CHAPTER 2

On Vector Agency

Emanuele Arielli

KEY TAKEAWAYS

- Agency in generative AI is *distributed* across humans, models, data, interfaces: there is no single locus.
- The issue of authorship is a legal one, while agency can be attributed to non-human components.
- Latent spaces embed *vector-like* tendencies (magnitude + direction) that steer outputs.
- Creative control progresses along a spectrum (prompting → curation → editing → model tuning), with a gradual rise in user intervention, not a binary on/off.
- Stochasticity enables novelty; "valuable surprise" (Boden, 2004) arises when randomness is structured by learned patterns.

2.1 INTRODUCTION

Consider a user who generates a series of images through a text prompt in an AI system, then carefully selects the most compelling output, evaluates it, refines the prompt, and iterates until the result satisfies their expectations. This also unfolds when someone interacts with a large language model: the prompt is formulated, the system generates a text, the user assesses, reformulates,

DOI: 10.1201/9781003740261-3

and tests again. What is striking in both cases is not simply the technological mediation, but the fact that the process involves a continual negotiation between human intention and machine-generated output.

If technological systems are generally understood as tools, what we encounter here is, however, a peculiar kind of instrumentality. Traditional tools operate in a largely deterministic way, like the *despeckle* command in Photoshop that smooths out noise in an image or the spellcheck function in a word processor. By contrast, generative AI systems embed a far higher degree of indeterminacy (see Manovich & Arielli, 2024; Young & Terrone, 2025). Their outputs, while constrained by data and architecture, cannot be reduced to a simple and predictable one-to-one mapping of command and effect. This is one of the reasons why it is increasingly common, especially in public discourse, to speak of the system's "autonomy," or even to reconsider the traditional notion of authorship (Arielli, 2023), despite the fact that attributing autonomy to AI is ambiguous and conceptually disputable (Bradshaw et al., 2013). Similarly, human-machine interaction is often described using terms like "cooperation," "partnership," or "teamwork" (Schmidt & Loidolt, 2023). Yet, while rhetorically appealing, these metaphors risk anthropomorphising machines without clarifying how such hybrid forms of agency are actually structured in practice (Nyholm, 2020; Placani, 2024).

The aim of this second chapter is to outline some general aspects of the debate on AI, authorship, and agency and to show how *agency* is not a binary attribute, toggling between agent and technology, but a *distributed* and gradual phenomenon that emerges dynamically from the interaction of users, algorithms and their designers, data, and contingency. Agency is not intrinsic to a specific acting source, but it is the product of a socially negotiated and technologically situated distribution across heterogeneous actors (see Anicker et al., 2024; Enfield & Kockelman, 2017). In this chapter, the discussion of agency concerns generative AI systems, particularly those producing visual or textual outputs through large models such as Midjourney, Stable Diffusion, or ChatGPT. The term *agency* here does not refer to what computer science calls *AI agents*, like autonomous software entities designed to perform tasks like driving cars or managing online services, but to the relational and distributed forms of action that emerge when humans interact with generative models. My focus is therefore not on autonomous decision-making systems, but on how creative processes unfold within networks of human and nonhuman contributors.

In this perspective, neural networks function as a form of cultural record: generative AI is trained on vast corpora drawn from the internet, books, and other repositories. They embody a collective sedimentation of culture: countless human contributions compressed into hidden statistical representations within the model. In other words, the agency of countless human creators, writers, artists, programmers, photographers, is embedded in the model as *latent space*, that is, a compressed mathematical space inside a trained AI model where patterns from the data are represented as vectors, allowing the system to combine and generate new outputs. The suggestion of this chapter is that latent space can also be understood as an actor-network (Gutiérrez, 2024; Latour, 2005): not a neutral geometry but the crystallisation of heterogeneous actors (data, algorithms, human past decisions, hardware constraints). It "acts" by structuring possible outputs, guiding what the system can generate or recognise. Each AI-generated output is therefore a remix of this accumulated knowledge, carrying forward traces of innumerable human decisions and expressions. Moreover, agency is inscribed not only in the data, but also in the architectures and infrastructures that support these systems. Programmers encode the algorithms, interface designers shape the modalities of interaction, and data annotators classify and filter the training sets. Together they constitute the background framework that determines how AI systems operate, constraining and enabling the apparent spontaneity of the output.

In this light, I propose the concept of *vector agency* to capture how agency is allocated, redistributed, and attributed (that is, how it comes to be *perceived* as acting) within socio-technical assemblages (in media theory, see Offert & Impett, 2025). The term *vector* here emphasises that agency operates not as a fixed property but as a *directional force*: each component of the system (human, technical, or institutional) exerts a certain orientation and intensity that shapes the collective outcome. In this regard, agency in AI-mediated creation is best understood as a gradient phenomenon rather than a discrete property or as the product of a specific and isolated subject. Agency, in this context, is rather an emergent pattern of coordination: a vector-like distribution that has both magnitude (how strongly a component steers activity) and direction (towards which goals it orients practice).

