastic-monkey" regime) reaches V with probability $|V|/|S| \approx 0$; a creative process induces a distribution concentrated on M and biased towards V . Original vector schematic drawn by the authors; no generative AI tools were used.

One generate-and-select loop, two constitutions

functional correspondence inside the loop, constitutional divergence around it

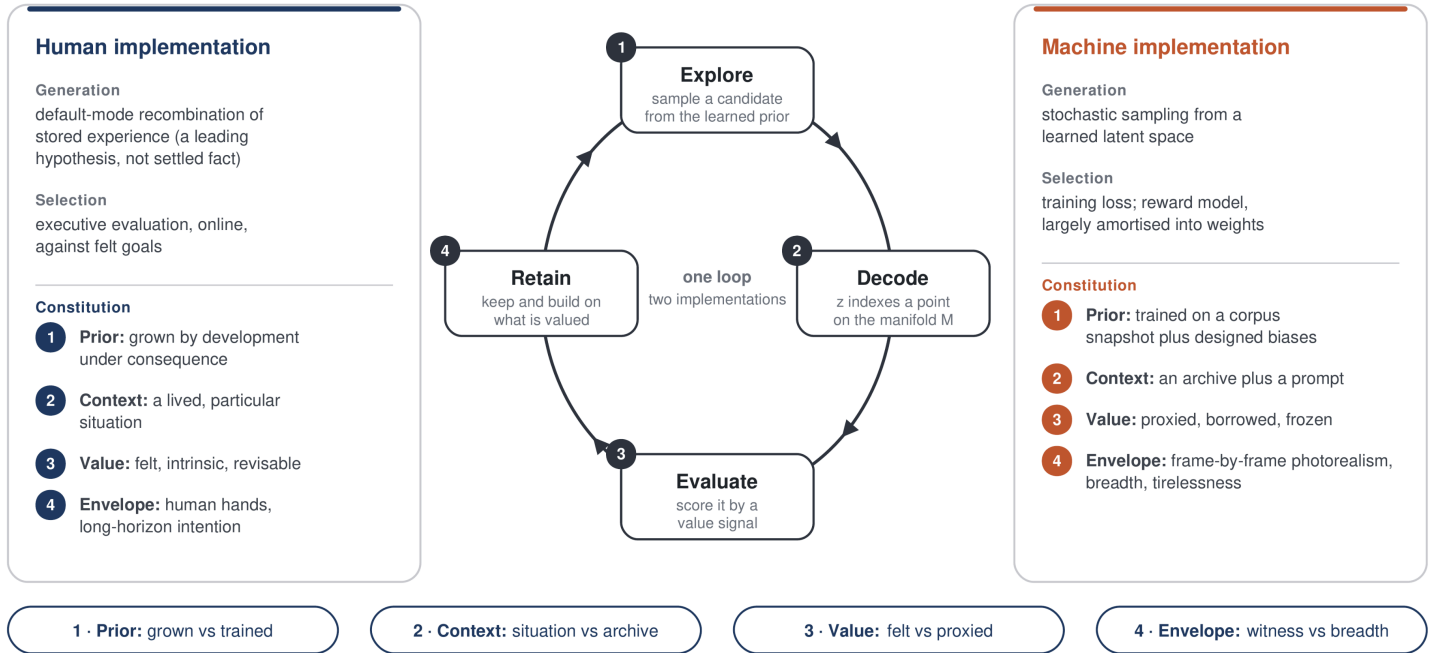
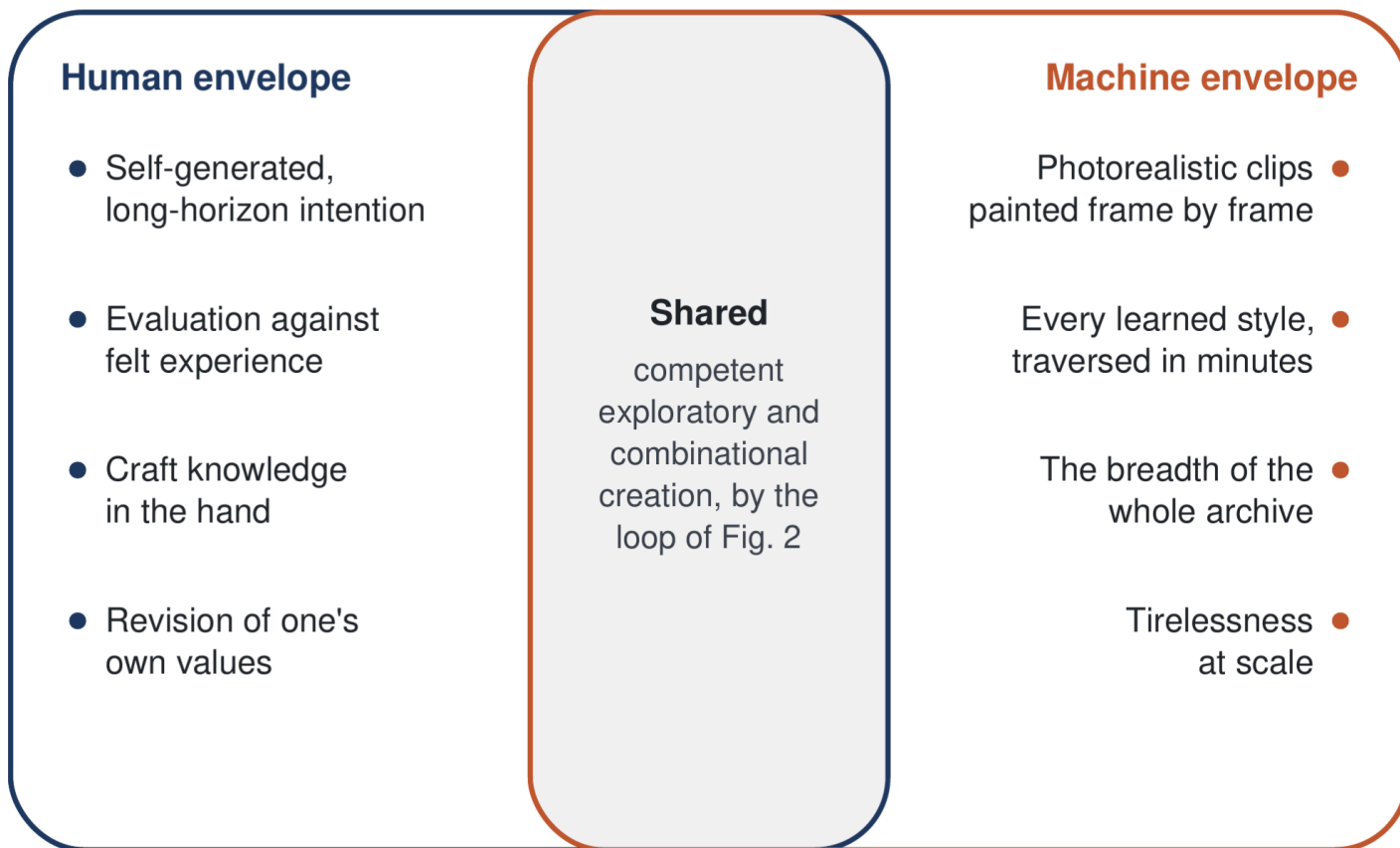