2.2 DISTRIBUTED AGENCY AND AUTHORSHIP

A traditional view ties agency to uniquely human faculties: the capacity to form intentions consciously, to choose among possible courses of action, and to justify those choices through moral or rational deliberation. An agent then acts upon the world through the use of tools and instruments. Artefacts participate in human activity by assisting, supporting, or guiding actions through features that are both cultural and technological. Anthropologists like Leroi-Gourhan (1993) described tools as carriers of cultural knowledge, while philosophers like Simondon (2017) analysed how technical objects co-constitute human experience. In psychology and cognitive science, Gibson (1986) introduced the concept of affordances to explain how objects and environments invite certain actions. Von Uexküll (1926) and Vygotsky (1978) extended this perspective by showing that mediation is not limited to material instruments: for von Uexküll, each organism inhabits an *Umwelt*, a perceptual world structured by biologically meaningful cues, which functions as a kind of natural tool for orienting action; for Vygotsky, human cognition is likewise mediated by cultural tools and symbolic systems such as language.

Together, these accounts underline that artefacts and mediating structures, whether biological or cultural, shape human activity through the affordances and meanings they embody. A canonical illustration is Hutchins' (1995) study of flight decks, where the unit that effectively flies the aircraft is not the pilot alone but the integrated system of pilots, cockpit's instruments, checklists, and maps. Cognition and agency are distributed across humans and artefacts rather than localised in a single head (Clark & Chalmers, 1998).

In light of these perspectives, recent discussions in philosophy of technology, media studies, and sociology have questioned whether the traditional agent-instrument dichotomy should still hold (see Enfield & Kockelman, 2017; Floridi & Sanders, 2004; Latour, 2005; Roine & Piippo, 2021). These debates converge on the idea that agency cannot be limited to intentional or conscious human subjects, since the effects of human action are often mediated and transformed by technological systems. However, this does not imply that production technologies possess agency *of their own*. Rather, agency emerges within configurations where humans and technologies act jointly, each contributing to the outcome in distinct ways. In this sense, I can speak of shared or *distributed* agency, where control and initiative are dynamically negotiated between human operators, technological affordances, and institutional frameworks. In Latour's terms, and from the perspective of actor-network theory (ANT), an *agent* or *actant* is any entity that has the capacity to "make a difference" in a network (Latour, 2005, p. 71). ANT deliberately places humans and nonhumans on an analytical continuum, treating both as potential sources of causal efficacy. An agent, in this sense, is anything that can modify the state of affairs of another, either by initiating causal processes or by redirecting ongoing ones. Such capacities are not confined to individual humans but are distributed across networks of people, machines, artefacts, and infrastructures. As in the pilot's example, the effective agency in accomplishing a task resides in the human-tool *assemblage* rather than in the person alone (see Roine & Piippo, 2021, for a discussion in digital media). From this relational viewpoint, one may also speak of *teleonomic* agents, that is, entities whose actions are guided by internal programs or goal-directed mechanisms, such as biological organisms governed by genetic codes, or artificial systems structured by algorithms (Enfield & Kockelman, 2017). Their activity follows an intrinsic orientation towards certain outcomes, even in the absence of conscious intention. Extending this perspective, Gell (1998) proposed that even artworks can be considered agents. They exercise influence not because they possess intention, but because they mediate relations and elicit responses, thereby transforming the behaviour or perception of others.

An AI model can thus be regarded as an agentive component within a relational system rather than as an autonomous agent in the technical sense. Its training data, architecture, and weights orient its behaviour, shifting the human-machine relationship from a straightforward user-technology relation towards a more iterative and responsive exchange. In previous technologies like Photoshop, a user operates a tool to achieve a predefined outcome with predictable results. By contrast, generative systems respond in ways the user cannot entirely foresee, introducing unanticipated forms, styles, or directions into the creative process. According to the view that agency is exclusively human, agency and authorship were assumed to coincide: to act was to intend, and to intend was to author. From the perspective introduced here, however, where agency is redistributed across networks of code, data, and interfaces, this coincidence no longer holds. Agency can now be exercised without implying authorship, since action and intention are dispersed among heterogeneous actors (see also Chapters 3 and 4, this volume). In this sense, generative AI systems channel and transform human creative capacities through algorithmic processes, producing outcomes that no single author can fully anticipate or control.

This expanded field of agency has also provoked legal and institutional challenges, particularly concerning authorship, originality, and copyright (Naqvi, 2020). Artists and rights holders have filed lawsuits against platforms such as Stability AI, Midjourney, DeviantArt, and Imagen, arguing that the use of copyrighted works in training datasets constitutes infringement. Recent analyses (MediaLaws, 2025; White House CRS, 2025) indicate that current legal frameworks, built on the assumption that authorship must stem from a human agent, are ill-suited to forms of production that emerge from distributed networks of human and nonhuman contributors. The question of AI authorship is widely debated and has attracted considerable public and legal attention. Yet, these discussions often overlook the fact that, while AI systems contribute to creative processes, they do not meet the conditions of authorship. Authorship still presupposes intention, accountability, and the legal capacity to hold or transfer rights, criteria that presuppose personhood. Current AI systems, however sophisticated, fulfil none of these requirements. They operate as mediating agents within human-devised infrastructures but cannot initiate claims, assert ownership, or bear responsibility for their outputs (Hayot, 2024). They exercise agency, but not authorship.

The distinction between exercising agency and achieving authorship has precedents in cases involving human agents, as illustrated by notable art world disputes, like the case between conceptual artist Maurizio Cattelan and Daniel Druet, a wax sculptor who produced many of the figures used in Cattelan's installations between 1999 and 2006. When Druet later claimed authorship of these works, French courts dismissed the claim and affirmed Cattelan as the sole author. Their reasoning reflected a conceptualist understanding of art: authorship belongs to the one who conceives, frames, and situates the work, not to the one who realises its material form under direction. In other words, Druet exercised agency, his fabrications required countless artistic decisions concerning anatomy, expression, and surface treatment. Yet, the court held that his contribution remained primarily technical and material, with little effect on the work's conceptual core.

For that reason, his agency did not meet the legal threshold of authorship, because the origination and meaning were attributed to Cattelan (de Brogniez, 2024; Jardonnet, 2002).

The cases are comparable, but with a crucial distinction: Druet lost his authorship claim because his contribution was deemed insufficiently creative, whereas an AI system cannot *claim* authorship at all, regardless of its creative capacity, since it lacks legal personhood and the capacity to assert rights. In other words, while a human agent can both exercise and claim authorship, an artificial system can only exercise a form of agency without any corresponding juridical or intentional claim. Paradoxically, an AI might operate with *more* autonomous agency than Druet, producing outputs that go well beyond a commissioner's intentions and conceptions. This is true in cases in which a user provides only a prompt or general instruction, and the system generates images, videos or texts that exceed the user's foresight, suggesting styles, images, or scenes not explicitly anticipated in the prompt. In this sense, the system does more than execute – it proposes directions. In human-AI collaborations, the system's generative operations can introduce unexpected directional vectors that reshape the creative intent while authorship remains human.

2.3 "ACTION" IN LATENT SPACE

There is often a risk in extending the notion of *agent* to every component of a system, thereby diluting its meaning. As discussed earlier with reference to Latour and Actor-Network Theory, such an approach usefully emphasises the active role of objects and infrastructures but risks turning the concept of agency into a universal property.

An agent is something that behaves with a certain degree of autonomy, meaning that it can regulate its own operations within the boundaries of its design. When an entity's actions are driven by an internal capacity to process inputs and generate behaviour beyond the immediate control of a human operator, it assumes control over some aspects of a situation. *Autonomy*, therefore, describes a form of *internal control* that allows a system to act according to its own rules or representations rather than merely responding to external commands. This definition applies not only to physical and material entities such as machines, robots, or organisms, but also to abstract ones like programs and algorithms (Bucher, 2018).

As will be discussed in Section 2.6, where stochasticity and controlled randomness are examined as key conditions for how we perceive agency in generative systems, agents understood as entities endowed with a certain level of autonomy exhibit behaviours that are not fully anticipated by their designers or users (Floridi & Sanders, 2004). Complete predictability would make them deterministic mechanisms rather than agents. The more a system's behaviour arises from its own information processing rather than from direct human prescription, the more control effectively shifts away from human operators. A purely reactive system, such as an alarm connected to a smoke detector, does nothing unforeseen on its own, whereas a more complex system can initiate processes or responses based on its internal state, which may at times surprise us.

Tools are both products and instruments of agency: they are created through the exercise of agency and subsequently serve as means through which further actions are carried out. They embody the intentions and skills of their makers, and also extend the agency of those who use them to pursue new ends.

In this way, a tool is *solidified* or "fossilised" agency, the material trace of human goals and actions hardwired into an object. A hammer's apparent agency is not "its own" but a frozen residue of the blacksmith's and carpenter's purposes. More complex artefacts are products and traces of distributed, societal agency. A traffic light at an intersection, for instance, embodies the coordinating intentions of urban planners, engineers, and policymakers (Suchman, 2007). Likewise, a well-designed door handle results from distributed agency: what appears as the decision of a single designer in fact depends on a network of shared norms, ergonomic research, stylistic traditions, manufacturing practices, and culturally ingrained gestures that determine how people grasp and turn it. The artefact's influence on action is thus the *crystallised intent* of human agents, expressed through layers of collective and individual design decisions. By understanding tools as derivative agents, we recognise that their apparent capacity to "act" on the world is borrowed from human designers and users.

On the one hand, AI systems are clearly artefacts designed and trained by humans: they originate from our goals and data, and therefore contain latent traces of human agency, as any tool does. On the other hand, GenAI, particularly machine-learning models that adapt and generate outputs, exhibits agent-like features, since it can produce unanticipated behaviours, respond flexibly to new inputs, and even appear to pursue goals in novel contexts. For example, DeepMind's *AlphaGo*, which defeated the Go champion Lee Sedol in 2016, famously played moves that experts described as creative and unpredictable. Sedol himself commented that he had learned "a new kind of creativity" from the system, acknowledging that its play revealed possibilities that no human player had previously imagined. Similar forms of apparent goal-seeking behaviour appear in generative image models that reinterpret prompts in unexpected ways, producing visual solutions that diverge from a user's initial intended images.