Figure 2

One loop, two constitutions. Centre: the shared generate-and-select loop (draw from a learned prior; decode to the manifold; evaluate; retain). Left: the human implementation (default-mode generation with executive selection, presented as a leading hypothesis) with its constitution: priors grown through embodied development under consequence; access to a lived, indexical situation; felt, intrinsic value; a human envelope of feasible operations. Right: the machine implementation (stochastic sampling from a learned latent space under a loss or reward signal) with its constitution: priors trained on a corpus; access to an archive plus a prompt; proxied, borrowed value; a non-human envelope (photorealistic clip synthesis, breadth, tirelessness). The four constitutional divergences are summarised beneath the panels. Original vector schematic drawn by the authors; no generative AI tools were used.

Two envelopes of the possible

what each creator can do; neither envelope contains the other



Superhuman along some axes, incapable along others;
no single scale ranks the two profiles.

Alternative, rather than lesser or greater

Figure 3

Two envelopes of the possible. Schematic of the operations feasible for a human creator (left) and for a generative model (right), with the shared ground of competent exploratory and combinational creation in the overlap. Neither envelope contains the other, which is the empirical core of the claim that machine creativity is alternative rather than lesser or greater. Original vector schematic drawn by the authors; no generative AI tools were used.

Where a machine can leave the human distribution on its own

the same generate-and-check loop under two ways of grounding value