When such an AI system acts or produces an output, it is not simply executing a fixed script written line by line by a programmer; rather, it draws on a complex internal structure of learned representations to respond to the situation. This responsiveness creates the impression of purposiveness or autonomy, even though the system's "goals" are emergent properties of its training rather than consciously held intentions.

The latent space or internal structure of a trained neural network is the direct product of training on human-curated data and objectives and contains agency in a distributed form: the network has internalised not just information, but ways of processing and responding that reflect human cognitive patterns and goals.

The agency embedded in these learned representations can then be activated to produce behaviours that, while rooted in human training, exceed direct human control.

2.4 VECTOR AGENCY

The purposeful structure encoded in a computational artefact like a trained neural network, therefore, is not merely inert information. When activated in context, it behaves like a *vector* of actions. A trained neural network is a multi-dimensional space of features or representations that the model has learned. This latent space is the crystallisation of countless human decisions, such as choices in model architecture, the selection and labelling of training data, and, more broadly, the ways humans classify and represent the world through images, language, and symbols. It encodes biases, strategies, and patterns that originated in human agency (see Chapters 3, 4, 5, and 6, this volume). Yet, when the model is running, these patterns are no longer passive data, but function as forces that shape the system's outputs. Each dimension of a latent vector can be thought of as exerting a push on the system's behaviour, with a specific *direction* (the type of outcome it favours) and a certain *magnitude* (the strength of its influence). In other words, the model's latent features carry an implicit orientation towards certain interpretations or actions, and their strength determines how decisively they steer the outcome. The artefact's internal state is thus vectorial in nature: it has direction and force, not only semantic content.

Suppose a user enters a prompt such as "a garden with people walking, in the style of Impressionism." The model's latent space contains a cluster of vectors associated with Impressionism, but this cluster is not uniform. It is a field of forces shaped by the work of individual painters that the model has internalised from its training data. Monet, Renoir, Degas, and others each contribute a directional pull, though with different magnitudes, exercising *historical influence*, that is, a continuing capacity to influence cultural production and aesthetic perception across time. What I call "Impressionism" in the model is therefore not a single stylistic vector, but the resultant of many painter-specific vectors, each with its own magnitude and orientation (see Chapter 6, this volume). A Monet-vector may dominate the field, shaping outputs towards water lilies, gardens, and soft atmospheric effects. A Degas-vector may exert less overall magnitude but still redirect the output towards dynamic figural compositions. The generated image emerges as the vector sum of these painterly forces, combined with the user's prompt and the system's architectural constraints. When users provide more specific and targeted prompts, their instructions introduce a stronger directional force that overrides the neural network's inherent tendencies.

The resultant agency of the system can be understood as the vector sum of all contributions the original authors of the content on which the system was trained, the designers and programmers who built it, the degree of intervention by the user, and the influence of stochastic or contingent factors, such as random sampling, temperature settings (i.e., a parameter that determines how predictable the AI's output will be, with low values producing precise answers and higher values generating more diverse results), or unanticipated interactions between prompt elements (see 2.6). The vectorial character of agency means that I can describe both how strongly a component influences an outcome (its magnitude) and what kind of outcome it is directed towards (its direction). This perspective, which I call *vector agency*, offers a way to analyse how control is shared, redistributed, and restructured within a generative system.

2.5 MODES OF HUMAN-AI INTERACTION

Given that AI systems contain these distributed vector forces, the question becomes: how much directional control can users actually exert? This depends on the specific ways they choose to interact with and intervene in the system's latent structure. Agency is therefore not a fixed property but a distribution that shifts with context, with the user acting as an agent who intervenes in and redirects these vectors. What matters is the ongoing interaction between system affordances and user choices, where users of generative systems assume, share, or relinquish control in different ways. This produces a spectrum of user involvement, ranging from simple prompt-based interaction, where users steer generation through textual or visual instructions, to more advanced technical interventions that modify the system's behaviour through parameter tuning, dataset adjustments, or model fine-tuning (Dellermann et al., 2019; Ensslin & Nelson, 2025; Haase & Pokutta, 2024).

Generative image systems such as Midjourney or Stable Diffusion can be used in radically different ways. Practices can be arranged on a spectrum of agency, from the most passive, where the user delegates almost entirely to the system, to the most active, where the system becomes a specialised instrument under the user's control. The tension between delegating action to the machine and appropriating it as a tool of intentional action can be described as distributing human agency at different levels (also see Chapter 4, this volume).

At the first level, often described as *zero-shot prompting*, the user provides a trivial or minimal prompt, such as a single word ("sunset") or a short phrase ("painting robot"), without specifying subjects, settings, or stylistic parameters. In this case, the system generates an image based on its default internal associations, with little external constraint. The action resembles initiating rather than directing a process: the user triggers the model but does not meaningfully guide its internal vector composition. From the standpoint of human agency, the contribution is limited to the initial conceptual impulse, the formulation of an idea, while the detailed realisation is delegated almost entirely to the system (Oppenlaender, 2023). At the second level, a user acts as *prompt crafter* by taking more control and engineering a richer prompt. By adding descriptors (details of the subject, style keywords, technical instructions), the user guides the AI towards a more specific vision. This reduces the AI's freedom to insert unintended elements.

A third level involves *iterative curation*. Here, the user exploits the system's generative multiplicity, producing many candidates and selecting among them. The action is closer to curation than to creation in the traditional sense: the system proposes, the user evaluates and chooses.

A fourth level centres on *post-processing and hybrid editing*. The AI provides the material that the user then modifies, corrects, or composes. Users who edit or retouch AI images reclaim a higher share of agency, no longer delegating fully but interrupting, redirecting, and supplementing. A common example is *inpainting* or *outpainting*: the user masks part of the AI-generated image (a distorted hand, an unwanted background) and prompts the AI again to fill or replace that area. Most Stable Diffusion interfaces allow inpainting, and Midjourney has recently added features like the "vary region" feature, which lets users select a specific area within an image and regenerate or modify only that portion while keeping the rest intact. Another example is exporting the AI image into external software (such as Photoshop) for manual touch-ups that allow to adjust lighting, clean artefacts, or compose several AI images together. At this stage, the user's role shifts from prompting to actively co-creating the image with the AI.

Finally, a fifth level consists of *advanced technical control* and customisation. Some users manipulate advanced parameters or even modify the model itself. Expert Stable Diffusion users, for instance, adjust the CFG (Classifier-Free Guidance) scale, which controls how strictly the image follows the prompt versus allowing randomness. They may also set seeds, ensuring reproducibility or controlled variation. Extensions like ControlNet add further constraints, such as input sketches, poses, or edge maps (Zhang & Agrawala, 2023). Beyond parameter tuning, advanced users fine-tune or re-train the model with new data, altering its behaviour. An artist might use DreamBooth to teach the system a personal style or a specific person's face, thereby customising the model into a bespoke tool. Programmatic uses also fall into this level, such as calling the API or writing scripts to generate images in batches or integrating the AI into larger creative coding pipelines. At this point, the user effectively becomes a co-creator of the model itself, with the system serving as a highly specialised instrument.

This typology of human-machine degrees of intervention shows that creative control in AI-assisted work is not simply lost or gained but redistributed. The human directs the process through prompting, curation, and editing, while the AI carries out the low-level realisation through vector operations in latent space. The role of the user shifts from mere initiator to curator or choreographer, steering outcomes without executing every detail. In this perspective, creative workflows become iterative feedback loops: the user prompts, the AI generates, and the user evaluates and refines. This co-adaptive cycle enables humans to learn the AI's tendencies and adjust strategies accordingly. The concept of vector agency highlights that generative systems operate through internal forces, statistical relations, and learned patterns that orient their outputs towards particular kinds of results. These orientations are not fixed; human intervention can strengthen, weaken, or redirect them through prompting, selection, or fine-tuning. In this sense, interaction with the system amounts to a continual reconfiguration of the directions and intensities of its internal forces. Latent structures define possible directions of action, while users determine how strongly those directions are activated, constrained, or recomposed.

2.6 UNPREDICTABILITY BY DESIGN

As visual generative AI systems gain complexity and autonomy, the gap between what the human specifies and what the system delivers tends to widen, although the extent of this distance depends on the user's level of intervention, as we have seen. This distance is itself a consequence of distributed agency: as agency becomes shared across human and machine components, the relation between initial instruction and final outcome becomes less direct. Generative systems and interfaces are thus characterised by a relative displacement between human intention and the resulting artefact, whether text (Bajohr, 2024 speaks of "writing at a distance"), image, video, or music. Moreover, a fully deterministic and predictable behaviour would not appear to us as agentive at all. Although not every unforeseen result is perceived as a sign of agency, a certain degree of unpredictability is a necessary condition for something to appear autonomous. When an outcome cannot be rigidly predicted, we often interpret it as if distinct creative sources were involved in the process (Manovich & Arielli, 2024). In this sense, we describe such technologies as "intelligent" precisely because they generate outputs that are not entirely foreseeable: the tool begins to appear as an active participant rather than a mere instrument.

Modern machine learning systems amplify the impression of machine agency because their inner workings are opaque. Unlike rule-based programmes, deep neural networks operate as black boxes, making their output difficult to trace. The process feels less like following a recipe and more like observing a natural phenomenon, such as organism growing (see Riemer & Peter, 2024; Young & Terrone, 2025). Crucially, this lack of transparency presupposes a degree of loss of direct control by the human author. In a simple rule-based system, inputs and outputs are tightly coupled: the process is deterministic, meaning that identical inputs predictable results, and it is also transparent, since the causal steps between input and output can be inspected and understood. Determinism concerns the *predictability* of outcomes, whereas transparency concerns the *intelligibility* of the process. By contrast, modern machine-learning approaches rely on stochastic processes, introducing degrees of randomness that make outcomes unpredictable. A traditional algorithm yields identical outputs each time, but advanced AI systems generate variations: results are statistically similar yet never exactly the same.