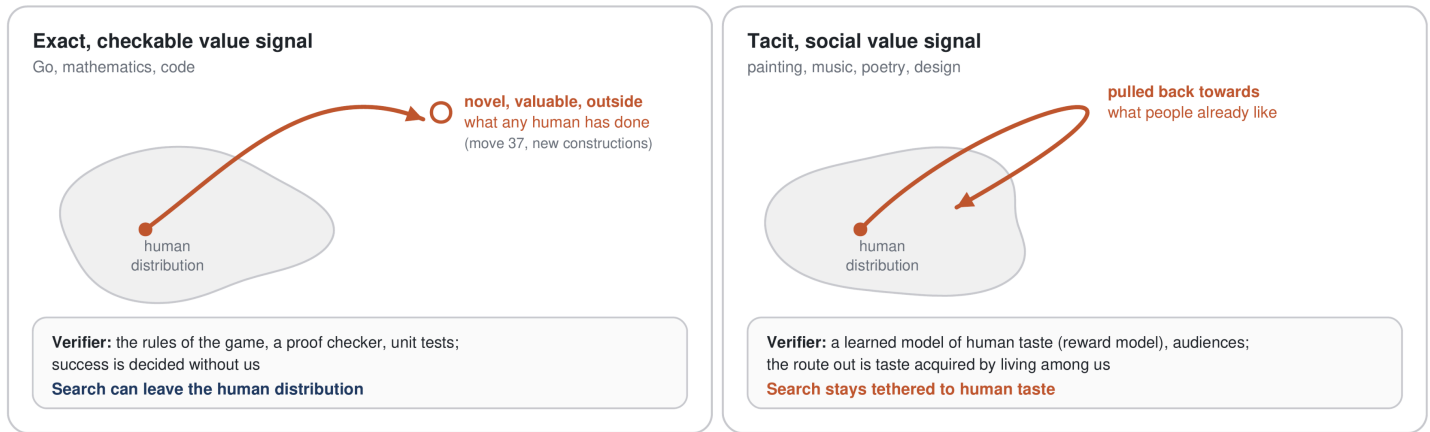
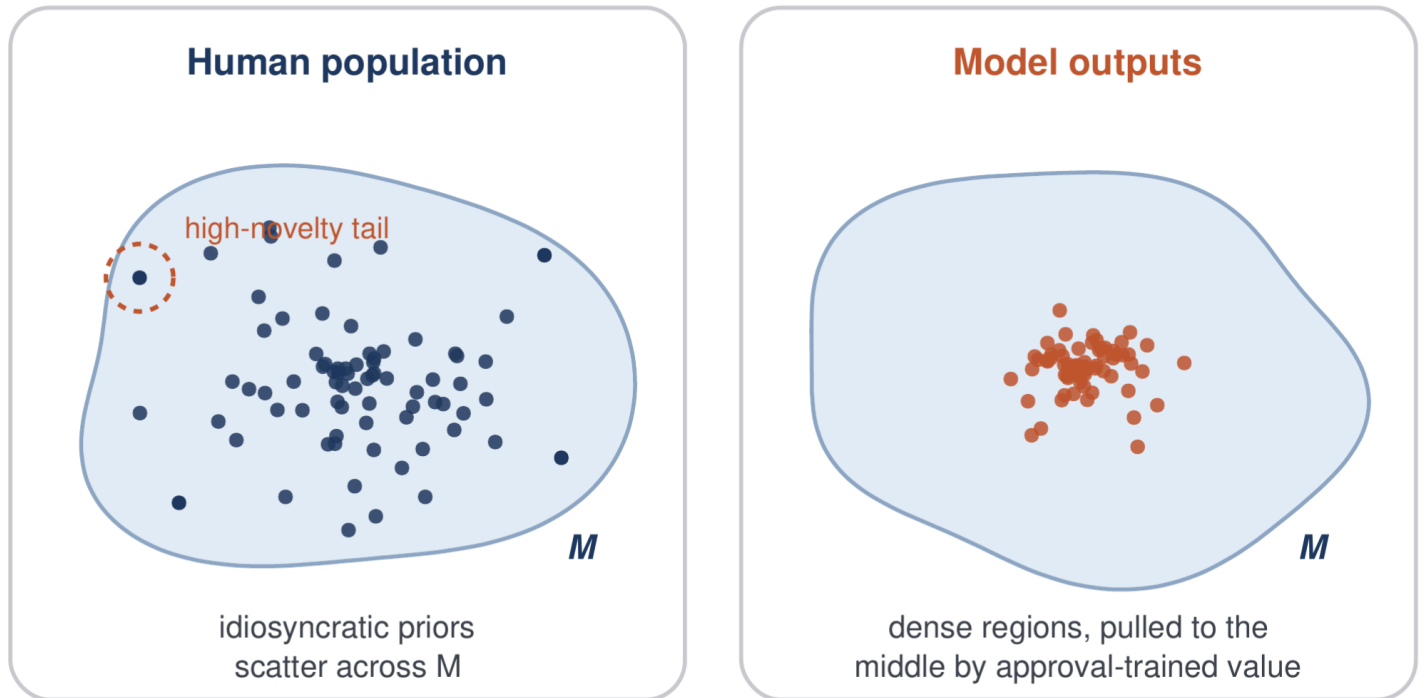


Figure 4

Where a machine can leave the human distribution on its own. The same generate-and-check loop under two ways of grounding value. Left: with an exact, checkable value signal (the rules of Go, a proof checker, unit tests), search can travel beyond the distribution of human practice and be confirmed without human judgement, as with AlphaGo's move 37 and the FunSearch constructions. Right: with a tacit, social value signal, the only available verifier is a learned model of human taste, which tethers the search to what people already like; the route out is taste acquired by living among us. Original vector schematic drawn by the authors; no generative AI tools were used.

Homogenisation as a mechanism

where many creators' outputs fall on the manifold M



Prediction

individual quality and novelty rise while the diversity of the pool contracts, more under preference tuning than under supervised fine-tuning

The durable human advantage sits in the high-novelty tail

Figure 5

Homogenisation as a mechanism internal to the technology. Schematic distributions of outputs on the manifold M for a human population, whose idiosyncratic priors scatter across the manifold and populate its high-novelty tail (left), and for a model that samples preferentially from dense regions and is pulled towards the comfortable middle by approval-trained value (right). The prediction is that individual quality and novelty rise while the diversity of the pool contracts, more under preference tuning than under supervised fine-tuning (Doshi and Hauser 2024; Kirk et al. 2024). Original vector schematic drawn by the authors; no generative AI tools were used.