The variable of unpredictability is one of the most direct means of achieving a deliberate relinquishment of control. In art history, chance was central to avant-garde movements such as Dada and Surrealism, which experimented with random collage, automatic writing, and aleatory music to inject uncertainty into creation. Tristan Tzara's instructions for composing a Dada poem by pulling words from a hat, or John Cage's use of coin flips and the *I Ching* to compose music, were explicit attempts to let the artwork "write itself." Cage (1961) famously said his goal in using chance operations was "to let things be themselves," removing his ego from the process so that the outcome could surprise even its creator. Here, artists intentionally ceded control to procedures and welcomed chance as a co-creator. The resulting works often felt less like deliberate human craft and more like discoveries, as if the system or process had its own voice (Piekut, 2013).

Randomness later became a structural principle in 1960s–1970s generative and computer art. Artists such as Vera Molnár, Frieder Nake, and Manfred Mohr employed algorithmic procedures to produce compositions whose outcomes were not entirely predictable, yet were still determined by well-defined mathematical rules (Galanter, 2003, 2016; Nake, 2005). Here, randomness manifests as *bounded indeterminacy*, used to activate variability within a controlled framework. This logic persists in contemporary machine-learning systems, where stochastic processes underpin both training and generation. In neural networks, randomness enters through mechanisms such as random weight initialisation, sampling, and temperature-based variation. These stochastic elements ensure that results are never perfectly reproducible, giving rise to the impression of spontaneity or creative emergence.

Recent technological practices extend this lineage. Experiments in jazz with algorithmic tools, for example, combine stochastic and deterministic processes to produce novel improvisations. Such systems generate "contextually meaningful contingency" (Wilf, 2014): by playing with those tools, musicians are provoked to modify their own repertoire of standardised intentions and cultivate more inventive improvisational strategies.

Randomness and stochasticity are not incidental but constitutive features of how today's deep-learning works and the element of uncertainty arises from processes built into both training and generation. Diffusion models, for instance, begin from random noise and iteratively refine it into a plausible image guided by a prompt. This built-in randomness enables novelty and surprise. If the process were fully deterministic, outputs would resemble static tools or databases, lacking the spontaneity that users interpret as creativity (see Boden, 1998, 2004).

Yet, randomness should always operate within a structure in order to be perceived as meaningful. To make an analogy, a kaleidoscope produces shifting patterns by chance, but the fascination lies in how this randomness is channeled through geometric principles: each turn creates genuine surprise within a framework that guarantees visual coherence and beauty. A Dadaist cut-up poem relies on chance juxtapositions, but generates meaning through the unexpected connections and apparent coherences that emerge from seemingly unrelated fragments. Similarly, modern AI models employ randomness constrained by learned statistical patterns. A text generator samples words stochastically, but always according to probability distributions derived from real language, ensuring the output still conveys sense. Likewise, an image generator begins with noise in latent space, but noise is steered by internal visual patterns learned from training data to form a plausible picture.

In all these cases, the apparent autonomy is not pure chaos but a guided random walk. A diffusion model does not invent images *ex nihilo*; it reconfigures statistical regularities extracted from its training set. What looks like chance is already structured by cultural and technical histories embedded in the data: styles, genres, iconographies, and conventions accumulated across millions of images (see Chapters 5 and 6, this volume). The system's "surprises" emerge because stochastic variation is constrained by the vectors and directions inscribed during training. In this sense, unpredictability is not free-floating but takes place within a structured field of forces.

2.7 CONCLUSION

Generative AI systems blur the previous distinction between agent and tool. Users neither control outputs completely nor predict results precisely: total unpredictability would make the systems useless-and total control would undermine the system's generative function. Unlike an airplane cockpit, which is a complex and sophisticated system that still allows complete control by the operator, generative AI systems express agency sedimented in human culture in partly unpredictable ways. This agency appears in the form of styles, forces, and vectors within latent space, activated through mechanisms that remain variably open to contingency and stochastic processes. In this, the act of creation involves both intervention and observation: the user makes use of a powerful tool yet simultaneously witnesses outcomes that could not have been fully anticipated. Unpredictability and stochasticity amplify this tension, eroding the sense of full control and revealing the collaborative distribution of agency within the human-machine ensemble.

The outcomes of systems like Midjourney or Stable Diffusion emerge from multiple vectors of influence: training data provided by artists (or, according to critical voices, *stolen* from them), design choices of engineers, user prompts and edits, and stochastic factors such as random seeds. Each is an agency vector with a magnitude and direction. The result is a vector sum: an emergent pattern rather than a property of any single entity.

The concept of vector agency helps us understand how control and creative intention gets distributed when the tools become active participants rather than passive instruments. By analyzing agency in latent space, I acknowledge that a trained model's structure is not just inert information: it is an opportunity for action. When I probe a latent space vector (say, tweaking an image-generating AI's latent code to produce a new face), I am effectively activating an embedded agentic tendency and the system responds by generating a specific kind of result. The extent to which users can do this, however, depends on the openness of the model and on user literacy. In proprietary systems, such as commercial image generators, access is usually limited to predefined interfaces and affordances, restricting human agency to prompting and curation. By contrast, open-source models permit deeper forms of intervention, such as fine-tuning or retraining on specific datasets, though these possibilities require substantial technical skill.

Understanding this helps us better design and interact with AI systems. We can map which latent vectors correspond to which "forces of action" and thereby make the artificial agent's behaviour more open to investigation and possibly more steerable. If we know that a particular component in the model consistently nudges an image or text towards a specific style, tone, or representation, we gain a handle on the otherwise hidden agentic tendencies of the system. Such interpretability turns the black box into a more steerable partner. This requires reaching a level of advanced technical control (see Section 2.5) that users without direct access to proprietary systems can nonetheless approach by participating in communities and shared practices, such as online forums and Discord groups dedicated to Midjourney, Leonardo AI, or Adobe Firefly, where collective experimentation compensates for limited system transparency. These collective spaces document how users empirically identify and describe the system's behavioural tendencies, contributing to a distributed form of interpretability and literacy that complements technical analysis. This also reminds us that our own agency as humans should not be removed from the equation: we might *engineer* those latent spaces, and we can intervene in them. At the same time, as McLuhan (1964, p. 7) famously stated, "we shape our tools, and thereafter our tools shape us." By consciously shaping generative AI systems, through their training data, model architecture, and user interface, we determine the vector field in which human users later act. Those systems in turn shape our creative process by opening new paths and constraints for expression.

FURTHER READING

Arielli, E. (2023). AI-aesthetics and the artificial author. *Proceedings of the European Society for Aesthetics*, *15*, 40–54. https://philarchive.org/rec/ARIAAT-8 (It examines how generative systems complicate traditional authorship, proposing the "artificial author" and showing how unpredictability and distributed creativity reshape agency.)

Manovich, L., & Arielli, E. (2024). *Artificial aesthetics: A critical guide to AI, media and design*. https://manovich.net/index.php/projects/artificial-aesthetics (This book provides a wide-ranging framework for understanding generative AI in visual culture and design, with practical and critical lenses on control, authorship, and human–machine co-creation.)

Rahwan, I., Cebrian, M., Obradovich, N., et al. (2019). Machine behaviour. *Nature*, *568*, 477–486. https://doi.org/10.1038/s41586-019-1138-y (This article argues for studying AI systems as behavioural agents in the wild, offering methods to analyse autonomy, adaptation, and social impact.)

Young, N., & Terrone, E. (2025). Growing the image: Generative AI and the medium of gardening. *The Philosophical Quarterly*, *75*(1), 310–319. https://doi.org/10.1093/pq/pqad111 (It develops a "gardening" metaphor for text-to-image systems, treating image making as a process of tending and emergence; useful for thinking about distributed agency, stochasticity, and iterative steering.)

REFERENCES

Anicker, F., Flaßhoff, G., & Marcinkowski, F. (2024). The matrix of AI agency: On the demarcation and social imputation of AI as an actor. *Sociological Theory*, *42*(4), 318–324. https://doi.org/10.1177/07352751241289925

Arielli, E. (2023). AI-aesthetics and the artificial author. *Proceedings of the European Society for Aesthetics*, *15*, 40–54. https://philarchive.org/rec/ARIAAT-8

Bajohr, H. (2024). Writing at a distance: Notes on authorship and artificial intelligence. *German Studies Review*, *47*(2), 315–337. https://doi.org/10.1353/gsr.2024.a927862

Bender, E. M., 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* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, *103*(1–2), 347–356. https://doi.org/10.1016/S0004-3702(98)00055-1

Boden, M. A. (2004). *The creative mind: Myths and mechanisms* (2nd ed.). Routledge. https://doi.org/10.4324/9780203508527

Bradshaw, J. M., Hoffman, R. R., Woods, D. D., & Johnson, M. (2013). The seven deadly myths of "autonomous systems. *IEEE Intelligent Systems*, *28*(3), 54–61. https://doi.org/10.1109/MIS.2013.70

Bucher, T. (2018). *If…Then: Algorithmic power and politics*. Oxford University Press. https://doi.org/10.1093/oso/9780190493028.001.0001

Cage, J. (1961). *Silence: Lectures and writings*. Wesleyan University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19. https://www.jstor.org/stable/3328150

de Brogniez, M. (2024). *Procès Cattelan: Rester sur sa faim ou imaginer d'autres mondes possibles, note sous Paris, 5 juin 2024*. *Auteurs et Média*, *2024*(2), 249–257. https://hdl.handle.net/2268/329148

Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2019). The future of human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems. In *Proceedings of the 52nd Hawaii international conference on system sciences (HICSS)* (pp. 1–10). https://doi.org/10.48550/arXiv.2105.03354

Dennett, D. C. (1987). *The intentional stance*. MIT Press.

Enfield, N. J., & Kockelman, P. (Eds.). (2017). *Distributed agency*. Oxford University Press.

Ensslin, A., & Nelson, J. (2025). Co-creative multimodal authorship as procedural performance with DALL·E. In W. Slocombe & G. Liveley (Eds.), *The Routledge handbook of AI and literature* (pp. 315–331). Routledge.

Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, *14*(3), 349–379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d

Galanter, P. (2003). What is generative art? Complexity theory as a context for art theory. In *Proceedings of the generative art conference* (pp. 1–10).

Galanter, P. (2016). Generative art theory. In C. Paul (Ed.), *A companion to digital art* (pp. 240–283). WILEY Blackwell. https://doi.org/10.1002/9781118475249.ch5

Gell, A. (1998). *Art and agency: An anthropological theory*. Clarendon Press. https://doi.org/10.1093/oso/9780198280132.001.0001

Gibson, J. J. (1986). *The ecological approach to visual perception*. Lawrence Erlbaum Associates.

Gutiérrez, J. L. M. (2024). On actor-network theory and algorithms: ChatGPT and the new power relationships in the age of AI. *AI and Ethics*, *4*, 1071–1084. https://doi.org/10.1007/s43681-023-00314-4

Haase, J., & Pokutta, S. (2024). Human-AI co-creativity: Exploring synergies across levels of creative collaboration. *arXiv preprint* arXiv:2411.12527. https://doi.org/10.48550/arXiv.2411.12527

Hayot, E. (2024). Only persons intend. *Poetics Today*, *45*(2), 275–281. https://doi.org/10.1215/03335372-11092898

Hutchins, E. (1995). *Cognition in the wild*. MIT Press. https://doi.org/10.7551/mitpress/1881.001.0001

Jardonnet, E. (2022, June 8). La justice donne raison à Maurizio Cattelan contre Daniel Druet. Le Monde. https://www.lemonde.fr/culture/article/2022/07/08/la-justice-donne-raison-a-maurizio-cattelan-face-a-daniel-druet_6134012_3246.html

Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory*. Oxford University Press. https://doi.org/10.1093/oso/9780199256044.001.0001

Leroi-Gourhan, A. (1993). *Gesture and speech* (B. B. Halpern, Trans.). MIT Press. (Original work published 1964–1965).

Manovich, L., & Arielli, E. (2024). *Artificial aesthetics: A critical guide to AI, media and design*. https://manovich.net/index.php/projects/artificial-aesthetics

MediaLaws. (2025, August 28). Copyright and generative AI: The current U.S. panorama. *MediaLaws*. https://www.medialaws.eu/copyright-and-generative-ai-the-current-us-panorama/

McLuhan, M. (1964). *Understanding media: The extensions of man*. McGraw-Hill.

Nake, F. (2005). Computer art: A personal recollection. *Communications of the ACM*, *48*(11), 35–40. https://doi.org/10.1145/1056224.1056234

Naqvi, Z. (2020). Artificial intelligence, copyright, and copyright infringement. *Marquette Intellectual Property Law Review*, *24*(1), Article 4, 15–51. https://scholarship.law.marquette.edu/iplr/vol24/iss1/4/

Nyholm, S. (2020). *Humans and robots: Ethics, agency, and anthropomorphism*. Rowman & Littlefield International.

Offert, F., & Impett, L. (2025). *Vector media*. Meson Press/University of Minnesota Press.

Oppenlaender, J. (2023). A taxonomy of prompt modifiers for text-to-image generation. *Behaviour & Information Technology*, *43*(15), 3763–3776. https://doi.org/10.1080/0144929X.2023.2286532

Piekut, B. (2013). *Chance and certainty: John Cage's politics of nature*. *Cultural Critique*, *84*, 134–163. https://doi.org/10.5749/culturalcritique.84.2013.0134

Placani, A. (2024). Anthropomorphism in AI: Hype and fallacy. *AI and Ethics*, *4*, 691–698. https://doi.org/10.1007/s43681-024-00419-4

Riemer, K., & Peter, S. (2024). Conceptualizing generative AI as style engines: Application archetypes and implications. *International Journal of Information Management*, *79*, Article 102824. https://doi.org/10.1016/j.ijinfomgt.2024.102824

Roine, H.-R., & Piippo, L. (2021). Authorship vs. assemblage in digital media. In S. Lindberg & H.-R. Roine (Eds.), *The ethos of digital environments: Technology, literary theory and philosophy* (pp. 60–76). Routledge. https://doi.org/10.4324/9781003123996

Schmidt, P., & Loidolt, S. (2023). Interacting with machines: Can an artificially intelligent agent be a partner? *Philosophy & Technology*, *36*, Article 55. https://doi.org/10.1007/s13347-023-00656-1

Simondon, G. (2017). *On the mode of existence of technical objects* (C. Malaspina & J. Rogove, Trans.). Univocal Publishing. (Original work published 1958)

Suchman, L. A. (2007). *Human–machine reconfigurations: Plans and situated actions* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511808418

von Uexküll, J. (1926). *Theoretical biology* (D. L. MacKinnon, Trans.). Harcourt, Brace & Co.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wilf, E. (2014). Semiotic dimensions of creativity. *Annual Review of Anthropology*, *43*, 397–412. https://doi.org/10.1146/annurev-anthro-102313-030020

Young, N., & Terrone, E. (2025). Growing the image: Generative AI and the medium of gardening. *The Philosophical Quarterly*, *75*(1), 310–319. https://doi.org/10.1093/pq/pqae120

White House Congressional Research Service. (2025, July 18). *Generative artificial intelligence and copyright law* (CRS Legal Sidebar LSB10922). U.S. Government Printing Office. https://www.congress.gov/crs-product/LSB10922

Zhang, L., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models (ControlNet). *arXiv preprint* arXiv: 2302.05543. https://doi.org/10.48550/arXiv.2302.05543